

Advances in Science, Technology & Engineering Systems Journal

VOLUME 3-ISSUE 6 | NOV-DEC 2018

www.astesj.com

ISSN: 2415-6698

EDITORIAL BOARD

Editor-in-Chief

Prof. Passerini Kazmerski
University of Chicago, USA

Editorial Board Members

Prof. Rehan Ullah Khan
Qassim University, Saudi Arabia

Prof. María Jesús Espinosa
Universidad Tecnológica
Metropolitana, Mexico

Dr. Hongbo Du
Prairie View A&M University, USA

Dr. Nguyen Tung Linh
Electric Power University,
Vietnam

Tariq Kamal
University of Nottingham, UK

Sakarya University, Turkey

**Dr. Mohmaed Abdel Fattah
Ashabrawy**
Prince Sattam bin Abdulaziz
University, Saudi Arabia

**Mohamed Mohamed Abdel-
Daim**
Suez Canal University, Egypt

Dr. Omeje Maxwell
Covenant University, Nigeria

Prof. Majida Ali Abed Meshari
Tikrit University Campus, Iraq

Dr. Heba Afify
MTI university, Cairo, Egypt

Regional Editors

Dr. Hung-Wei Wu
Kun Shan University, Taiwan

Dr. Maryam Asghari
Shahid Ashrafi Esfahani, Iran

Dr. Shakir Ali
Aligarh Muslim University, India

Dr. Ahmet Kayabasi
Karamanoglu Mehmetbey
University, Turkey

Dr. Ebubekir Altuntas
Gaziosmanpasa University,
Turkey

Dr. Sabry Ali Abdallah El-Naggar
Tanta University, Egypt

Dr. Shagufta Haneef
Aalborg University, Denmark

Dr. Gomathi Periasamy
Mekelle University, Ethiopia

Dr. Walid Wafik Mohamed Badawy
National Organization for Drug Control
and Research, Egypt

Aamir Nawaz
Gomal University, Pakistan

Abdullah El-Bayoumi
Cairo University, Egypt

Ayham Hassan Abazid
Jordan university of science and
technology, Jordan

Dr. Abhishek Shukla
R.D. Engineering College, India

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 60 accepted papers in Computer Science and Civil Engineering domains.

Editor-in-chief

Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 3 Issue 6

November-December 2018

CONTENTS

<i>A Comparative Study of a Hybrid Ant Colony Algorithm MMACS for the Strongly Correlated Knapsack Problem</i>	01
Wiem Zouari, Ines Alaya, Moncef Tagina	
<i>Low-Dimensional Spaces for Relating Sensor Signals with Internal Data Structure in a Propulsion System</i>	23
Catherine Cheung, Nicolle Kilfoyle, Julio Valdés, Srishti Sehgal, Richard Salas Chavez	
<i>Fuzzy Uncertainty Management in Multi-Shift Single-Vehicle Routing Problem</i>	33
Francesco Nucci	
<i>Holistic Access Control and Privacy Infrastructure in Distributed Environment</i>	46
Uche Magnus Mbanaso, Gloria A Chukwudebe	
<i>A Holistic User Centric Acute Myocardial Infarction Prediction System With Model Evaluation Using Data Mining Techniques</i>	56
Procheta Nag, Saikat Mondal, Arun More	
<i>Similarity-based Resource Selection for Scientific Workflows in Cloud Computing</i>	67
Takahiro Koita, Yu Manabe	
<i>Visualizing Affordances of Everyday Objects Using Mobile Augmented Reality to Promote Safer and More Flexible Home Environments for Infants</i>	74
Miho Nishizaki	
<i>Enhanced Ship Energy Efficiency by Using Marine Box Coolers</i>	83
Abdallah Aijjou, Lhoussain Bahatti, Abdelhadi Raihani	
<i>A Resolution-Reconfigurable and Power Scalable SAR ADC with Partially Thermometer Coded DAC</i>	89
Hao-Min Lin, Chih-Hsuan Lin, Kuei-Ann Wen	
<i>A Novel Technique for Enhancing Color of Undersea Deblurred Imagery</i>	97
Chrispin Jiji, Nagaraj Ramrao	
<i>Analysis and Methods on The Framework and Security Issues for Connected Vehicle Cloud</i>	105
Lin Dong, Akira Rinoshika	

<i>Non-bearing Masonry Walls Behavior and Influence to High Reinforced Concrete Buildings</i> Sorina Constantinescu	111
<i>Slender Confined Masonry Buildings in High Seismic Areas</i> Sorina Constantinescu	118
<i>Masonry Walls Behavior in Predominant Frames Structures</i> Sorina Constantinescu	124
<i>Impacts of Synchronous Generator Capability Curve on Systems Locational Marginal Price through a Convex Optimal Power Flow</i> Italo Fernandes	131
<i>A Development of Agility Mode in Cardiopulmonary Resuscitation Learning Support System Visualized by Augmented Reality</i> Keisuke Fukagawa, Yuima Kanamori, Akinori Minaduki	136
<i>Analysis Refactoring with Tools</i> Zhala Sarkawt Othman	140
<i>Management Tool for the “Nephele” Data Center Communication Agent</i> Angelos Kyriakos, Thomas Tsavalos, Dionysios Reisis	144
<i>Actual Use and Continuous Use of Retail Mobile App: A Model Comparison Perspective</i> Sunday Adewale Olaleye, Ismaila Temitayo Sanusi, Bisola Adepoju	151
<i>cv4sensorhub – A Multi-Domain Framework for Semi-Automatic Image Processing</i> Kristóf Csorba, Ádám Budai	159
<i>Medium Height Dual Buildings with Masonry and Concrete Walls in High Seismic Areas</i> Sorina Constantinescu	165
<i>Linear Evaluation on Weak Story Medium Rise Structures Placed in High Seismic Areas</i> Sorina Constantinescu	173
<i>A Wi-Fi based Architecture of a Smart Home Controlled by Smartphone and Wall Display IoT Device</i> Tareq Khan	180
<i>Prospects of Wind Energy Injection in the Brazilian National Interconnected System and Impacts Analysis Through a Quasi-Steady Power Flow</i> Italo Fernandes, David Melo, Gabriel Santana, Fernando Brito, Allisson Almeida	185

<i>Exploring the use of Manual Liquid Based Cytology, Cell Block with Immunomarkers p16/ki67, VIA and HPV DNA Testing as a Strategy for Cervical Cancer Screening in LMIC</i>	190
Nandini Nandish Manoli, Devananda Devegowda, AshokaVarshini, Pushkal Sinduvadi Ramesh, Sherin Susheel Mathew, Nandish Siddappa Manoli	
<i>Guidance Law Based on Line-of-Sight Rate Information Considering Uncertain Modeled Dynamics</i>	195
Saori Nakagawa, Takeshi Yamasaki, Hiroyuki Takano, Isao Yamaguchi	
<i>Attitude Control Simulation of a Legged Aerial Vehicle Using the Leg Motions</i>	204
Yoshiyuki Higashi, Soonki Chang	
<i>iSensA – A System for Collecting and Integrating Sensor Data</i>	213
João Manuel Leitão Pires Caldeira, Vasco Nuno da Gama de Jesus Soares, Pedro Miguel de Figueiredo Dinis Oliveira Gaspar, Joel José Puga Coelho Rodrigues, Ricardo Manuel Valentim Fontes, José Luís Lopes Silva	
<i>A Novel Fair and Efficient Resource Allocation Scheduling Algorithm for Uplink in LTE-A</i>	222
Havva Esra Bilisik, Radosveta Sokullu	
<i>Computational Techniques to Recover Missing Gene Expression Data</i>	233
Negin Fraidouni, Gergely Zaruba	
<i>Closed Approach of a Decoder Mobile for the 406 Mhz Distress Beacon</i>	243
Billel Ali Srihen, Jean-Paul Yonnet, Malek Benslama	
<i>An Improved Cross-Connection Abatement Algorithm with RSSI Using In-Band Magnetic Field Control in Densely Located LC Wireless Charger Environments</i>	247
Nam Yoon Kim, Jinsung Cho, Chang-Woo Kim	
<i>3D Reconstruction of Monuments from Drone Photographs Based on The Spatial Reconstruction of The Photogrammetric Method</i>	252
Andras Molnar	
<i>Student Performance Evaluation Using Data Mining Techniques for Engineering Education</i>	259
Veena Deshmukh, Srinivas Mangalwede, Dandina Hulikunta Rao	
<i>Parallelizing Combinatorial Optimization Heuristics with GPUs</i>	265
Mohammad Harun Rashid, Lixin Tao	
<i>Probabilistic Method for Anomalies Detection Based on the Analysis of Cyber Parameters in a Group of Mobile Robots</i>	281
Elena Basan, Alexander Basan, Oleg Makarevich	

<i>An Empirical Study of Icon Recognition in a Virtual Gallery Interface</i> Denise Ashe, Alan Eardley, Bobbie Fletcher	289
<i>Emergence of fun emotion in computer games -An experimental study on fun elements of Hanafuda-</i> Yuki Takaoka, Takashi Kawakami, Ryosuke Ooe	314
<i>Metaheuristics for Solving Facility Location Optimization Problem</i> Chika Yinka-Banjo, Babatunde Opesemowo	319
<i>Contract Price Model Under Active Demand Response</i> Zhijian. Liu, Ni. Xiao, Hui. Xu	324
<i>Machine Learning Applied to GRBAS Voice Quality Assessment</i> Zheng Xie, Chaitanya Gadepalli, Farideh Jalalinajafabadi, Barry M.G. Cheetham, Jarrod J. Homer	329
<i>MRI images Enhancement and Brain Tumor Segmentation</i> Aye Min, Zin Mar Kyu	339
<i>Semi-Autonomous Robot Control System with an improved 3D Vision Scheme for Search and Rescue Missions. A joint research collaboration between South Africa and Argentina</i> Jorge Alejandro Kamlofsky, Nicol Naidoo, Glen Bright, Maria Lorena Bergamini, Jose Zelasco, Francisco Ansaldo, Riaan Stopforth	347
<i>A Practical Approach for Extending DSMLs by Composing their Metamodels</i> Anas Abouzahra, Ayoub Sabraoui, Karim Afdel	358
<i>Modeling an Energy Consumption System with Partial-Value Data Associations</i> Nong Ye, Ting Yan Fok, Oswald Chong	372
<i>Using Fuzzy PD Controllers for Soft Motions in a Car-like Robot</i> Paolo Mercorelli	380
<i>Robot Self-Detection System</i> Ivaylo Penev, Milena Karova, Mariana Todorova, Danislav Zhelyazkov	391
<i>Analysis of Garri Frying Machine Manufacturing in Nigeria: Design Innovation</i> Rufus Ogbuka Chime, Odo Fidelis O	403
<i>Two Degree-of-Freedom Vibration Control of a 3D, 2 Link Flexible Manipulator</i> Waweru Njeri, Minoru Sasaki, Kojiro Matsushita	412
<i>Adaptation of Electronic Book Publishing Technology by The Publishers in Southeast Nigeria</i> Godson Emeka Ani, Chike Ogboh	425

<i>An Integrated & Secure System for Wearable Devices</i>	432
Callum Owen-Bridge, Stewart Blakeway, Emanuele Lindo Secco	
<i>Study on CD ROADM Contention Blocking</i>	438
Guangzhi Li, Kerong Yan, Li Huang, Bin Xia, Fanhua Kong, Yang Li	
<i>Real Time Eye Tracking and Detection- A Driving Assistance System</i>	446
Sherif Said, Samer AlKork, Taha Beyrouthy, Murtaza Hassan, OE Abdellatif, M Fayek Abdraboo	
<i>Simulation-Optimisation of a Granularity Controlled Consumer Supply Network Using Genetic Algorithms</i>	455
Zeinab Hajiabolhasani, Romeo Marian, John Boland	
<i>CNN-based Automatic Coating Inspection System</i>	469
Lili Liu, Estee Tan, Zhi Qiang Cai, Xi Jiang Yin, Yongda Zhen	
<i>Multi-Objective Path Optimization of a Satellite for Multiple Active Space Debris Removal Based on a Method for the Travelling Serviceman Problem</i>	479
Masahiro Kanazaki, Yusuke Yamada, Masaki Nakamiya	
<i>Perfect Molding Challenges and The Limitations “A Case Study”</i>	489
Tan Lay Tatt, Lim Boon Huat, Rosli Muhammad Tarmizi, T. Joseph Sahaya Anand	
<i>Omni-directional Dual-Band Patch Antenna for the LMDS and WiGig Wireless Applications</i>	496
Mourad S. Ibrahim	
<i>Building an Online Interactive 3D Virtual World for AquaFlux and Epsilon</i>	501
Omar Al Hashimi, Perry Xiao	
<i>PID-Type FLC Controller Design and Tuning for Sensorless Speed Control of DC Motor</i>	515
Abdullah Y. Al-Maliki, Kamran Iqbal	

A Comparative Study of a Hybrid Ant Colony Algorithm MMACS for the Strongly Correlated Knapsack Problem

Wiem Zouari*, Ines Alaya, Moncef Tagina

COSMOS Laboratory, National School of Computer Science University of Manouba, 2010, Tunisia

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 22 October, 2018

Online: 01 November, 2018

Keywords:

Ant colony optimization

Hybrid metaheuristic

Stochastic greedy approach

Local search

Strongly Correlated Knapsack Problem

Empirical study

ABSTRACT

Metaheuristic hybridization has recently been widely studied and discussed in many research works as it allows benefiting from the strengths of metaheuristics by combining adequately the different algorithms. MMACS is a new hybrid ant colony optimization algorithm based on the foraging behavior of ants. This algorithm presents two hybridization levels. The first hybridization consists in integrating the Ant Colony System selection rule in MAX-MIN Ant System. The second level of hybridization is to combine the hybridized ACO and an algorithm based on a local search heuristic, then both algorithms are operating sequentially. The optimal performance of MMACS algorithm depends mainly on the identification of suitable values for the parameters. Therefore, a comparative study of the solution quality and the execution time for MMACS algorithm is presented. The aim of this study is to provide insights towards a better understanding of the behavior of the MMACS algorithm with various parameter settings and to develop parametric guidelines for the application of MMACS to the Strongly Correlated Knapsack Problems. Results are compared with well-known Ant Colony Algorithms and recent methods in the literature.

1 Introduction

MMACS algorithm is a hybrid metaheuristic that was proposed in a previous work [1] and employed to solve one of the most complex variants of the knapsack problem which is the Strongly Correlated Knapsack Problem (SCKP). The proposed approach combines a proposed Ant Colony Optimization algorithm (ACO) with a 2-opt algorithm. The proposed ACO scheme combines two ant algorithms: the MAX-MIN Ant System and the Ant Colony System. At a first stage, the proposed ACO aims to solve the SCKP to optimality. In case an optimal solution is not found, a proposed 2-opt algorithm is used. Even if the 2-opt heuristic fails to find the optimal solution, it would hopefully improve the solution quality by reducing the gap between the found solution and the optimum.

An optimal resolution of a combinatorial optimization problem by applying an approximate method requires an adequate balance between exploitation of the best available solutions and wide exploration of the research space. On the one hand, the aim of exploitation is to intensify the research around the most promising

areas of the research space, which are in most cases close to the best-found solutions. On the other hand, it comes to diversifying the research by encouraging the exploration in order to discover new and better areas of the research space. The behavior of ants in relation to this duality between exploitation and exploration can be affected by the adjustment of the parameter values.

A comparative study was conducted on the hybrid ant colony algorithm MMACS. Firstly, this study is intended to present the behavior of MMACS algorithm and its dependencies on the values given to parameters while solving SCKP. Secondly, a comparison of performances of MMACS algorithm and two well-known Ant Colony Algorithms: the Max-Min Ant System (MMAS) and the Ant Colony System (ACS) was provided. Finally, MMACS algorithm was compared with two recent state of art algorithms that show significant results when solving the SCKP to optimality.

The paper has been organized as follows. In the next section, we define the Strongly Correlated Knapsack Problem. We present the studied Ant Colony Opti-

*Corresponding Author: Wiem Zouari, wiem.zouari@ensi-uma.tn

mization algorithms (ACO) in section 2. In section 3, we introduce the parameters in ACO. Then, we define the local search in section 4. Finally, experiments and results are presented in section 5.

2 Strongly Correlated Knapsack Problem (SCKP)

The SCKP problem is a NP-hard problem, whose goal is to find a subset of items that maximizes an objective function while satisfying resource constraints. In SCKP, the profit of each item is linearly related to its weight, in other words, the profit of an item is equal to its weights plus a fixed constant. The complexity of this problem compared to the classical knapsack problem resides in the strong correlation between the variables that characterize the problem. According to Pisinger in [2], the strongly correlated instances are hard to solve for two reasons. First, they are badly conditioned in the sense that there is a large gap between the continuous and integer solution of the problem. Then, sorting the items according to decreasing efficiencies corresponds to a sorting according to the weights. Thus, for any small interval of the ordered items (i.e. a "core"), there is a limited variation in the weights, making it difficult to equally satisfy the capacity constraint. SCKP can be formulated as follows:

$$\max \sum_{i=1}^n (w_i + k)x_i \quad (1)$$

subject to constraint:

$$\sum_{i=1}^n w_i x_i \leq c$$

$$x_i \in \{0, 1\}, i = 1, \dots, n$$

where x_i is a decision variable associated with an item i , which has value 1 if the item is selected and 0 otherwise, w_i is the weight of the item i uniformly random $[1, R]$, k is a positive constant, c is the knapsack capacity and n is the number of items. The capacity of the knapsack c is proposed by Pisinger in [2] and it is obtained as follows :

$$c = \frac{i}{S+1} \sum_{j=1}^n w_j \quad (2)$$

where S is the series of instances as such, for each instance, a series of $S = 100$ instances is performed and $i = 1, \dots, S$ corresponds to the test instance number. From equation (1), the profit $profit(x)$ of a solution x can be described as follows:

$$profit(x) = \sum_{i=1}^n w_i x_i + kb \quad (3)$$

where b is the number of items in x . According to equation (3), the maximization of the

$profit(x)$ means the maximization of the number of the selected items. In other words, items with the lowest weights should be first selected until the sum of weights is about to exceed the capacity c . This can be achieved through the use of greedy algorithms [3, 4]. However, greedy does not guarantee optimal solutions since it chooses the locally most attractive item with no concern for its effect on global solutions.

The convergence to local optima caused by greedy algorithms, called stagnation, should be avoided, hence the idea of alternation between greedy and stochastic approaches.

The proposition of Pisinger in [5] is one of the most well-known works that shows significant results when solving the SCKP to optimality. Pisinger proposed a specialized algorithm for this problem where he used a surrogate relaxation to transform the problem into a Subset-sum problem. He started the resolution by applying a greedy algorithm, then he used a 2-optimal heuristic and a dynamic programming algorithm to solve the problem to optimality. More recently, Han [6] proposed an evolutionary algorithm inspired by the concept of quantum computing. The study in [6] shows that the proposed algorithm, called Quantum-Inspired Evolutionary Algorithm (QEA), can find high quality results when solving the strongly correlated knapsack problems.

3 Ant Colony Optimization (ACO)

The ACO [7, 8] is a constructive population-based metaheuristic inspired from the real ants' behavior, seeking an adequate path between their colony and a food source, which is often the shortest path. The communication between ants is mediated by trails of a chemical substance called pheromone. Several ant colony optimization algorithms have been proposed in the literature. In this section, we present the Max-Min Ant System proposed by Stützle and Hoos [9, 10], the Ant Colony System proposed by Gambardella and Dorigo [11] and a recent hybrid ant colony algorithm called MMACS.

3.1 MMAS

The Max-Min Ant System [9, 10] is one of the most effective solvers of certain optimization problems. In MMAS, ants apply a random proportional rule to select the next item. The probabilistic action choice rule is defined as follows:

$$P_{ij}^k = \frac{[\tau_{ij}]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}]^\alpha \cdot [\eta_{il}]^\beta} \quad (4)$$

where τ and η are successively the pheromone factor and the heuristic factor, α and β are two parameters that determine the relative influence of the pheromone trail and the heuristic information and N_i^k is the feasible neighborhood of an ant k that selected an item i and chooses to select an item j . Besides, MMAS exploits the best solutions found by letting only the best

ant deposit pheromone. This best ant can be the one which found the best solution during the last iteration or the one which found the solution from the beginning of the execution. The pheromone update can be formulated as follows:

1. Pheromone evaporation applied to all components:

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} \quad (5)$$

2. Pheromone update applied to components selected by the best ant:

$$\tau_{ij} \leftarrow \tau_{ij} + \Delta_{ij}^{bs} \quad (6)$$

Then, MMAS introduces bounds to limit the range of pheromone trails to $[\tau_{min}, \tau_{max}]$ in order to escape a stagnation that can be caused by an excessive growth of pheromone trails. These pheromone trails are initialized, at the beginning, to upper pheromone trail limit to ensure the exploration of the research space, and reinitialized when system approaches stagnation.

3.2 ACS

The ACS algorithm [11] achieves performance improvement through the use of a more aggressive action choice rule. In ACS, ants choose items according to an aggressive action choice rule called pseudorandom proportional rule given as follows:

$$s = \begin{cases} \mathit{argmax}_{i \in N_i^k} \{ \tau_{il} \cdot [\eta_{il}]^\beta \} & \text{if } q \leq q_0 \\ S & \text{otherwise} \end{cases} \quad (7)$$

where q is a random variable uniformly distributed in $[0, 1]$, q_0 ($0 \leq q_0 \leq 1$) and S is a random variable selected according to the probability distribution given in equation (4). Besides, only one ant called the best-so-far-ant is allowed to deposit pheromone after each iteration. Thus, the global pheromone trail update is given as follows:

$$\tau_{ij} \leftarrow (1 - \rho) \cdot \tau_{ij} + \rho \cdot \Delta_{ij}^{bs}(s) \quad (8)$$

This pheromone trail update is applied to only components in the best-so-far solution, where the parameter ρ represents pheromone evaporation. In addition to the global pheromone update, ants use a local pheromone update rule that is applied immediately after choosing a new item during the solution construction, given as follows:

$$\tau_{ij} \leftarrow (1 - \epsilon) \cdot \tau_{ij} + \epsilon \cdot \tau_0 \quad (9)$$

where $0 < \epsilon < 1$ and τ_0 are two parameters such that the τ_0 value is equal to the initial value of the pheromone trails which is 0.1. The local update happens during the solution construction in order to prevent other ants to make the same choices. This increases the exploration of alternative solutions.

3.3 MMACS

The MMACS algorithm combines Max-Min Ant System with Ant Colony System and an algorithm based on the 2-opt heuristic. In fact, the scheme of MMACS is based on ACO scheme presented in [12] and ACS scheme presented in [11] in a way that it uses an MMAS pheromone update rule and a choice rule inspired from the ACS aggressive action choice rule. In MMACS, the minimum and the maximum pheromone amounts are limited to an interval $[\tau_{min}, \tau_{max}]$, like MMAS, in order to avoid premature stagnation. Initially, the pheromone trails are set to τ_{max} . After the construction of all solutions in one cycle, the best ant updates the pheromone trails by applying a rule similar to MMAS pheromone update rule. Indeed, once all ants finish the solutions construction, the pheromone trails are decreased to simulate evaporation by multiplying each component by a pheromone persistence ratio equal to $(1 - \rho)$ where $0 \leq \rho \leq 1$ as given by equation (5). After that, an amount of pheromone is laid on the best solution found by ants by applying the pheromone update rule given in equation (6), where the amount of pheromone trails is calculated as follows:

$$\Delta \tau_{S_k} = \frac{1}{1 + \mathit{profit}(S_{best}) - \mathit{profit}(S_k)} \quad (10)$$

S_{best} represents the best solution built since the beginning and S_k is the best solution of a cycle.

Besides, in MMACS, each ant constructs a solution by applying the choice rule, where the decision making is based on both:

1. A random proportional rule that selects a random item using the probability distribution.
2. A guided selection rule that chooses the next item as the best available option.

Like ACS, MMACS balances between greedy and stochastic approaches by applying the pseudorandom proportional rule (7). Actually, at each construction step, an ant k chooses a random variable q uniformly distributed in $[0, 1]$. If q is less than a fixed parameter q_0 such as $0 \leq q_0 \leq 1$, the ant makes the best possible choice as indicated by the pheromone trails and the heuristic information (exploitation) else, with a probability $1 - q_0$, the ant applies the random proportional rule (4) to select the next item (biased exploration).

The heuristic factor used in the probability rule (4) is given as follows:

$$\eta_{S_k}(o_j) = \frac{d_{S_k}}{w_j} \quad (11)$$

where d_{S_k} is the remaining capacity when an ant k built a solution S_k and it is given as follows:

$$d_{S_k} = c - \sum_{g \in S_k} w_g \quad (12)$$

As shown in equation (11), the heuristic information value and the item weight are inversely proportional. Consequently, the more the weight value decreases the

more the heuristic information value increases. Added to that, the closer the remaining capacity and the item weight are, the more the heuristic information value increases. This can be helpful at the end of the execution when the knapsack is about to be filled. At last, the execution of MMACS algorithm ends either when an optimum is found, or in the worst cases, it ends after a fixed number of iterations. The pseudo-code of MMACS algorithm is represented by algorithm 1.

Algorithm 1 MMACS pseudo-code applied to SCKP

```

Initialize pheromone trails to  $\tau_{max}$ 
repeat
  repeat
    Construct a solution
    Update  $S_{best}$ 
  until maximum number of ants is reached or optimum is found
  Update pheromone trails
until maximum number of cycles is reached or optimum is found
Apply a local search algorithm

```

S_{best} is the best solution found all along the execution.

The construction procedure can be represented by algorithm 2.

Algorithm 2 Construct Solution

```

Select randomly a first item
Remove from candidates each item that violates resource constraints
while  $Candidates \neq \emptyset$  do
  if a randomly chosen  $q$  is greater than  $q_0$  then
    Choose item  $o_j$  from Candidates with probability  $P_{ij}^k$ 
  else
    Choose the next best item
  end if
  Remove from candidates each item that violates resource constraints
end while

```

4 Parameters in ACO

In the ACO algorithm, relevant parameters that request reasonable settings are the heuristic information parameter β , the pheromone parameter α and the pheromone evaporation rate ρ . Those parameters can influence the algorithm performance by improving its convergence speed and its global optimization ability.

4.1 The heuristic information parameter β

The ants' solution construction is biased by a heuristic value η that represents the attractiveness of each item.

The parameter β determines the relative importance of this heuristic value. In our case, the heuristic information makes items characterized by little weights as desirable choices. In other words, the increase in the β value can be triggered by the selection of items which have little weights. This behavior is close to that of greedy algorithm. However, the decrease in the value of β makes the heuristic factor unprofitable. As a result, ants fall easily in local optima.

4.2 The pheromone parameter α

Besides heuristic value, the ants' solution construction is influenced by the pheromone trails. The pheromone parameter α determines the relative influence of the pheromone trails τ . Indeed, the parameter α reflects the importance of the amplification of pheromone amounts. In other words, the increase of α favors the choice of items associated with uppermost pheromone trails values. In case the value of α is considerable, ants tend to choose the same solution components. This behavior is caused by the strong cooperation between them so ants drift towards the same part of the search space. In such case, the convergence speed of algorithm accelerates and consequently, it causes the fall in local optima. This gives rise to the need to prevent this premature convergence. Then, several tests were conducted to evaluate the influence of α on solutions' quality. Additional experiments were carried out to examine the similarity of solutions in one cycle. The analysis of the similitude of solutions allows appropriately assigning the value of α in order to avoid the premature stagnation of the search that can be caused by the excessive reliance upon pheromone trails at the expense of the heuristic information.

4.3 The pheromone evaporation rate ρ

The amount of pheromone decreases to simulate evaporation by multiplying each component by a constant evaporation ratio equal to $1 - \rho$. This pheromone trails reduction gives ants the possibility of abandoning bad decisions previously taken. In fact, the pheromone value of an unchosen item decreases exponentially with the number of iterations.

5 Local search

Local search algorithms are usually used in most applications of ACO to combinatorial optimization problems in order to improve solutions found by ants. Among those algorithms, we cite 2-opt heuristic. The 2-opt [13] is a simple local search algorithm. When applied to knapsack problems, it consists of exchanging an item present in the current solution with another that is not part of this solution in order to improve it. The new solution should satisfy constraints and it would be better or equal to the old one. In other words, the 2-opt algorithm takes a current solution as input and returns a better accepted solution to the problem,

Table 1: Default parameter settings for MMAS, ACS and MMACS algorithms

ACO algorithm	α	β	ρ	<i>ants</i>	<i>cycles</i>	q_0
MMAS	1	2	0.02	n	20	-
MMAS + 2-opt	1	2	0.2	25	20	-
ACS	-	2	0.1	10	20	0.9
ACS + 2-opt	-	2	0.1	10	20	0.98
MMACS	1	5	0.02	20	20	0.9

if it exists. The 2-opt algorithm is used once the ants have completed their solution construction, thereby improving the solution by approaching the best one or even reaching it. Our proposed 2-opt algorithm can be written as represented by algorithm 3.

Algorithm 3 A 2-opt pseudo-code applied to SCKP

```

Initialize Candidates by observing  $S_{best}$ 
repeat
  for each item  $o_j \in Candidates$  do
    for each item  $o_i \in S_{best}$  do
       $S'_{best} = \text{Swap}(o_i, o_j)$ 
      if constraints are satisfied by  $S'_{best}$  and  $S'_{best}$  is
      better than  $S_{best}$  then
        Update best solution
      end if
    end for
  end for
until no improvement is made

```

6 Computational results

In this section, we study the results of a set of experiments that was carried out to determine the efficacy of the MMACS algorithm. The proposed algorithm was programmed in C++, and compiled with GNU g++ on an Intel Core i7-4770 CPU processor (3.40 GHz and 3.8 GB RAM).

Through the experimentations, we analyze the influence of the parameters' selection on the MMACS performances. Then, we identify the convenient parameter settings that produce better results. Those parameter settings are employed for the rest of the experiments. At a later stage, we compare the results of MMACS with those of the two well-known ant colony algorithms: MMAS [9, 10] and ACS [11]. After that, the results of MMACS are compared with those of the evolutionary algorithm QEA in [6] and to the optimal values.

6.1 Benchmark instances

In order to evaluate the performance of the MMACS algorithm, experiments were conducted on two sets of instances.

6.1.1 Pisinger Set

The first set contains 100 different instances with n items, where the number of items n varies from 50 to 2000. Those benchmark instances used in comparison with algorithms in [5] and [6] are available at the website (<http://www.diku.dk/pisinger/codes.html>).

6.1.2 Generated Set

The second set regroups 3 different instances having the number of items equal to 100, 250 and 500, respectively. Those instances used in comparison with the proposed algorithm in [6], were randomly generated using a generator similar to the one in [2].

6.2 Parametric analysis of MMACS

In order to evaluate the influence of parameters' values on MMACS performances, we conducted tests for different values of parameters and compared the obtained results. The experiments were realized on the Pisinger set instances of size 50, 100, 200 and 500, and for each instance, we applied 10 runs (10 runs * 100 instances * 4 knapsack problems). In each of these experiments, we fixed the parameters to their default values and we made the variation of the studied parameter. In fact, the MMAS and ACS default parameter settings were recommended by the authors in [14]. The default parameter settings are given in Table 1 and the various values for each parameter are presented in Table 2. Tables 3- 29 report the results of MMACS algorithm in response to the variation of the parameters. N is the number of items and R is the range of coefficients. Then, the presentation of the data is visualized using different curves. Figures 1- 5 present the effect of the studied parameters on MMACS performances. The abscissa axis of curves presented in those figures shows the instances' size that varies between 50 and 500. In each figure, left and right plots show the behavior of MMACS algorithm while solving the SCKP instances having the range of coefficients equal to 1000 and 10000, respectively. In those curves, we examine the percentage of exact solutions, the relative deviation of the best solution found by MMACS from the optimal solution value and the execution time in terms of the studied parameter.

Table 2: Parameter settings used in experiments for MMACS algorithm

Control parameter	Value
α	1, 2, 3, 4, 5
β	1, 2, 3, 4, 5
ρ	0.01, 0.02, 0.4, 0.5, 0.8, 1
q_0	0, 0.5, 0.75, 0.9, 0.99
m	1, 5, 10, 20, 50, 1000

6.2.1 Influence of parameter α

We set the value of β to 5 and ρ to 0.02. After that, we make the change of the value of α in order to study its influence on the solutions' quality and the execution time. Tables 3- 6 represent the results after applying MMACS to SCKP where α varies between 1 and 5. Results are visualized using curves in Figure 1. In fact, the differences among the various settings of α are almost insignificant. However, the value of α equal to 1 gives better results in terms of the most reduced execution time. Additional experiments are conducted to analyze the similitude of solutions in one cycle in order to appropriately assign the value of α . In fact, we propose to calculate a similarity ratio proposed in [15]. The similarity ratio associated with a set of solutions S is defined as follows:

$$ratio = \frac{\sum_{v_i \in V} (freq[i] \cdot (freq[i] - 1))}{(|S| - 1) \cdot \sum_{S_k \in S} |S_k|} \quad (13)$$

where V is a set of items and $freq[i]$ is the frequency of the object v_i in the solutions of S . Thus, this ratio is equal to 1 if all the solutions of S are identical and it is equal to 0 if the intersection of solutions of S is empty. In those experiments, the number of cycles varies from 10 to 25. For each number of cycles, each ant's solution was compared with others. Experiments are conducted on the Pisinger set instances of size 100 and for both range of coefficients 1000 and 10000. Figure 2 shows the influence of the pheromone trails on the similarity of the solutions built during the execution of the MMACS algorithm, and for different values of the α parameter that determine the influence of the pheromone on the behavior of the ants. The curves show that the solutions are remarkably similar where the similarity ratio varies between the values 0.6 and 0.8 beginning with the number of cycles at about 15. In other words, ants are not deeply influenced by pheromone trails and they are obviously focusing in a short time on a very small area of the research space that they explore intensely.

6.2.2 Influence of parameters β and ρ

In this section, we examine the influence of different values of the heuristic information parameter β and the pheromone evaporation rate ρ . In this context, we fix the value of α to 1 then we make a simultaneous variation of the two variables β and ρ . We modify the values of β in order to control the influence of the heuristic information and to examine its effect on the MMACS performances. Besides, we make the variety of ρ values in order to study its influence on the items' selection and consequently on the MMACS performances. Tables 7- 21 present results after applying MMACS to SCKP where β varies between 2 and 5 and ρ values are 0.01, 0.02, 0.4, 0.5, 0.8 and 1.

Results of MMACS algorithms with the fixed parameter β are almost similar to all values of ρ .

The MMACS algorithms with the fixed parameter β_1 and β_2 work in a similar way. In fact, the results

show that the percentages of exact solutions decrease considerably in terms of number of items for both ranges of coefficients 1000 and 10000. Consequently the values of gap increases. Besides, the execution time results vary in the same way for the six values of ρ .

The MMACS algorithms with the fixed parameter β_3 and β_4 behave in a similar way for almost all problems with both ranges of coefficients 1000 and 10000. Results show that the percentages of exact solutions for both algorithms increase for number of items between 50 and 200. Then these percentages decrease considerably for the large number of items 500. Consequently, the gap results progress in a reverse way.

However, the MMACS algorithm with the fixed heuristic parameter β_5 shows acceptable results. Results are presented in Figure 3, each curve represents a value of ρ . The curves have almost the same evolutions with insignificant differences between the values. In fact, the percentage of exact solution presents an increase in terms of number of items. Besides, the gap results show a remarkable decrease for large instances. Then, the execution time has the same variation for all values of ρ .

However, the value of ρ_2 equal to 0.02 selected by MMACS can be modified to ρ_1 in order to make a slight improvement in gap values for large instances with a range of coefficients equal to 10000.

6.2.3 Influence of parameter q_0

In MMACS, we study the effect of q_0 that represents the probability of selecting the best available choices in equation 7. Results are given in Tables 22- 25 and presented in curves of Figure 4. The curves in Figure 5 are associated with the values of q_0 : 0, 0.5, 0.75, 0.9 and 0.99. For all instances, among those values only 0.9 and 0.99 show an increase in terms of the percentage of the exact solutions. As regards the other values of q_0 MMACS does not succeed in finding, in most cases, the exact solution. This is clearly represented by its decreasing curves. As to the execution time, the five curves are growing in almost the same way. However, the curve that corresponds to q_0 value equal to 0.9 reaches the lowest values. For large instances, the differences between the curves are significant. The best results that correspond to the lowest gap values is represented by the curve associated to the value of q_0 equal to 0.9. We conclude that MMACS has the same behavior as ACS regarding the q_0 parameter, where the values close to 1 present the good ones as suggested in the literature.

6.2.4 Influence of colony size

Tables 26- 29 show the effect of colony size on the quality of the solutions. In these tables, the ant colony size m varies between 1 and 100. Results are compared in Figure 5. As shown by curves in Figure 5, the increase in the ants' number improves the percentage of exact solutions. This increase causes the growth of the execution time, although in practice, we generally seek to

Table 3: Percentage of exact solutions found by MMACS. The value of α varies between 1 and 5.

N	R	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$
50	1000	82.0%	79.0%	83.0%	77.0%	80.0%
	10000	52.0%	48.0%	51.0%	48.0%	46.0%
100	1000	90.0%	91.0%	93.0%	88.0%	91.0%
	10000	63.0%	60.0%	58.0%	65.0%	59.0%
200	1000	92.0%	94.0%	92.0%	90.0%	92.0%
	10000	79.0%	82.0%	82.0%	80.0%	84.0%
500	1000	98.0%	97.0%	97.0%	9.0%	93.0%
	10000	79.0%	78.0%	76.0%	76.0%	81.0%

Table 4: Percentage of perfect solutions found by MMACS. The value of α varies between 1 and 5.

N	R	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$
50	1000	52.0%	50.0%	54.0%	48.0%	52.0%
	10000	13.0%	12.0%	15.0%	12.0%	10.0%
100	1000	82.0%	84.0%	86.0%	81.0%	83.0%
	10000	46.0%	44.0%	41.0%	48.0%	41.0%
200	1000	88.0%	87.0%	88.0%	86.0%	87.0%
	10000	72.0%	75.0%	74.0%	73.0%	78.0%
500	1000	96.0%	94.0%	95.0%	96.0%	92.0%
	10000	78.0%	76.0%	75.0%	74.0%	80.0%

Table 5: Averages of solutions found by MMACS. The value of α varies between 1 and 5.

N	R	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$
50	1000	0.02934	0.05073	0.06385	0.02732	0.03439
	10000	0.02485	0.02534	0.02009	0.02104	0.02113
100	1000	0.00887	0.03124	0.01363	0.01113	0.01537
	10000	0.00450	0.00445	0.00534	0.00668	0.00497
200	1000	0.00248	0.00239	0.00149	0.00213	0.00198
	10000	0.00087	0.00143	0.00102	0.00154	0.00125
500	1000	0.00059	0.00102	0.00056	0.00220	0.00445
	10000	0.00020	0.00049	0.00064	0.00038	0.00312

Table 6: Execution time of MMACS. The value of α varies between 1 and 5.

N	R	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$
50	1000	0.05370	0.08514	0.07676	0.08809	0.08404
	10000	0.11360	0.20702	0.14372	0.14228	0.15597
100	1000	0.36749	0.35563	0.49085	0.34433	0.36030
	10000	0.73181	1.13194	1.67862	0.84895	0.97173
200	1000	1.64040	3.23461	2.36788	2.40549	2.40237
	10000	4.02055	7.20106	8.16290	4.69110	7.43489
500	1000	10.5397	48.9903	46.4050	37.3577	28.8331
	10000	38.1680	71.5076	64.0864	72.4135	59.0820

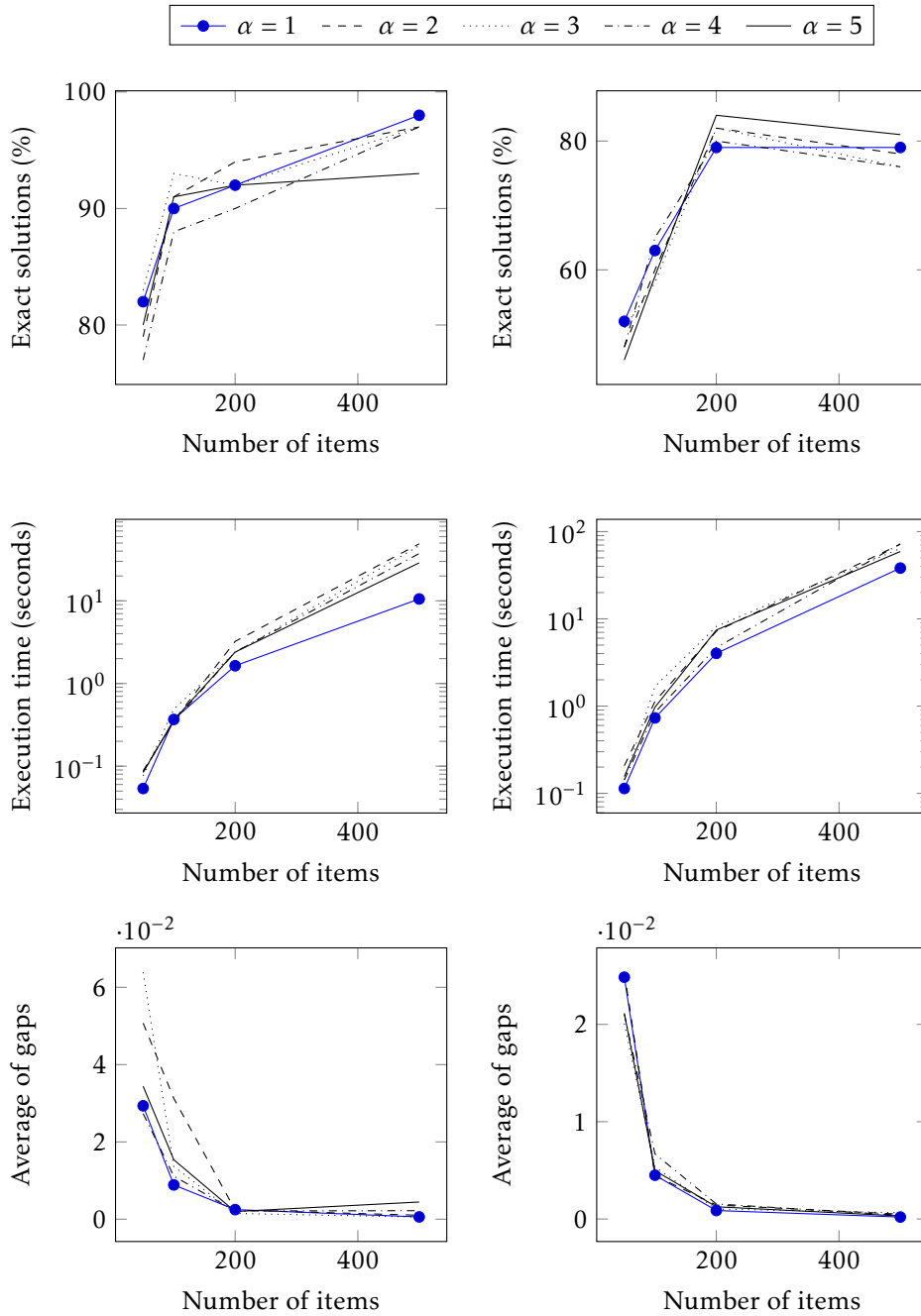


Figure 1: MMACS with various values of α . Plots on the left show results for SCKP instances with a range of coefficients equal to 1000 and plots on the right show results for SCKP instances with a range of coefficients equal to 10000.

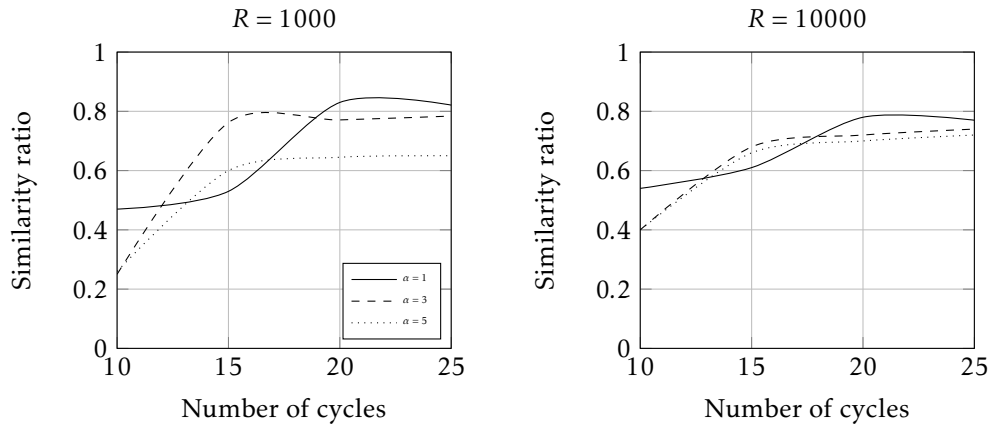


Figure 2: Influence of α on the similarity of solutions found by MMACS algorithm for SCKP instances with 100 items.

Table 7: Percentage of exact solutions found by MMACS. The heuristic information value is fixed to β_1 and the value of ρ varies between 0.01 and 1.

N	R	β_1					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	82.0%	81.0%	83.0%	84.0%	83.0%	84.0%
	10000	55.0%	54.0%	48.0%	55.0%	50.0%	57.0%
100	1000	91.0%	86.0%	90.0%	87.0%	88.0%	87.0%
	10000	53.0%	51.0%	57.0%	53.0%	54.0%	55.0%
200	1000	84.0%	84.0%	82.0%	82.0%	84.0%	81.0%
	10000	55.0%	49.0%	46.0%	52.0%	53.0%	51.0%
500	1000	35.0%	40.0%	41.0%	37.0%	37.0%	34.0%
	10000	24.0%	28.0%	23.0%	27.0%	22.0%	27.0%

Table 8: Averages of solutions found by MMACS. The heuristic information value is fixed to β_1 and the value of ρ varies between 0.01 and 1.

N	R	β_1					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.08984	0.02556	0.03288	0.03032	0.03363	0.03963
	10000	0.02220	0.02653	0.01738	0.02172	0.01726	0.01966
100	1000	0.02126	0.01814	0.01206	0.02616	0.01718	0.04464
	10000	0.00870	0.00930	0.01639	0.00883	0.00897	0.01018
200	1000	0.00832	0.04333	0.02262	0.04705	0.00998	0.02377
	10000	0.00676	0.01265	0.00538	0.00577	0.01255	0.01741
500	1000	0.00676	0.16979	0.15236	0.11632	0.15827	0.17096
	10000	0.11085	0.12811	0.13676	0.12798	0.11106	0.12794

Table 9: Execution time of MMACS. The heuristic information value is fixed to β_1 and the value of ρ varies between 0.01 and 1.

N	R	β_1					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.05492	0.05765	0.06821	0.05596	0.05465	0.05635
	10000	0.05567	0.05591	0.05756	0.05331	0.05382	0.05712
100	1000	0.39064	0.38973	0.76695	0.38287	0.39028	0.39951
	10000	0.38998	0.38882	0.77140	0.38358	0.39128	0.39009
200	1000	2.98575	3.00137	2.94311	2.93495	2.91098	2.95657
	10000	3.00461	2.95496	2.93763	2.91634	2.91860	2.95932
500	1000	48.0960	47.6133	46.3170	54.1633	46.7279	47.2380
	10000	47.9146	46.8829	46.5491	46.5299	46.7395	47.3001

Table 10: Percentage of exact solutions found by MMACS. The heuristic information value is fixed to β_2 and the value of ρ varies between 0.01 and 1.

N	R	β_2					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	80.0%	84.0%	82.0%	81.0%	84.0%	82.0%
	10000	52.0%	47.0%	51.0%	55.0%	54.0%	52.0%
100	1000	91.0%	93.0%	91.0%	90.0%	90.0%	91.0%
	10000	59.0%	59.0%	66.0%	60.0%	60.0%	55.0%
200	1000	86.0%	92.0%	90.0%	89.0%	90.0%	93.0%
	10000	62.0%	67.0%	68.0%	69.0%	65.0%	65.0%
500	1000	60.0%	63.0%	52.0%	58.0%	44.0%	59.0%
	10000	41.0%	37.0%	36.0%	36.0%	36.0%	33.0%

Table 11: Averages of solutions found by MMACS. The heuristic information value is fixed to β_2 and the value of ρ varies between 0.01 and 1.

N	R	β_2					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.03299	0.03350	0.03077	0.05647	0.05426	0.03865
	10000	0.01685	0.01709	0.01783	0.01535	0.02255	0.02233
100	1000	0.01535	0.02211	0.01249	0.01724	0.01435	0.01480
	10000	0.00785	0.00968	0.01008	0.00698	0.00798	0.01020
200	1000	0.00337	0.00443	0.00610	0.00417	0.00520	0.00423
	10000	0.00116	0.00208	0.00217	0.00180	0.00181	0.00275
500	1000	0.06840	0.08212	0.06738	0.06309	0.07076	0.06730
	10000	0.03134	0.05077	0.03474	0.03055	0.03297	0.02785

Table 12: Execution time of MMACS. The heuristic information value is fixed to β_2 and the value of ρ varies between 0.01 and 1.

N	R	β_2					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.05313	0.05536	0.05604	0.05722	0.05452	0.05769
	10000	0.05358	0.05588	0.05365	0.05522	0.05618	0.05340
100	1000	0.37927	0.39185	0.38238	0.38306	0.38370	0.38917
	10000	0.38958	0.38976	0.39358	0.38518	0.38692	0.38878
200	1000	2.91574	2.96851	2.92166	2.91118	2.96750	2.97729
	10000	2.93331	2.91021	2.91412	2.95463	2.92098	2.94065
500	1000	46.3217	68.3414	46.0147	48.9379	47.6850	48.1014
	10000	46.0396	45.7421	46.2623	48.0479	46.4917	54.1009

Table 13: Percentage of exact solutions found by MMACS. The heuristic information value is fixed to β_3 and the value of ρ varies between 0.01 and 1.

N	R	β_3					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	81.0%	84.0%	83.0%	83.0%	79.0%	85.0%
	10000	50.0%	48.0%	49.0%	50.0%	54.0%	53.0%
100	1000	95.0%	89.0%	89.0%	91.0%	90.0%	94.0%
	10000	59.0%	63.0%	62.0%	61.0%	56.0%	57.0%
200	1000	92.0%	89.0%	93.0%	94.0%	92.0%	93.0%
	10000	72.0%	80.0%	77.0%	72.0%	73.0%	74.0%
500	1000	84.0%	88.0%	85.0%	84.0%	80.0%	86.0%
	10000	52.0%	50.0%	53.0%	46.0%	46.0%	49.0%

Table 14: Averages of solutions found by MMACS. The heuristic information value is fixed to β_3 and the value of ρ varies between 0.01 and 1.

N	R	β_3					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.02858	0.03836	0.03311	0.03606	0.05618	0.03919
	10000	0.02232	0.02145	0.02609	0.02073	0.02204	0.02508
100	1000	0.02088	0.01689	0.01302	0.01218	0.01536	0.02067
	10000	0.00652	0.00897	0.00696	0.00931	0.00977	0.01037
200	1000	0.00353	0.00350	0.00293	0.00233	0.00261	0.00330
	10000	0.00093	0.00095	0.00117	0.00118	0.00161	0.00128
500	1000	0.02795	0.05289	0.04132	0.03866	0.03520	0.03838
	10000	0.01185	0.01029	0.01414	0.01199	0.00983	0.00629

Table 15: Execution time of MMACS. The heuristic information value is fixed to β_3 and the value of ρ varies between 0.01 and 1.

N	R	β_3					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.07581	0.07366	0.07465	0.07367	0.07425	0.07242
	10000	0.07422	0.07476	0.07528	0.07497	0.08080	0.07237
100	1000	0.46370	0.46119	0.46399	0.46445	0.45845	0.45893
	10000	0.46346	0.45978	0.46745	0.46602	0.46544	0.46089
200	1000	3.27010	3.26748	3.29506	3.30715	3.28653	3.24697
	10000	3.26838	3.27883	3.28464	3.30160	3.27514	3.24717
500	1000	50.0693	49.6441	49.2067	48.8005	48.5047	48.1483
	10000	49.8877	48.5299	48.9702	51.1495	48.8320	48.0786

Table 16: Percentage of exact solutions found by MMACS. The heuristic information value is fixed to β_4 and the value of ρ varies between 0.01 and 1.

N	R	β_4					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	84.0%	84.0%	82.0%	81.0%	80.0%	84.0%
	10000	51.0%	54.0%	51.0%	51.0%	51.0%	55.0%
100	1000	89.0%	90.0%	91.0%	91.0%	90.0%	93.0%
	10000	59.0%	61.0%	62.0%	67.0%	60.0%	53.0%
200	1000	91.0%	93.0%	92.0%	93.0%	91.0%	94.0%
	10000	83.0%	80.0%	81.0%	83.0%	78.0%	85.0%
500	1000	90.0%	93.0%	92.0%	89.0%	94.0%	85.0%
	10000	70.0%	70.0%	77.0%	65.0%	71.0%	74.0%

Table 17: Averages of solutions found by MMACS. The heuristic information value is fixed to β_4 and the value of ρ varies between 0.01 and 1.

N	R	β_4					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.03218	0.03467	0.03252	0.03324	0.03187	0.03196
	10000	0.01684	0.02137	0.02540	0.02021	0.02979	0.02631
100	1000	0.01555	0.02284	0.01867	0.02045	0.02287	0.01441
	10000	0.00541	0.01040	0.00831	0.00895	0.00965	0.00892
200	1000	0.00263	0.00192	0.00176	0.00350	0.00185	0.00218
	10000	0.00104	0.00155	0.00142	0.00118	0.00127	0.00097
500	1000	0.02596	0.02232	0.03127	0.01763	0.01168	0.01989
	10000	0.00606	0.00439	0.00595	0.01001	0.00683	0.01049

Table 18: Execution time of MMACS. The heuristic information value is fixed to β_4 and the value of ρ varies between 0.01 and 1.

N	R	β_4					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.07929	0.08029	0.07424	0.07557	0.10963	0.07293
	10000	0.07846	0.07948	0.07611	0.07561	0.11752	0.07161
100	1000	0.49581	0.48432	0.48316	0.48317	0.48473	0.45665
	10000	0.49517	0.47001	0.49464	0.48414	0.48985	0.45644
200	1000	3.38250	3.28004	3.46215	3.45131	3.09758	3.22518
	10000	3.51687	3.29262	3.30075	3.45131	3.46598	3.23389
500	1000	48.3098	48.9813	47.9923	47.9734	49.1355	47.8999
	10000	48.7001	48.5943	48.7463	48.2574	48.2564	47.7792

Table 19: Percentage of exact solutions found by MMACS. The heuristic information value is fixed to β_5 and the value of ρ varies between 0.01 and 1.

N	R	β_5					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	82.0%	81.0%	81.0%	83.0%	79.0%	82.0%
	10000	49.0%	50.0%	47.0%	51.0%	44.0%	51.0%
100	1000	89.0%	91.0%	89.0%	88.0%	90.0%	89.0%
	10000	55.0%	58.0%	61.0%	55.0%	53.0%	59.0%
200	1000	91.0%	92.0%	90.0%	93.0%	92.0%	91.0%
	10000	79.0%	76.0%	81.0%	79.0%	76.0%	81.0%
500	1000	97.0%	98.0%	96.0%	95.0%	98.0%	97.0%
	10000	80.0%	78.0%	79.0%	79.0%	81.0%	77.0%

Table 20: Averages of solutions found by MMACS. The heuristic information value is fixed to β_5 and the value of ρ varies between 0.01 and 1.

N	R	β_5					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.03247	0.03550	0.04044	0.06754	0.03757	0.03616
	10000	0.01910	0.03156	0.01838	0.01617	0.02353	0.02000
100	1000	0.01684	0.01483	0.01291	0.01588	0.01642	0.01395
	10000	0.00378	0.00786	0.00950	0.00682	0.00966	0.00802
200	1000	0.00218	0.00241	0.00284	0.00188	0.00210	0.00213
	10000	0.00129	0.00101	0.00097	0.00116	0.00158	0.00061
500	1000	0.00058	0.00188	0.00050	0.00265	0.00097	0.00283
	10000	0.00073	0.00460	0.00055	0.00069	0.00033	0.00348

Table 21: Execution time of MMACS. The heuristic information value is fixed to β_5 and the value of ρ varies between 0.01 and 1.

N	R	β_5					
		ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
50	1000	0.07230	0.06982	0.07270	0.07362	0.07413	0.07302
	10000	0.07589	0.06936	0.07335	0.07472	0.07412	0.07157
100	1000	0.46598	0.47126	0.43751	0.46137	0.45676	0.45724
	10000	0.47470	0.46507	0.46672	0.46696	0.45895	0.45802
200	1000	3.34288	3.31321	3.31792	3.29320	3.27194	3.24836
	10000	3.09774	3.33303	3.35658	3.33049	3.28618	3.23154
500	1000	53.2248	48.9947	49.9506	48.2442	45.1045	44.6440
	10000	48.7456	49.1683	49.5907	49.3678	45.7457	48.2131

Table 22: Percentage of exact solutions found by MMACS. The values of q_0 are 0, 0.5, 0.75, 0.9 and 0.99.

N	R	q_1	q_2	q_3	q_4	q_5
50	1000	83.0%	93.0%	84.0%	82.0%	66.0%
	10000	46.0%	61.0%	53.0%	52.0%	39.0%
100	1000	50.0%	86.0%	94.0%	90.0%	77.0%
	10000	11.0%	53.0%	67.0%	63.0%	36.0%
200	1000	15.0%	59.0%	88.0%	92.0%	88.0%
	10000	05.0%	23.0%	70.0%	79.0%	62.0%
500	1000	02.0%	12.0%	56.0%	98.0%	97.0%
	10000	01.0%	06.0%	24.0%	79.0%	85.0%

Table 23: Percentage of perfect solutions found by MMACS. The values of q_0 are 0, 0.5, 0.75, 0.9 and 0.99.

N	R	q_1	q_2	q_3	q_4	q_5
50	1000	52.0%	62.0%	54.0%	52.0%	39.0%
	10000	12.0%	19.0%	16.0%	13.0%	6.0%
100	1000	44.0%	80.0%	87.0%	82.0%	68.0%
	10000	06.0%	36.0%	49.0%	46.0%	22.0%
200	1000	13.0%	56.0%	85.0%	88.0%	82.0%
	10000	05.0%	22.0%	66.0%	72.0%	56.0%
500	1000	01.0%	11.0%	55.0%	96.0%	95.0%
	10000	01.0%	06.0%	24.0%	78.0%	84.0%

Table 24: Averages of solutions found by MMACS. The values of q_0 are 0, 0.5, 0.75, 0.9 and 0.99.

N	R	q_1	q_2	q_3	q_4	q_5
50	1000	0.02926	0.04270	0.07026	0.02934	0.31059
	10000	0.02430	0.01672	0.01380	0.02485	0.02395
100	1000	0.08169	0.03278	0.01605	0.00887	0.01448
	10000	0.04577	0.00236	0.00286	0.00450	0.00636
200	1000	0.16829	0.08070	0.00426	0.00248	0.00327
	10000	0.15506	0.03231	0.00146	0.00087	0.00090
500	1000	0.33937	0.13807	0.06764	0.00059	0.00447
	10000	0.29663	0.12393	0.13807	0.00020	0.00026

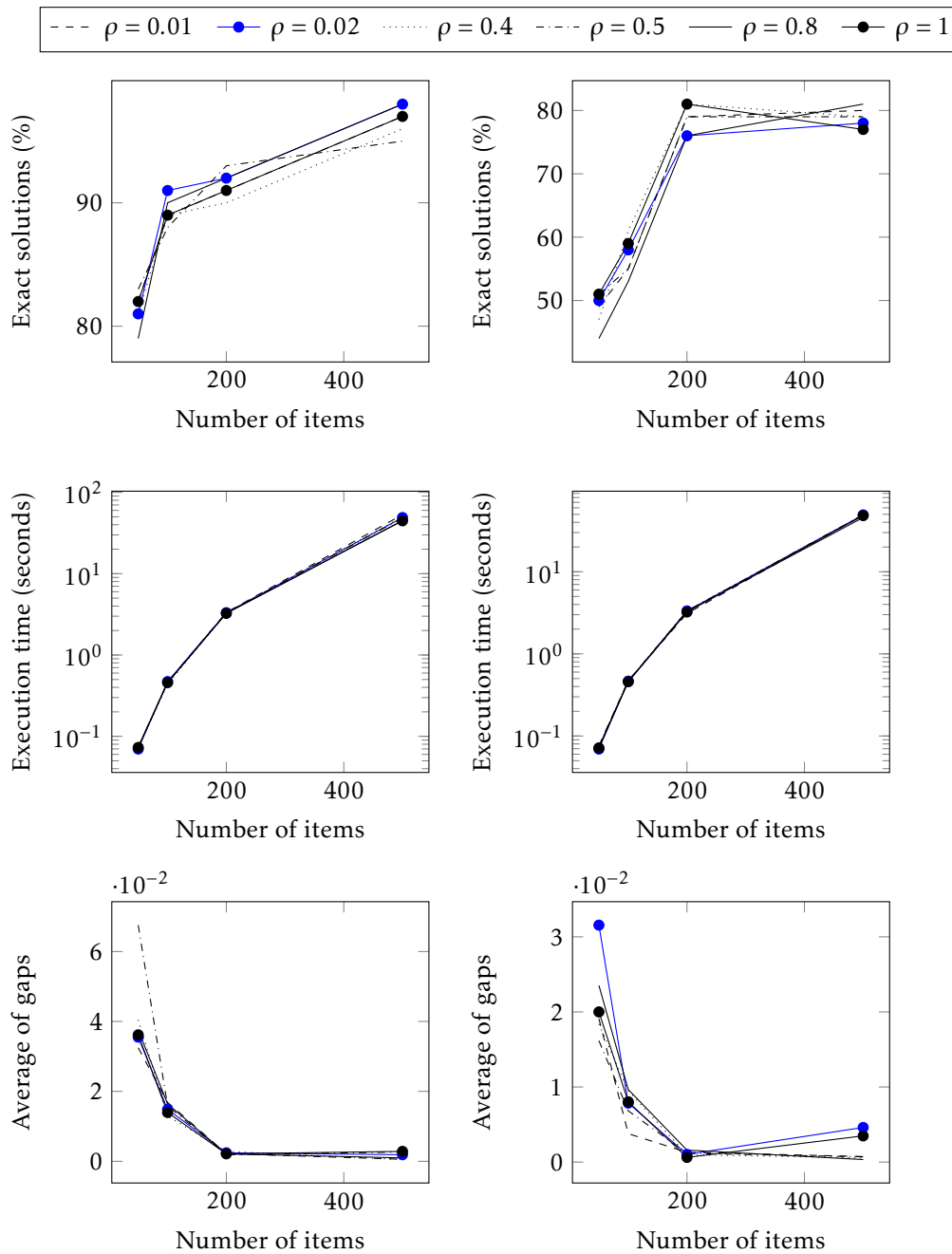


Figure 3: MMACS with a fixed heuristic parameter β_5 and various values of ρ . Plots on the left show results for SCKP instances with a range of coefficients equal to 1000 and plots on the right show results for SCKP instances with a range of coefficients equal to 10000.

Table 25: Execution time of MMACS. The values of q_0 are 0, 0.5, 0.75, 0.9 and 0.99.

N	R	q_1	q_2	q_3	q_4	q_5
50	1000	0.11231	0.06590	0.07087	0.05370	0.17519
	10000	0.14675	0.13272	0.24615	0.11360	0.15682
100	1000	1.11215	0.45967	0.59802	0.36749	1.04415
	10000	1.40685	1.06039	1.16371	0.73181	2.03097
200	1000	11.0831	7.03010	6.07585	1.64040	4.58164
	10000	11.3605	9.8870	11.0704	4.02055	11.2200
500	1000	189.264	162.474	114.984	10.5397	14.0793
	10000	179.640	175.398	162.474	38.1680	47.3076

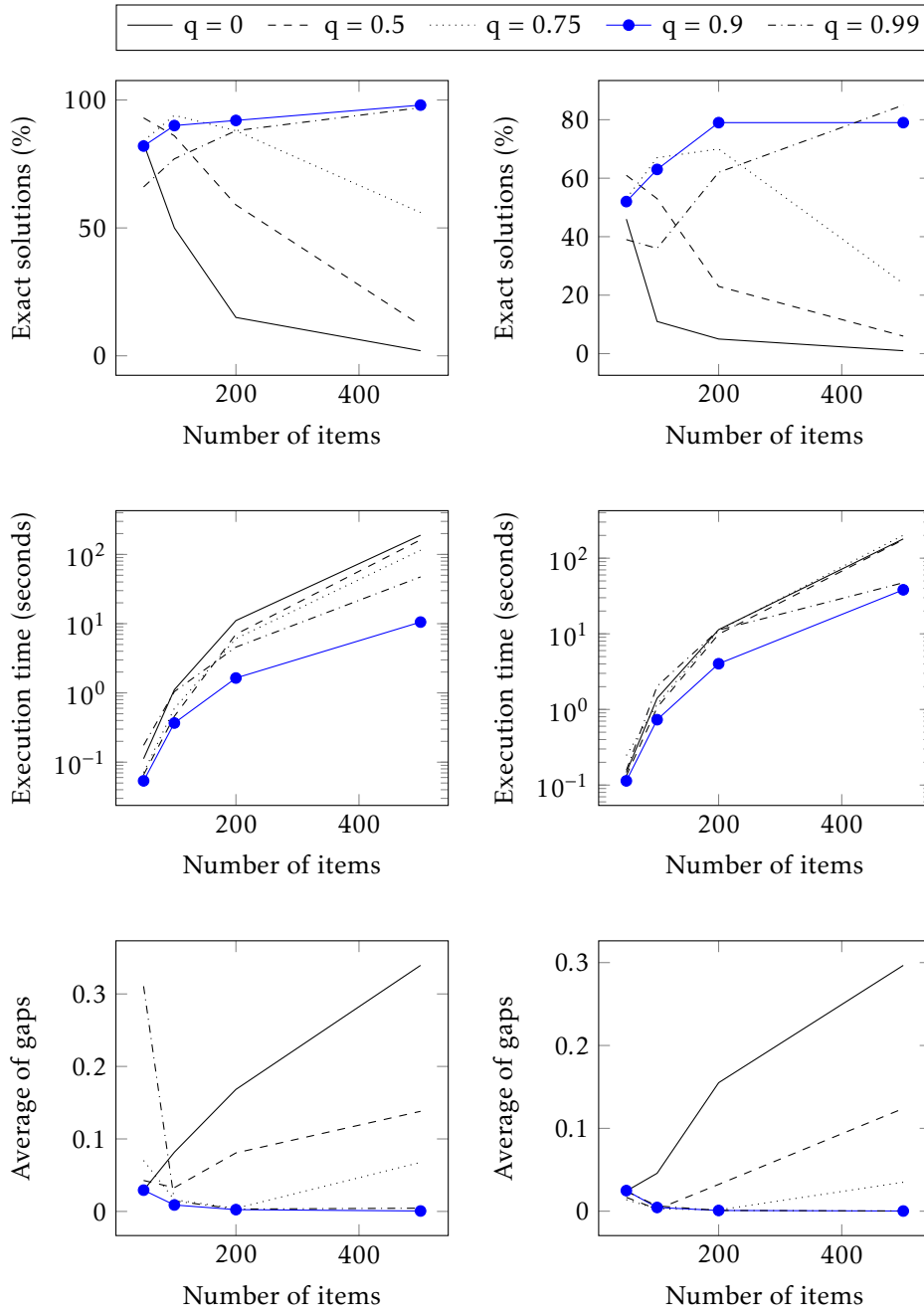


Figure 4: MMACS with various values of q_0 . Plots on the left show results for SCKP instances with a range of coefficients equal to 1000 and plots on the right show results for SCKP instances with a range of coefficients equal to 10000.

Table 26: Percentage of exact solutions found by MMACS. The number of ants varies between 1 and 100.

N	R	$m1$	$m2$	$m3$	$m4$	$m5$	$m6$
50	1000	41.0%	67.0%	77.0%	82.0%	86.0%	89.0%
	10000	25.0%	39.0%	45.0%	52.0%	59.0%	66.0%
100	1000	56.0%	84.0%	84.0%	90.0%	95.0%	93.0%
	10000	20.0%	42.0%	48.0%	63.0%	73.0%	84.0%
200	1000	72.0%	88.0%	90.0%	92.0%	94.0%	93.0%
	10000	18.0%	52.0%	67.0%	79.0%	89.0%	89.0%
500	1000	68.0%	86.0%	93.0%	98.0%	98.0%	98.0%
	10000	19.0%	53.0%	72.0%	79.0%	87.0%	90.0%

Table 27: Percentage of perfect solutions found by MMACS. The number of ants varies between 1 and 100.

N	R	m1	m2	m3	m4	m5	m6
50	1000	18.0%	40.0%	48.0%	52.0%	57.0%	59.0%
	10000	2.0%	5.0%	10.0%	13.0%	20.0%	24.0%
100	1000	48.0%	75.0%	77.0%	82.0%	87.0%	86.0%
	10000	7.0%	29.0%	29.0%	46.0%	54.0%	62.0%
200	1000	63.0%	83.0%	85.0%	88.0%	88.0%	88.0%
	10000	13.0%	47.0%	62.0%	72.0%	82.0%	81.0%
500	1000	66.0%	84.0%	90.0%	96.0%	97.0%	97.0%
	10000	16.0%	52.0%	71.0%	78.0%	86.0%	89.0%

Table 28: Averages of solutions found by MMACS. The number of ants varies between 1 and 100.

N	R	m1	m2	m3	m4	m5	m6
50	1000	0.17092	0.04399	0.05144	0.02934	0.03614	0.04252
	10000	0.06369	0.02088	0.02278	0.02485	0.01487	0.02389
100	1000	0.19670	0.02178	0.01721	0.00887	0.01573	0.01388
	10000	0.00962	0.01011	0.00449	0.00450	0.00639	0.00295
200	1000	0.01496	0.00345	0.00237	0.00248	0.00221	0.00198
	10000	0.00847	0.00139	0.00126	0.00087	0.00110	0.00127
500	1000	0.05480	0.01032	0.01423	0.00059	0.00679	0.00511
	10000	0.01608	0.00523	0.00092	0.00020	0.00028	0.00027

Table 29: Execution time of MMACS. The number of ants varies between 1 and 100.

N	R	m1	m2	m3	m4	m5	m6
50	1000	0.01501	0.03193	0.04690	0.05370	0.13695	0.23765
	10000	0.01918	0.04428	0.10274	0.11360	0.29489	0.53658
100	1000	0.06359	0.16153	0.35330	0.36749	0.60109	1.06442
	10000	0.07816	0.27944	0.50458	0.73181	1.55836	2.23375
200	1000	0.35064	0.81911	1.22310	1.64040	4.42083	8.29943
	10000	0.55900	2.20860	5.59126	4.02055	6.85303	10.5081
500	1000	5.07768	20.2446	20.4123	10.5397	34.7523	65.2062
	10000	8.64695	28.3465	44.5976	38.1680	118.269	153.611

Table 30: Number of problems solved by MMACS, MMAS, MMAS with 2opt, ACS, ACS with 2opt and QEA, in percentage

N	R	MMACS		MMAS		MMAS-2OPT		ACS		ACS-2OPT		QEA	
		BS	PS	BS	PS	BS	PS	BS	PS	BS	PS	BS	PS
50	1000	82.0%	52.0%	75.0%	48.0%	77.0%	47.0%	49.0%	28.0%	60.0%	33.0%	0.0%	0.0%
	10000	52.0%	13.0%	41.0%	5.0%	44.0%	5.0%	31.0%	4.0%	34.0%	5.0%	0.0%	0.0%
100	1000	90.0%	82.0%	38.0%	36.0%	45.0%	40.0%	63.0%	60.0%	69.0%	0.0%	0.0%	0.0%
	10000	63.0%	46.0%	10.0%	5.0%	14.0%	5.0%	28.0%	15.0%	22.0%	9.0%	0.0%	0.0%
200	1000	92.0%	88.0%	12.0%	12.0%	16.0%	13.0%	75.0%	73.0%	74.0%	70.0%	0.0%	0.0%
	10000	79.0%	72.0%	2.0%	2.0%	3.0%	3.0%	36.0%	32.0%	26.0%	21.0%	0.0%	0.0%
500	1000	98.0%	96.0%	2.0%	2.0%	3.0%	1.0%	65.0%	65.0%	69.0%	66.0%	0.0%	0.0%
	10000	79.0%	78.0%	0.0%	0.0%	1.0%	0.0%	29.0%	28.0%	34.0%	33.0%	0.0%	0.0%
1000	1000	100%	96.0%	0.0%	0.0%	2.0%	1.0%	38.0%	38.0%	40.0%	40.0%	0.0%	0.0%
	10000	90.0%	90.0%	0.0%	0.0%	2.0%	2.0%	20.0%	20.0%	20.0%	20.0%	0.0%	0.0%
2000	1000	100%	98.0%	0.0%	0.0%	1.0%	1.0%	40.0%	40.0%	10.0%	10.0%	0.0%	0.0%
	10000	96.0%	94.0%	0.0%	0.0%	0.0%	0.0%	30.0%	30.0%	0.0%	0.0%	0.0%	0.0%

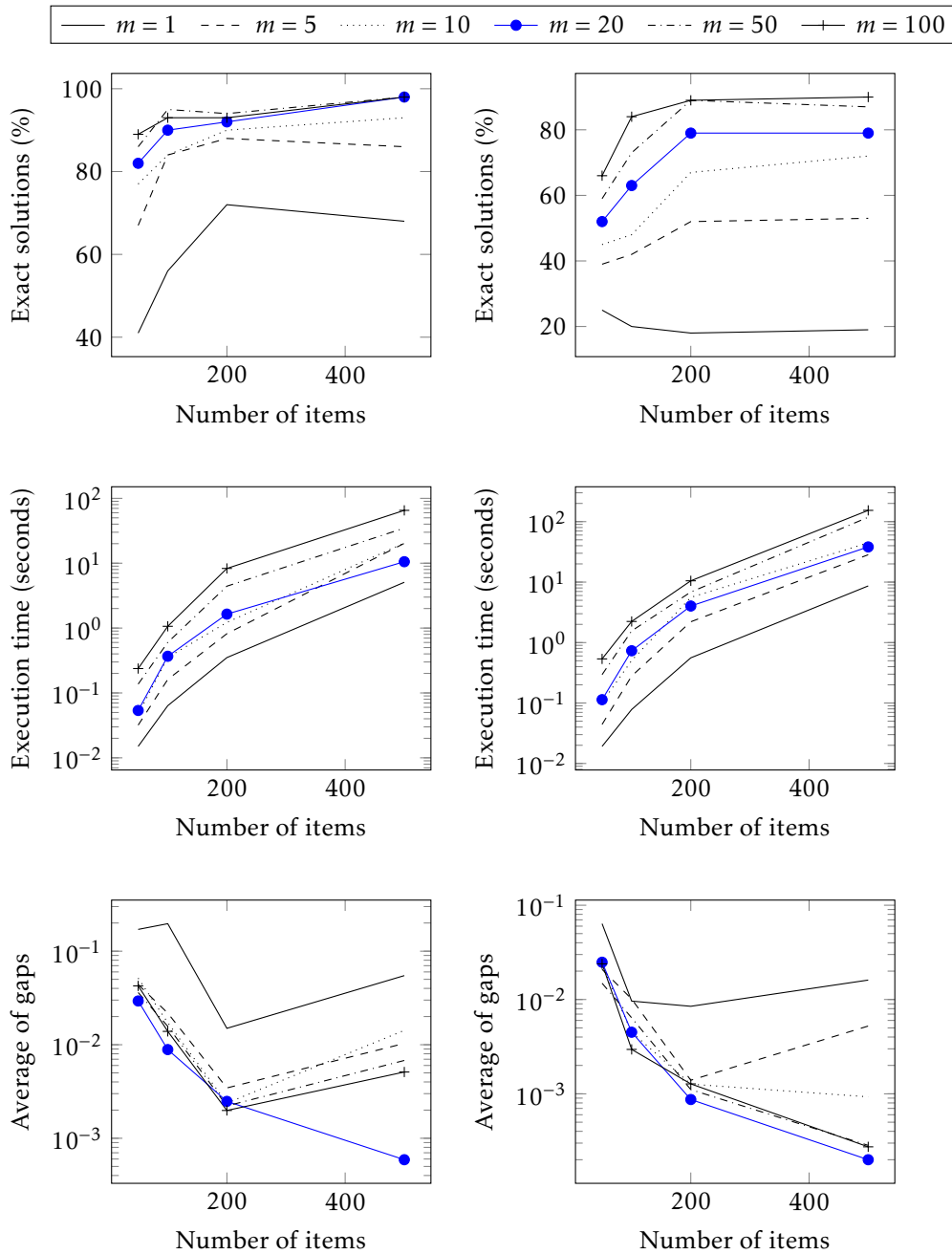


Figure 5: MMACS with various values of the number of ants. Plots on the left show results for SCKP instances with a range of coefficients equal to 1000 and plots on the right show results for SCKP instances with a range of coefficients equal to 10000.

Table 31: Average of GAPS of MMACS, MMAS, MMAS with 2opt, ACS, ACS with 2opt and QEA

N	R	MMACS	MMAS	MMAS-2OPT	ACS	ACS-2OPT	QEA
50	1000	0.02934	0.23080	0.02355	0.35243	0.02780	15.9902
	10000	0.02485	0.09708	0.02444	0.29466	0.03918	19.0041
100	1000	0.00887	0.76940	0.08039	0.13530	0.01213	39.4817
	10000	0.00450	0.50409	0.04409	0.08419	0.00915	39.6268
200	1000	0.00248	0.16937	0.16207	0.20180	0.02418	55.7324
	10000	0.00087	0.15295	0.14329	0.03192	0.00513	50.7264
500	1000	0.00059	0.33796	0.33374	0.04334	0.04135	92.6842
	10000	0.00020	0.30745	0.30680	0.02888	0.02113	91.554
1000	1000	0.00000	0.79903	0.42312	0.04723	0.09261	109.042
	10000	0.00015	0.83977	0.43098	0.10385	0.15489	115.452
2000	1000	0.00000	1.27071	0.49220	0.10819	0.09249	123.848
	10000	0.00001	1.33153	1.37336	0.08955	0.16696	130.323

Table 32: Execution time of MMACS, MMAS, MMAS with 2opt, ACS, ACS with 2opt and QEA, in seconds

N	R	MMACS	MMAS	MMAS-2OPT	ACS	ACS-2OPT	QEA
50	1000	0.05370	0.10212	0.10072	0.11601	0.11575	0.73706
	10000	0.11360	0.25768	0.15341	0.15797	0.14653	0.70629
100	1000	0.36749	1.72373	1.06384	0.68345	0.55707	2.30542
	10000	0.73181	1.37522	1.38279	1.10934	1.06852	2.31311
200	1000	1.64040	19.1191	9.92326	3.65467	3.44395	8.40170
	10000	4.02055	11.0615	10.2807	7.45693	8.76620	8.21929
500	1000	10.5397	189.923	182.438	67.8814	61.3986	46.9157
	10000	38.1618	195.993	164.348	131.969	127.623	46.3203
1000	1000	47.9736	829.651	1284.21	898.149	326.323	174.192
	10000	191.046	867.371	1281.49	730.028	420.990	183.470
2000	1000	340.109	3332.80	1437.79	3648.84	4474.66	722.961
	10000	1318.38	3359.86	3152.59	4253.97	4566.97	781.133

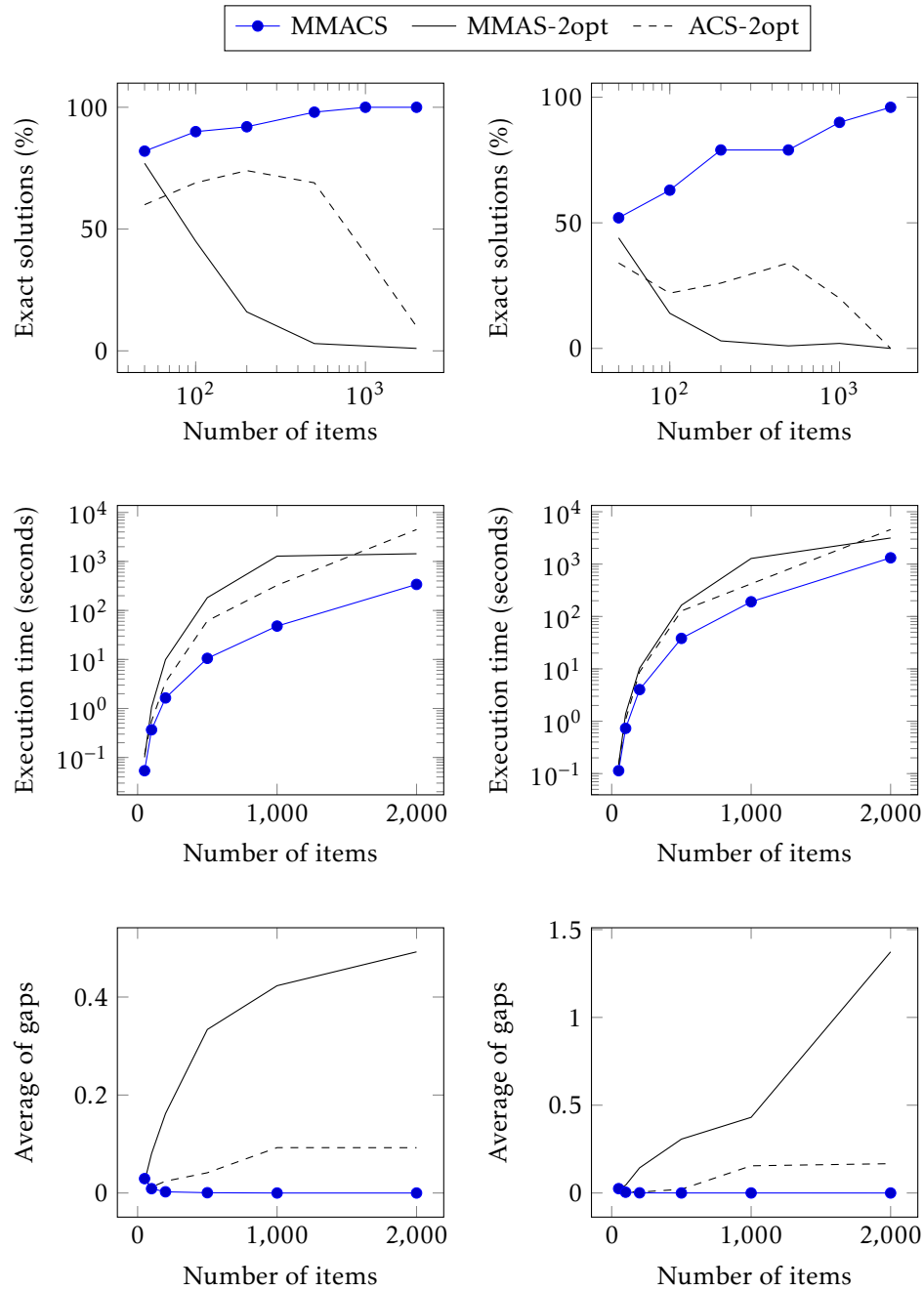


Figure 6: Comparison of MMACS, MMAS with 2-opt and ACS with 2-opt results. Plots on the left show results for SCKP instances with a range of coefficients equal to 1000 and plots on the right show results for SCKP instances with a range of coefficients equal to 10000.

reduce the time. The values 20, 50 and 100 give almost the same percentages. Thus, the value 20 gives the best compromise between a reasonable execution time and acceptable solutions. The solutions are satisfying regarding the considerable percentage of exact solutions and the reduced gap values.

6.3 MMACS experimental settings

We have fixed the parameter values after a set of experimental tests. We have set α to 1, β to 5 and ρ to 0.02, where α and β are the two parameters that determine the relative importance of the pheromone and the heuristic factors and ρ is the evaporation factor. The number of cycles were set to 20 and 20 is the number of ants. As for pheromone trails bounds, we have set τ_{max} to 6 and τ_{min} to 0.01. Finally, the fixed parameter q_0 was set to 0.9.

6.4 MMACS results

After fixing the parameter values of MMACS, empirical results are presented in this section. At a first stage, MMACS results are compared with those of MMAS [9, 10] and ACS [11]. After that, the performances of MMACS while solving SCKP are evaluated and compared to recent methods in the literature: the 2-optimal heuristic [5] and the QEA algorithm [6].

6.4.1 Comparison of MMACS, MMAS and ACS

We test MMACS and the two ACO algorithms: MMAS [9, 10] and ACS [11]. Then, we compare the obtained results. Table 1 gives the default values of the ACO parameters. MMACS, MMAS and ACS were performed on Pisinger instances of size 50, 100, 200, 500, 1000 and 2000. MMAS and ACS were tested with and without the employment of a local search. Results are given in Tables 30- 32. Their performance was compared in terms of percentage of exact and perfect solutions in Table 30, deviation of the best solution found from the optimal solution in Table 31 and the execution time in Table 32. Then, the three ACO algorithms are compared in Figure 6. Results show that for all instances, MMACS algorithm outperform the two ACO algorithms in terms of percentage of exact solutions, execution time and deviation of the best solution found from the optimal solution.

6.4.2 Comparison of MMACS with state of art algorithms

Experiments were conducted on two sets of instances and presented in this section. In the first part of experiments, MMACS, 2-optimal heuristic and QEA solve the instances of the Pisinger set. In the second part, MMACS an QEA were used to solve the instances of the generated set.

Experimental results on the instances of the Pisinger set: We present in this part the results of the experiments realized on the Pisinger set instances of the Strongly Correlated Knapsack Problems. Table 33 shows that in most cases, MMACS turned out to outperform both state of art algorithms. In fact, QEA could not solve these problems to optimality, unlike 2-optimal heuristic which showed better results than MMACS in one case out of four. Besides, our proposed algorithm MMACS reached one hundred percent of solved problem starting with the number of items equal to 1000 and a range of coefficients equal to 1000.

Table 33: Percentage of problems solved by 2-optimal heuristic, MMACS algorithm and QEA algorithm

N	R	2-Optimal	QEA	MMACS
100	1000	68.9%	0%	90%
	10000	13.9%	0%	63%
1000	1000	99.6%	0%	100%
	10000	96.7%	0%	90%
2000	1000	100%	0%	100%
	10000	100%	0%	100%

Experimental results on the instances of the generated set: Additional experiments were conducted on the instances of the generated set. We compare the results of the MMACS algorithm with a state of art algorithm QEA [6]. In QEA, the population size, the maximum number of generations, the global migration period in generation, the local group size and the rotation angle were set to 10, 1000, 100, 2 and 0.01π , respectively.

Both algorithms MMACS and QEA were run under the same computational conditions on instances of the generated set.

In experiments of this part, the SCKP numeric parameters were set to the following values: $R = 10$, $k = 5$ and the number of items are 100, 250 and 500. The generated instances used here are similar to those presented in [6].

The exact solutions of generated instances were obtained using a dynamic programming algorithm [16] that we implemented.

Table 34 shows that MMACS found 30/30 exact solutions for all instances where MMACS BFS (best found solutions) are equal to the optima. Those significant results were given within an acceptable execution time when compared with QEA. The execution time (CPU) and the gap between the found solution and the optimum (Gap) were averaged over 30 runs.

Besides, MMACS and QEA were compared using Wilcoxon Signed Rank Test [17]. This nonparametric test shows that the two groups of data are different according to z-statistic value and p-value at the 0.01 significance level, as shown in Table 35.

Table 34: Comparative results of MMACS algorithm with a state of art algorithm QEA

N	QEA				MMACS			
	BFS	Gap	CPU	BS	BFS	Gap	CPU	BS
100	572	7.43099	1.04186	0%	607	0.00	0.00423	100%
250	1407	11.7935	4.48270	0%	1547	0.00	0.05806	100%
500	2115	20.5593	15.5481	0%	2499	0.00	0.98446	100%

Table 35: Comparative results of MMACS and QEA using Wilcoxon Signed Rank Test

N	z-satistic	p-value
100	-4.7821	0.00000
250	-4.7821	0.00000
500	-4.7821	0.00000

7 Conclusion

The paper presents a comparative study of the proposed hybrid algorithm MMACS while solving the strongly correlated knapsack problems. We gave an experimental analysis of the impact of different parameters on the behavior of MMACS algorithm. Experiments show that the default parameter settings proposed in the literature, gave the best possible results essentially in terms of execution time. It is also noticed from the results that ants in MMACS construct solutions relying mainly on the heuristic information rather than pheromone trails. In fact, initializing pheromone trails to the upper bound helps ants to start the search in promising zones. Besides, the MMACS balances between exploitation and exploration by the employment of a choice rule that alternate between greedy and stochastic approaches. Then, MMACS results were compared to those of MMAS and ACS, the three algorithms show very different behaviors when solving SCKP. The MMACS algorithm outperforms both ant algorithms. In the second part of experiments, we compared MMACS to other recent metaheuristics. Basically, MMACS gave better quality of solutions. As perspective, we propose to draw more attention to the exploitation of the best solutions, in order to avoid early search stagnation. This can achieve the best performances of MMACS.

Conflict of Interest The authors declare no conflict of interest.

References

- [1] W. Zouari, I. Alaya, M. Tagina, "A hybrid ant colony algorithm with a local search for the strongly correlated knapsack problem" in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet Tunisia, 2017. <https://doi.org/10.1109/aiccsa.2017.61>
- [2] D. Pisinger, "Core problems in knapsack algorithms" *Oper. Res.*, 47 (4), 570–575, 1999. <https://doi.org/10.1287/opre.47.4.570>
- [3] T. Cormen, C. Leiserson, R. Rivest, C. Stein, "Greedy algorithms" *Introduction to algorithms*, 1990. <https://doi.org/10.2307/2583667>
- [4] Stützle, Thomas and Ruiz, Rubén, *Iterated greedy*, *Handbook of Heuristics*, Springer, 2018. https://doi.org/10.1007/978-3-319-07124-4_10
- [5] D. Pisinger, "A fast algorithm for strongly correlated knapsack problems" *Discrete Appl. Math.*, 89 (1-3), 197–212, 1998. [https://doi.org/10.1016/s0166-218x\(98\)00127-9](https://doi.org/10.1016/s0166-218x(98)00127-9)
- [6] K.-H. Han, J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization" *IEEE Trans. Evol. Comput.*, 6 (6), 580–593, 2002. <https://doi.org/10.1109/tevc.2002.804320>
- [7] K. O. Jones, "Ant colony optimization, by marco dorigo and thomas stützle" *ROBOTICA*, 2005. <https://doi.org/10.1017/s0269888905220386>
- [8] Dorigo, Marco and Stützle, Thomas, *Ant colony optimization: overview and recent advances*, *Handbook of metaheuristics*, Springer, 2019. https://doi.org/10.1007/978-3-319-91086-4_10
- [9] T. Stützle, H. Hoos, "Max-min ant system and local search for the traveling salesman problem" in *IEEE International Conference on Evolutionary Computation*, Indianapolis USA, 1997. <https://doi.org/10.1109/icec.1997.592327>
- [10] T. Stützle, H. H. Hoos, "Max-min ant system" *Future Gener. Comput. Syst.*, 16 (8) 889–914, 2000. [https://doi.org/10.1016/s0167-739x\(00\)00043-1](https://doi.org/10.1016/s0167-739x(00)00043-1)
- [11] M. Dorigo, L. M. Gambardella, "Ant colony system: a cooperative learning approach to the traveling salesman problem" *IEEE Trans. Evol. Comput.*, 1 (1), 53–66, 1997. <https://doi.org/10.1109/4235.585892>
- [12] I. Alaya, C. Solnon, K. Ghedira, "Ant colony optimization for multi-objective optimization problems" in *19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007)*, Patras Greece, 2007. <https://doi.org/10.1109/ictai.2007.108>
- [13] G. A. Croes, "A method for solving traveling-salesman problems" *Oper. Res.*, 6 (6), 791–812, 1958. <https://doi.org/10.1287/opre.6.6.791>
- [14] T. Stützle, M. López-Ibáñez, P. Pellegrini, M. Maur, M. M. De Oca, M. Birattari, M. Dorigo, *Parameter adaptation in ant colony optimization*, *Autonomous search*, Springer, 2011. https://doi.org/10.1007/978-3-642-21434-9_8
- [15] R. W. Morrison, K. A. De Jong, "Measurement of population diversity" in *International Conference on Artificial Evolution (Evolution Artificielle)*, Berlin Heidelberg, 2001. https://doi.org/10.1007/3-540-46033-0_3
- [16] P. Toth, "Dynamic programming algorithms for the zero-one knapsack problem" *Computing*, 25 (1), 29–45, 1980. <https://doi.org/10.1007/bf02243880>
- [17] F. Wilcoxon, S. Katti, R. A. Wilcox, *Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test*, *Selected tables in mathematical statistics 1*, 1970.

Low-Dimensional Spaces for Relating Sensor Signals with Internal Data Structure in a Propulsion System

Catherine Cheung^{*1}, Nicolle Kilfoyle², Julio Valdés³, Srishti Sehgal¹, Richard Salas Chavez¹

¹National Research Council Canada, Aerospace, Ottawa, Canada

²Department of National Defence, Ottawa, Canada

³National Research Council Canada, Digital Technologies, Ottawa, Canada

ARTICLE INFO

Article history:

Received: 17 August, 2018

Accepted: 28 October, 2018

Online: 01 November, 2018

Keywords:

Low-dimensional spaces

Condition indicators

Failure prediction

Intrinsic dimension

ABSTRACT

Advances in technology have enabled the installation of an increasing number of sensors in various mechanical systems allowing for more detailed equipment health monitoring capabilities. It is hoped the sensor data will enable development of predictive tools to prevent system failures. This work describes continued analysis of sensor data surrounding a seizure of a turbocharger within a propulsion system. The objective of the analysis was to characterize and distinguish healthy and failed states of the turbocharger. The analysis approach included mapping of multi-dimensional sensor data to a low-dimensional space using various linear and nonlinear techniques in order to highlight and visualize the underlying structure of the information. To provide some physical insight into the structure of the low-dimensional representation, the transformation plots were analyzed from the perspective of several engine signals. By overlaying operating ranges of individual sensor signals, certain regions of the mappings could be associated with distinct operational states of the engine, and several anomalies could be related to various points in the turbocharger seizure. Although the failed points did not map to an obvious outlier location in the transformations, incorporating expert domain knowledge with the data mining tools significantly enhanced the insight derived from the sensor data.

1. Introduction

Rapid developments in sensor technology, data processing tools and data storage capability have helped fuel an increased appetite for equipment health monitoring in mechanical systems. As a result, the number of sensors and amount of data collected for health monitoring has grown tremendously. It is hoped that by collecting large quantities of operational data, predictive tools can be developed that will provide operational, maintenance and safety benefits. Data mining and machine learning techniques are important tools in addressing the ensuing challenge of extracting useful results from the data collected. However, incorporating as much physical domain knowledge to the analysis as possible is also necessary to ensure the results are relevant and practical for the operator and end-user.

This work describes continued analysis of sensor data for the turbocharger subsystem of a diesel engine system. The engine has

hundreds of sensors monitoring both the inputs of the engine operator and the resulting equipment outputs. A turbocharger seizure was recorded by the diesel engine sensor system. Therefore, data analysis of this incident including the lead up to the event allows for monitoring and identification of changes in equipment condition indicators with a known outcome.

This paper is an extension of work originally presented at the 2017 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS) [1]. The initial data analysis of this event included intrinsic dimension analysis and relied on clustering techniques for data reduction to transform the high-dimensional sensor data to a low-dimensional space. In this extended version of the paper, the data has not been reduced using clustering techniques in order to minimize information loss and t-Distributed Stochastic Neighbour Embedding (t-SNE) is included as an additional mapping method. An important addition in this paper is further analysis to relate individual sensor signals to the internal data structure of the low-dimensional mappings.

*Catherine Cheung, 1200 Montreal Rd, Bldg M-14, Ottawa, ON, Canada K1A 0R6, +1-613-998-1541, Email: catherine.cheung@nrc-cnrc.gc.ca

Failure detection in mechanical systems using data-driven models is an area that has been the focus of much published research in the last decade or so. Machinery failures are hard to predict due to the complex nature of the structure and functions of the system. Using data driven models helps reduce maintenance costs, improve productivity and increase machine availability [2]. For example, fuzzy support vector machines have been applied to identify faults in induction motors [3]. In another case, various data-driven models such as support vector machines, decision trees and kernel methods were used and compared to diagnose faults in shaft and bearings of rotating machinery [4]. Additionally, a model was developed to continuously monitor the health of wind turbine gearboxes [5]. In brief, data-driven models demonstrate great potential for failure detection and preventive maintenance in mechanical applications.

Other related work revolves around understanding the trends data-driven models produce. In several fields, data is being acquired at an astounding rate [6]. There is a significant need for the development of methods and techniques to obtain useful knowledge from these growing sets of available data. Currently, there exist few processes that combine expert knowledge of a system with algorithm-generated prediction models of the system's datasets. However, in cases where expert knowledge is combined with machine learning techniques, there is a notable improvement in the results. For example, a methodology was proposed to detect web attacks [7]. In this procedure, features that represented the tendencies of the system were chosen by experts and were combined with output provided by *n-grams*, a feature extraction algorithm. In another instance, similar to mechanical applications, a framework to combine the clinical intuitions and experience of medical experts with machine learning models was used to overcome the lack of ideal training sets [8]. Medically trained experts provided a task to accomplish, the desired outcome, the data, and helped construct relationships between variables with the algorithm designers. The relationships that were provided aligned with the intuition of the medical expert and their understanding of how factors played out in predicting a certain outcome [8]. However, when expert-knowledge is not always readily available, or if all the variables cannot be identified for a particular outcome, or when proper variable relationships cannot be constructed for a particular outcome, the need to be able to obtain useful knowledge from the results of data-driven models still exists.

In this work, a multi-disciplinary approach to gain knowledge from high-dimensional and voluminous datasets generated by complex real-world systems is explored. Data mining and machine learning techniques are implemented to gather useful insights from the large amounts of sensor data collected for this diesel engine system. By incorporating expert domain knowledge with the low-dimensional representations of the data, a more practical understanding of the data structure presented in the mappings is provided which helps ensure that the results are relevant and accessible to the operator and end-user.

This paper is organized as follows: description of the turbocharger sensor data and the turbocharger seizure are provided in Section 2; details of the implemented data analysis tools and techniques are given in Section 3; the data pre-processing steps and the experimental settings are outlined in Section 4; the intrinsic

dimension analysis and low-dimensional mappings are included in Section 5; Section 6 provides the results of several sensor signals overlaid on the visualizations; and finally, Section 7 summarizes the findings of the paper.

2. Turbocharger data

The analyzed turbocharger system contains twin air-cooled turbochargers providing the air-charge to the medium-speed diesel engine system. The diesel engine system consists of two banks of 10-cylinders, denoted Bank A and Bank B. A single turbocharger is assigned to each 10-cylinder bank; the two turbochargers are differentiated as Turbo A and Turbo B, where the letter 'A' or 'B' identifies their respective cylinder bank.

The incident recorded by the engines sensor system and analyzed within this work relates to the seizure of Turbo A [1]. The sensor data recorded for this particular incident was available for the month leading up to and including the time of the incident. From the resulting investigation of the Turbo A seizure, a series of key events leading to the incident were noted. The chronology of the incident's events is detailed below, with the time of occurrence indicated as (hh:mm).

- Noted loss of sensor reading on Turbo A speed sensor
- Engine shut down for inspection
- Turbo A and B speed sensors switched
- Engine restarted at idle, still no Turbo A speed reading indicated (01:12)
- Engaged diesel engine (01:41)
- Higher speed setting requested (01:42), engine exhaust temperatures increased beyond alarm threshold, with no speed increase achieved (01:43 - 01:44)
- Diesel engine disengaged and shut down (01:44 - 01:45)

Following the incident, an inspection of Turbo A was conducted. From the inspection it was determined that the seizure of Turbo A occurred due to a sensor installation error, which occurred when the speed sensors for Turbo A and B were switched. Insufficient spacing between the speed sensor and the turbine's thrust collar led to rubbing and eventually the turbocharger seizure [1].

Although this failure was caused by installation error rather than gradual deterioration of a system element, the ability to characterize and distinguish the healthy state from the seized state of the turbocharger system using data analysis tools is of significant value. This type of analysis could aid in establishing failure models for further predictive work.

2.1. Turbocharger sensor signals

The sensor system of the diesel engine is comprised of 238 sensors that capture information related to operator inputs, equipment outputs, and sensor data. The sensor system provides a means for staff to control system components, monitor the systems status, or be notified via alarm when various pieces of equipment operate outside of pre-set threshold values. In addition, the sensor system allows for the recording and archiving of the operational data measured from various instruments at rates up to 2 Hz. From the 238 sensors relevant to the diesel engine, a subset of 31 signals relating to the operation of the turbochargers was selected. The 31

sensors considered within this analysis, listed in Table 1 [1], encompass parameters such as speeds, temperatures (inlet, outlet, and exhaust), pressures, and shaft torque.

Table 1: 31 turbocharger input parameters

Signal #	Signal
1	Average cylinder exhaust temperature Bank A
2	Average cylinder exhaust temperature Bank B
3	Turbo A speed
4	Turbo B speed
5	Turbo B inlet temperature
6	Turbo B outlet temperature
7	Turbo A inlet temperature
8	Turbo A outlet temperature
9	Charge air manifold pressure
10	A1 cylinder exhaust gas temperature
11	B1 cylinder exhaust gas temperature
12	A2 cylinder exhaust gas temperature
13	B2 cylinder exhaust gas temperature
14	A3 cylinder exhaust gas temperature
15	B3 cylinder exhaust gas temperature
16	A4 cylinder exhaust gas temperature
17	B4 cylinder exhaust gas temperature
18	A5 cylinder exhaust gas temperature
19	B5 cylinder exhaust gas temperature
20	A6 cylinder exhaust gas temperature
21	B6 cylinder exhaust gas temperature
22	A7 cylinder exhaust gas temperature
23	B7 cylinder exhaust gas temperature
24	A8 cylinder exhaust gas temperature
25	B8 cylinder exhaust gas temperature
26	A9 cylinder exhaust gas temperature
27	B9 cylinder exhaust gas temperature
28	A10 cylinder exhaust gas temperature
29	B10 cylinder exhaust gas temperature
30	Shaft torque position 1
31	Shaft torque position 2

The data from the month leading up to and including the turbocharger seizure was analyzed. With the data down-sampled to 1-minute intervals, there were 9968 data points during the time period. The data points prior to the turbocharger seizure were designated as ‘healthy’, while the points from midnight of the date of the incident were considered ‘failed’ points. These failed points include all of the data points after the seizure as they correspond to data related to the seized subsystem. As a result, there were 9875 healthy points and 93 failed points identified.

3. Analytical techniques

An important aspect of this work was the characterization of the relationship between the healthy and failed states of the turbocharger system, particularly the transition between the two states. The original sensor data is described by a multidimensional time-series composed of the 31 signals. Typical from these kind of data is the presence of noise, irrelevancies and redundancies between the descriptor variables, as in reality the core of the data represents a subspace of lower dimension embedded within the higher dimensional descriptor space. In such situations, transformations to lower dimensional spaces are useful for

highlighting and visualizing the underlying structure of the information.

To that end, a suitable transformation, preferably with an intuitive metric [9] would produce a mapping of the original high dimensional objects into a lower dimension one, such that a certain property of the data is preserved by the representation. Desired properties characterizing data structure could be conditional probability distributions around neighbourhoods, similarity relations and others. Under normal circumstances, such transformations imply some information loss or error that should be minimized. If successful, the transformation would generate a new set of features out of the original ones which would preserve the desired property, but in a lower dimension representation space that mitigates the curse of dimensionality.

3.1. Intrinsic dimensionality analysis

When choosing the dimension of the target space, it is important to consider the intrinsic dimensionality of the original information which is typically understood as the minimum number of variables required to account for the observed properties of the data [10-13]. Given a functional measure of information loss, it is the minimum number of dimensions (descriptor variables) required to describe the data that minimizes that measure. This concept could be understood in several ways, which results in different algorithms aiming at producing such an estimation. Some approaches focus on local properties of the data, whereas other techniques emphasize the analysis of global properties of the data. Most complex systems in the real world exhibit nonlinear relations among their state variables, which make linear estimators of intrinsic dimensionality at a global scale less powerful than their nonlinear estimation counterparts. However, some of them are computationally expensive.

From the practical point of view, the smaller the choice of the target dimension with respect to the intrinsic one, the higher the representation error would be. On the other hand, choosing values higher than the chosen dimension introduce unnecessary attributes, redundancies and possibly noise. For visualization purposes, three or two dimensions are forcibly required. In these cases, the value of the intrinsic dimension provides a useful guide to the level of caution required when making inferences based on the visualization space.

In this work, the intrinsic dimension of the turbocharger data is estimated using four nonlinear methods and one linear technique: maximum likelihood estimation (MLE), correlation dimension (CD), geodesic minimum spanning tree (GMST), nearest neighbour estimator, and principal component analysis (PCA).

The maximum likelihood estimator [14] is based on the assumption of a Poisson distribution for the k neighbour points and a constant behavior of the probability density function around a given point. The actual estimate of the dimension is derived from the log-likelihood function.

Correlation dimension [15] is one type of fractal dimension and it is one of the most commonly used techniques for estimating the intrinsic dimension. The idea is to compare objects from the point of view of their pairwise distances, producing a normalized count of those pairs whose distance does not exceed a given threshold (the correlation integral). The estimate is given by the slope of a

log-log linear regression of the correlation integral values vs. the different distance thresholds r .

The geodesic minimum spanning tree (GMST) estimator [16] assumes that i) the set of multivariate objects are in a smooth manifold embedded within the higher dimensional space determined by the original descriptor variables and ii) these objects are realizations of a random process from an unknown multivariate probability density distribution. This technique produces an asymptotically consistent estimate of the manifold dimension without requiring the reconstruction of the manifold or the estimation of the multivariate distribution of the objects. The first step is to construct a graph based on k -neighbourhood density (or neighbourhood distances) where every object is connected with the others nearby. The second step is to build a minimal spanning tree (MST). The distances along its edges and the overall length are used to estimate parameters of the manifold, like entropy and dimension.

The nearest neighbour estimator [17] presents some similarities with the correlation dimension. It is motivated by the possibility of approximating the unknown probability density of the set of multidimensional objects, by normalizing the relative number of nearest neighbours by the volume of the hypersphere containing the objects. The procedure computes the smallest radius r required to cover k nearest neighbours via a linear log-log regression of the average minimum radius vs. k .

Principal component analysis (PCA) is an unsupervised, classical method that is among others, it is used to estimate intrinsic dimensionality. The estimation is simply constructed by obtaining the number of eigenvalues whose relative contributions to the overall variance exceeds a predetermined threshold (e.g. 0.975). Singular value decomposition techniques or diagonalization of covariance/correlation matrices are the typical approaches used for finding the components, which are linear combination of the original set of features. The former approach was used in this paper following the algorithm described in [18].

3.2. Transformation from high- to low-dimensional space

The specificities of the data determine its intrinsic dimension and, in particular, when the estimates are not too different from three, mappings targeting that number of dimensions could portray appropriate representations of the data. In these cases, the visualizations obtained with different mapping techniques typically exhibit low errors or information loss measures. They would reveal patterns corresponding to valid relationships within the data like regularities, showing up as clustering structure, as well as abnormalities, less frequent elements and outliers.

Clearly, for machine learning purposes, mitigating the curse of dimensionality is important and often the dimension of the target spaces must go beyond the ones required by visual inspection. In this paper, Principal Components, Sammon mapping, t-SNE and Isomap were used as representatives of linear and nonlinear transformation techniques.

Principal Components

A low-dimensional representation of the data can be produced using the first few principal components found through principal component analysis, which are mutually orthogonal (described in

Section 3.1). Their main features were presented in the previous section. From the point of view of visually inspecting the data, the first few principal components (up to three) are used as a baseline low-dimensional representation. They are linearly uncorrelated and the amount of variance contained in each new component decreases monotonically. However, the cumulated variance contained in the first three components is not sufficiently high and it is a crucial element to consider when working with principal components visualizations.

Sammon Mapping

The idea of constructing low dimensional spaces where the distance distribution maximally matches the one in the original space is very intuitive and has been at the core of multi-dimensional scaling methods (MDS) [19-22]. Different variants of this approach have been used for creating visual representations of metric and non-metric data. On representations that aim at preserving distances in the original and the target spaces, nearby/distant objects in the original data space are placed at nearby/distant locations from each other in the low-dimensional target space. Some variants preserve the actual distance values, while others aim at preserving their ranks or their ordering relation.

In the first case, measures based on squared differences between dissimilarities on both spaces are commonplace and are variations of objective functions like

$$\sum_{1 \leq i, j \leq N} w_{ij} (F(\delta_{ij}^p) - d_{ij}^p)^2$$

where N is the number of objects, w_{ij} is a weight associated to every pair of objects i, j , F is a monotonically increasing function, δ_{ij} is a dissimilarity measure between objects i, j in the original data space and d_{ij} is their dissimilarity/distance in the target space, with p as an exponent of the difference term.

From this general formulation, several mapping techniques are derived, in particular, Sammon's nonlinear mapping [23], conceived as a transformation of vectors of two spaces of different dimension ($D > m$) by means of a function $\varphi: R^D \rightarrow R^m$, which maps vectors $\vec{x} \in R^D$ to vectors $\vec{y} \in R^m$, $\vec{y} = \varphi(\vec{x})$. The actual objective function to minimize is given by Equation 1:

$$Sammon\ error = \frac{1}{\sum_{i < j} \delta_{ij}} \sum_{i < j} \frac{(\delta_{ij} - d(\vec{y}_i, \vec{y}_j))^2}{\delta_{ij}}, \quad (1)$$

where typically d is an Euclidean distance in R^m . The weight term δ_{ij}^{-1} highlights the importance of smaller distances and therefore, the behavior around close neighbourhoods exerts a larger influence on the error function.

t-SNE

A probabilistic principle is used by the Stochastic Neighbour Embedding (SNE) [24], where the goal is to preserve neighbour identities. A dissimilarity or distance between the objects in the original space is used for creating an asymmetric probability for each object with respect to its potential neighbours, with a pre-set neighbourhood notion (the perplexity). The same is performed for the objects in the target space.

The goal is to match the two distributions as much as possible, which is achieved by minimizing the sum of Kullback-Leibler divergences. The rationale is to center a Gaussian on each object in the original space and to use the distances (or given similarities) for constructing a local probability density function on the neighbourhood. The same operation is performed in the transformed, low dimensional space and the purpose is to match the two as much as possible, so that neighbourhoods are preserved.

The t-Distributed Stochastic Neighbour Embedding (t-SNE) [24-26] is an improvement of the original SNE. There are two main distinguishing features of t-SNE: *i*) a simpler symmetric objective function is introduced; and *ii*) instead of a Gaussian distribution, a t-Student distribution is used for the points in the low-dimensional space. These modifications allow for better capturing the local structure of the high-dimensional data and also revealing the presence of clusters at several scales, as indicators of global structure.

Isomap

Isomap [27-30] is a flexible technique oriented to learn non-linear manifolds and overcomes some difficulties inherent to classical methods like principal components or MDS-related. In contrast to the latter, the Isomap technique aims to preserve pairwise geodesic (or curvilinear) distances rather than plain (Euclidean) ones. Geodesic distances are those measured along the low-dimensional manifold containing the data and therefore, not necessarily objects that are close in the Euclidean sense will be so when geodesic distances are considered.

There are three steps in the procedure: *i*) build a graph (the neighbourhood graph) that connects all points according to their pair-wise Euclidean distances; *ii*) estimate the geodesic distances between all pairs of points by calculating their shortest path distances in the neighbourhood graph; and *iii*) compute a geodesic distance preserving mapping using MDS with Euclidean distance as metric for the low-dimensional space.

4. Experimental settings

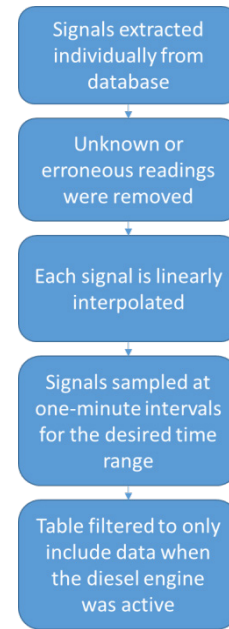
For the investigated time period leading up to and including the turbocharger seizure, there were 9968 data points. These points were sampled at intervals of one-minute. Of the 9968 total points, 9875 were designated as ‘healthy’ and the remaining 93 designated ‘failed’.

To determine the low-dimensional transformation using the Isomap method, 12 nearest neighbours were specified. For the t-SNE transformation, the perplexity was set at 30.

4.1. Data pre-processing

The engine sensor system was originally set up for real-time equipment health monitoring, and not specifically for maintenance or safety purposes. As such, a number of data pre-processing and data consolidation steps were necessary before implementing the data analysis tools. The data pre-processing steps followed are illustrated in Figure 1.

From the full signal database, each of the 31 signals was extracted separately, then unknown and erroneous readings were removed. Afterwards, each signal was linearly interpolated and



sampled at one-minute intervals for the desired time range, ensuring that the time range fell within the interpolated values. Finally, a filter was applied to only consider data recorded during active operation of the diesel engine.

The input parameters were standardized in order to fairly compare variables measured in different units and different ranges. The standardized variables had a mean value of zero and standard deviation equal to one.

5. Low-dimensional mappings

5.1. Intrinsic dimension results

Estimates of the intrinsic dimension of the turbocharger data were calculated using the five techniques detailed in Section 3.1. These estimates are listed in Table 2. The first three principal components represent 0.981 of the total variance in the data.

Table 2: Intrinsic dimension estimates

Estimation Method	Estimate
Maximum Likelihood Estimator	5.239
Correlation Dimension	1.285
Geodesic Minimum Spanning Tree	4.202
Nearest Neighbour Dimension	0.307
Principal Component Analysis Eigenvalues	3.000

Almost all of the estimates fall in the range of 1 to 5. Since the original sensor space corresponds to a dimension of 31, these estimates show that the information contained in that 31-D high dimensional space could be sufficiently explained more simply by the combination of a few factors. A target dimension of 3 was selected, taking into account the range of estimates.

5.2. Transformation to 3-dimensions

From the intrinsic dimension results, low-dimensional representations of the turbocharger data with three dimensions

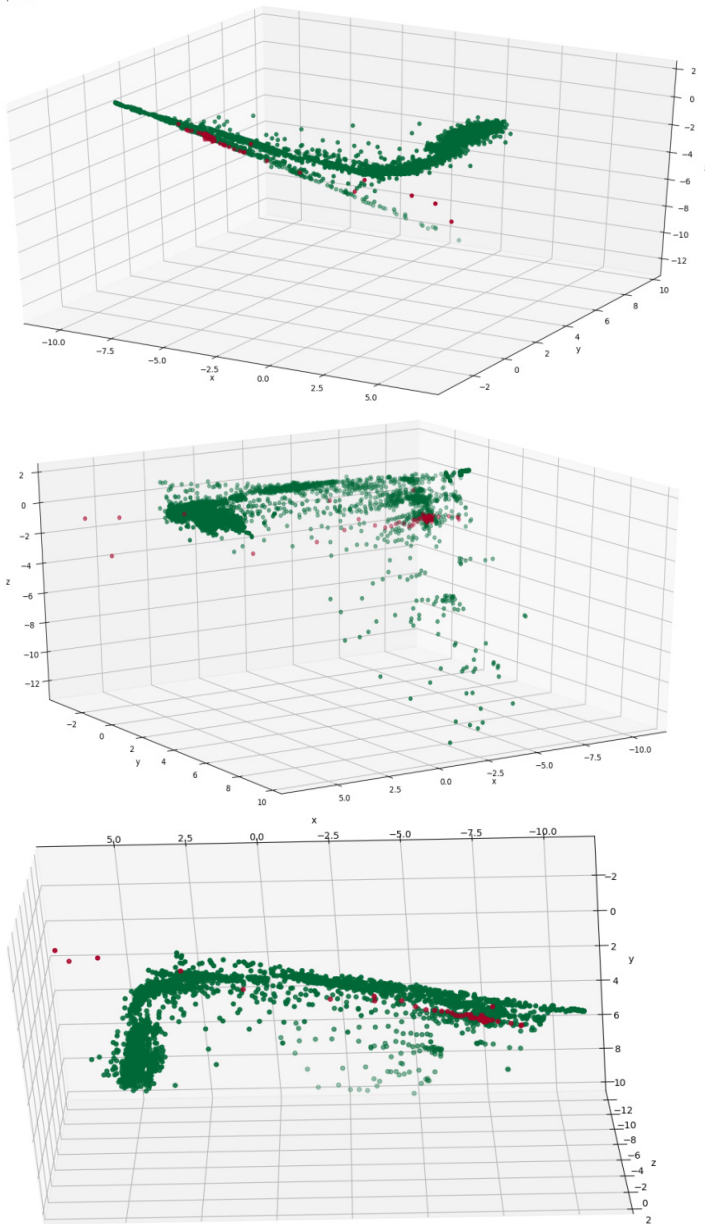


Figure 2: PCA mappings to 3-D

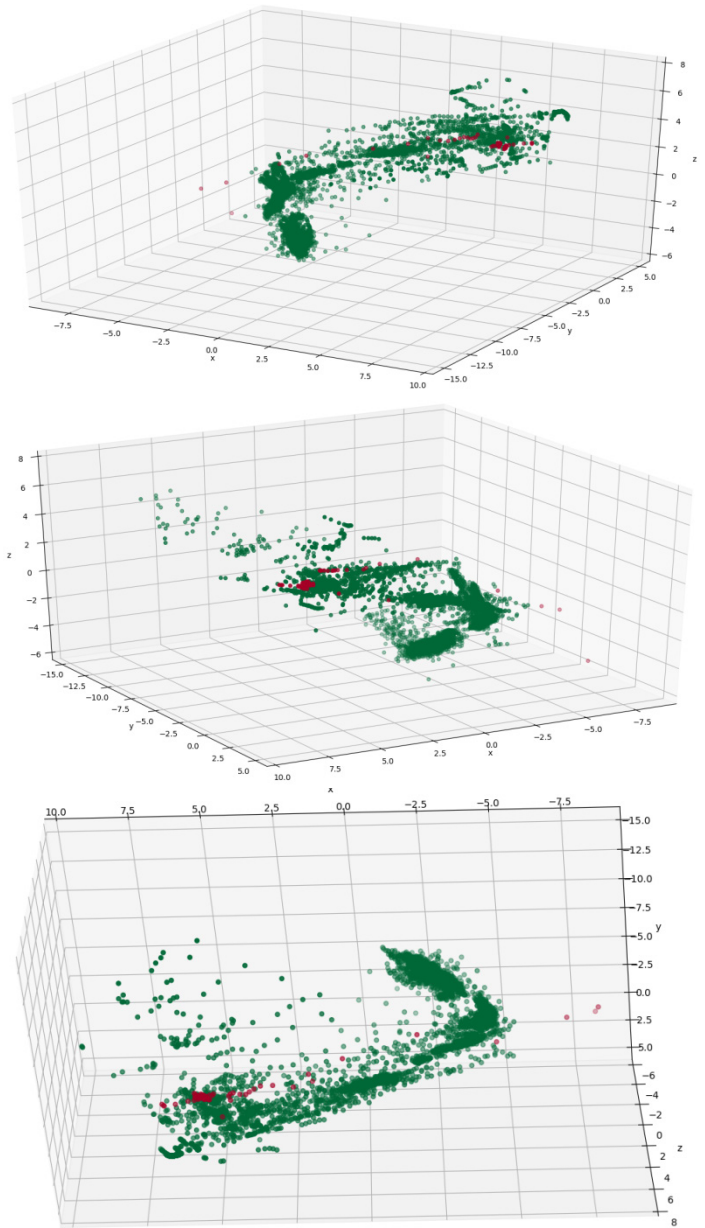


Figure 3: Sammon mappings to 3-D

were sought. The transformation of the original 31-D space, corresponding to the 31 turbocharger sensors, to 3-dimensions was performed using the four methods described in Section 3.2 (principal component analysis, Sammon, t-SNE, and Isomap). The full data set from the time period around the turbocharger event was mapped, consisting of 9968 total points of which 93 were designated ‘failed’ and the rest ‘healthy’.

Images of the 3-D spaces obtained from the different mappings are presented in Figures 2-5. In these images, the healthy points are coloured green, while the failed points are red. Since the ratio of healthy to failed points is tremendously imbalanced (9875:93), it may be difficult to see the failed points. Representing 3-D scenes by 2-D images is clearly not ideal because of the limitations in exploring different perspectives and distances between objects in the scene. Several snapshots of the 3-D space are included to help overcome that limitation.

Figure 2 shows three views of the low-dimensional mapping using the first three principal components determined through PCA. These three components are orthogonal and are linear combinations of the 31 turbocharger variables. Figure 3 depicts three views of the Sammon mapping to 3-dimensional space. Figure 4 shows the t-SNE mapping to 3-D. Figure 5 illustrates the Isomap transformation to 3-dimensions.

In the PCA mapping (Figure 2), the data is distributed mostly in a 2-dimensional plane in a boomerang-like shape. The failed points are concentrated at one end of that boomerang shape ($-5 \leq x \leq 10$). The Sammon mapping (Figure 3) also shows the distribution of the data in a boomerang-like shape. Again, the failed points are clustered at one end of that shape ($3 \leq x \leq 7$).

The t-SNE transformation (Figure 4) has a markedly different data structure with the data organized in more distinct clusters as opposed to the continuous distribution of data in a particular shape

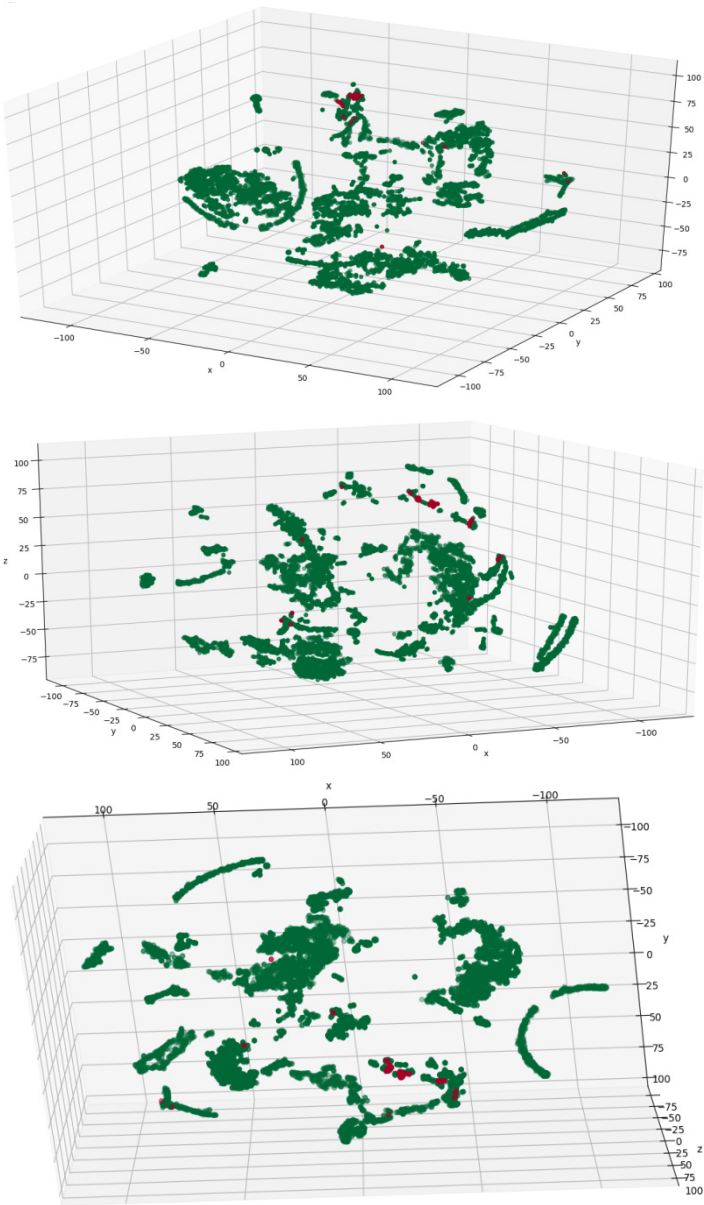


Figure 4: t-SNE mappings to 3-D

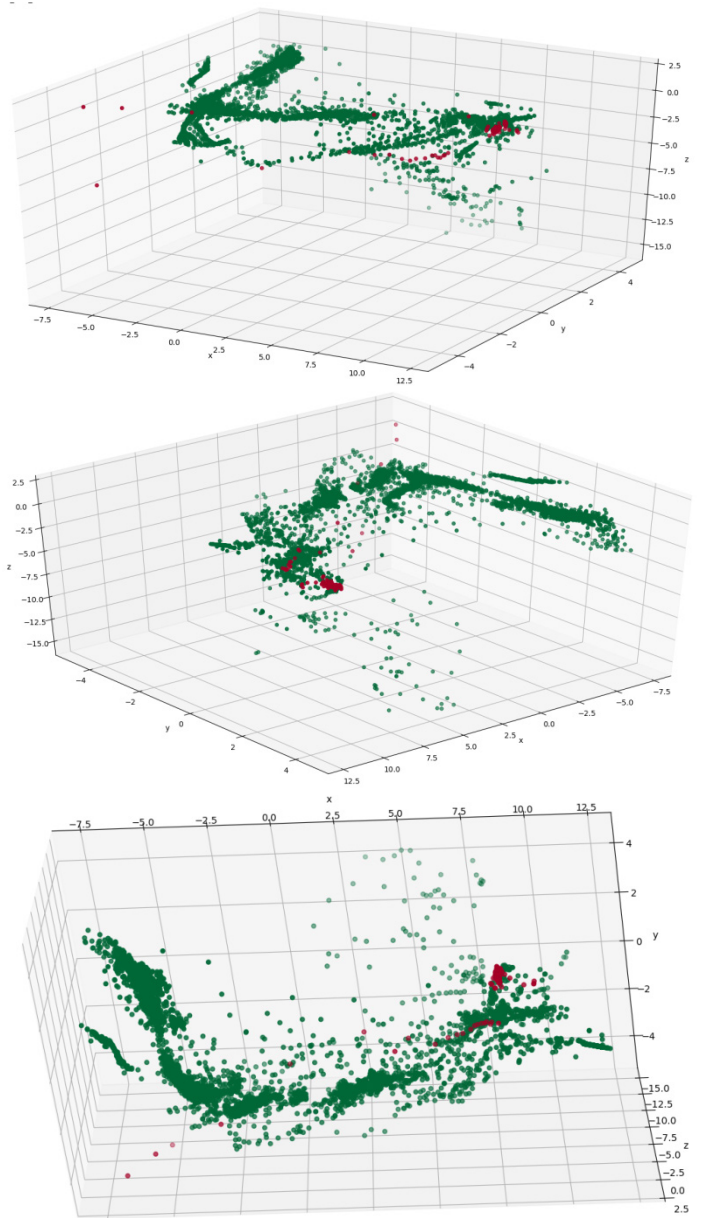


Figure 5: Isomap mappings to 3-D

(e.g. boomerang). The failed points are located in several areas of the structure in the t-SNE mapping. This is a consequence of the property expressed by the mapping (Section 3.2, t-SNE), which focuses on preserving conditional probability distributions within neighbourhoods, rather than distances. Exposing cluster structure more clearly is one of the strengths of t-SNE.

The Isomap plots (Figure 5) show a similar data structure to PCA and Sammon with much of the data falling along a boomerang-like shape. Also similar to PCA and Sammon, the failed points are all concentrated at one end of that main boomerang shape ($5 \leq x \leq 10$). However, more prevalent in the Isomap plot, than with the other plots, is one dense island of healthy data points just offset from that main shape, which will be discussed further in Section 6.2. The similar data structure seen with the PCA, Sammon and Isomap plots can be attributed to the fact that nonlinear effects are mild.

The transformation of the 31-dimension space representing the 31 turbocharger parameters to a low-dimensional 3-D space shows a distinct data structure in the various mapping techniques. The failed points are not mapped to an obvious outlier location in the transformations that can be easily distinguished from the healthy points.

6. Relating sensor signals to internal data structure

In order to better understand these data structure visualizations in a more physical sense, e.g. from the perspective of the operator or maintainer of the engine, the next step was to select a handful of the engine signals and examine how each of these signals impacted and was represented in the data structure. One of the aims was to determine if certain regions of the mappings could be associated with distinct operational states of the engine. Another aim was to demonstrate that individual signals were appropriately represented in the low-dimensional mapping.

6.1. Selected signals

As detailed in Table 1, a variety of sensors recording speeds, temperatures, pressures and torque were included in the analysis. A small subset of signals was selected to investigate the internal data structure of the mappings. The selected signals were: Turbo A speed, Turbo B speed, A1 cylinder exhaust gas temperature, B2 cylinder exhaust gas temperature, and engine speed.

As there was a known failure of the Turbo A speed signal, the Turbo B speed signal was selected to compare the recorded values of the two turbos. In a similar manner, at the time of failure, the exhaust cylinder temperature sensors went into alarm. To enable comparison of the exhaust temperatures, one cylinder exhaust temperature signal was selected from each bank. Finally, engine speed was selected, although it was not part of the 31 turbocharger parameters used for the transformations. This sensor was included as it is the main driving factor for change in the values of the other sensor signals.

6.2. Representation of signals in low-dimensional mappings

Each of the low-dimensional mappings was then overlaid with a heat map of one of these sensor signals. Where instead of the data points coloured green and red for healthy and failed points respectively, they were coloured according to the relative value of the particular signal in its operating range, with blue corresponding to the low end of the range and red for the high end of the operating range.

In the following section of the paper, two regions of interest are highlighted: i) the points surrounding the time frame where the Turbo A speed sensor experienced a fault and failed to record; and ii) the points surrounding the five minutes to the turbocharger seizure.

These two regions were chosen for further investigation because they appear to stand out the most in the visualizations. The Isomap low-dimensional mappings are shown in this paper with an overlay of the heat map for the five selected signals. Figure 6

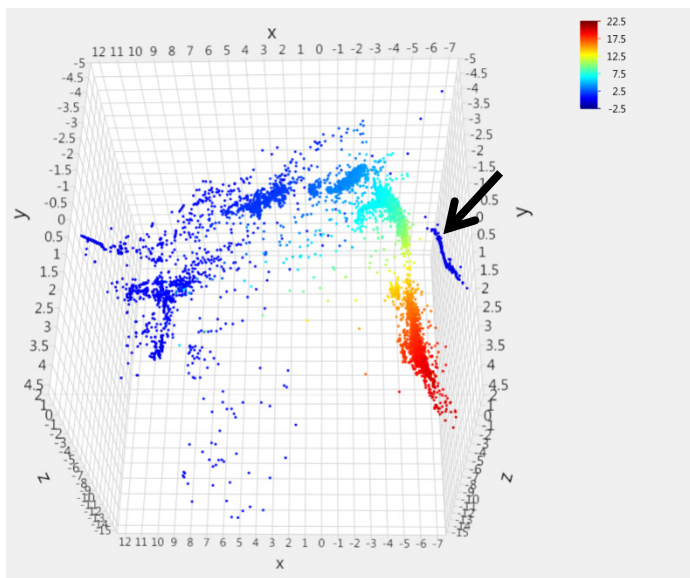


Figure 6: Turbo A speed sensor value [kRPM] overlay onto Isomap transformation

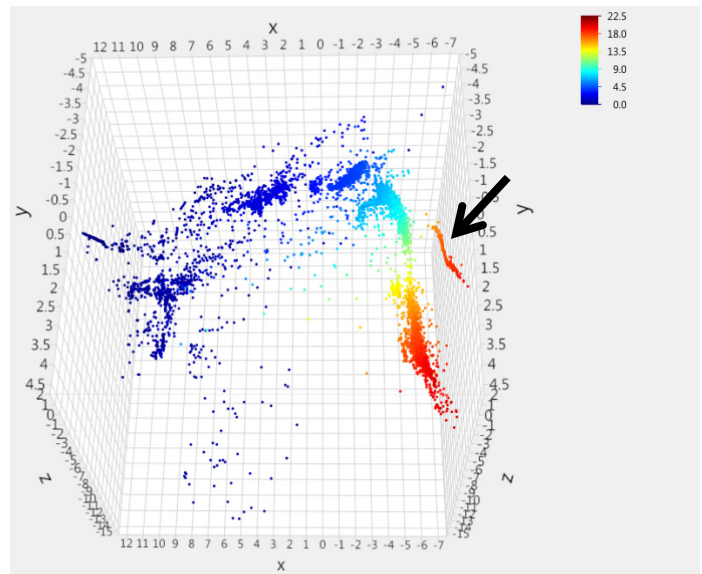


Figure 7: Turbo B speed sensor value [kRPM] overlay onto Isomap transformation

shows the overlay of the Turbo A speed signal on the Isomap transformation. Figure 7 shows the overlay of the Turbo B speed signal on the Isomap transformation. Since both these signals are speed sensors on turbochargers, the expectation is that these two plots should be very similar. For the most part the two plots are indeed very similar, however, it should be noted that the speed ranges are slightly different in the legend.

The distribution of the data points in the mapping indicates that almost all the near-zero/low-speed data points are found in one region of the data structure ($x \geq -2$) while the data points with high speed values are concentrated in a different region ($-7 \leq x \leq -5$). There is a strong gradient between these two regions containing a transition region of the intermediate speeds.

However, an abnormality to the general structure of the data is evident in the high power region of the plots. In Figure 6, the heat map indicates zero values for the Turbo A speed sensor for the data points in the area indicated by the arrow. In the same area, indicated by the arrow in Figure 7, the Turbo B speed sensor portrays high speed values. The overlays of the other examined signals displayed expected values for those data points in that high power region. Thus this group of points is likely related to the loss of the Turbo A speed sensor. Indeed the data points in this area were later found to coincide with the known initial loss of the Turbo A speed sensor readings.

Figure 8 shows the overlay of the Turbo A Cylinder 1 exhaust gas temperature signal on the Isomap transformation. Figure 9 shows the overlay of the Turbo B Cylinder 2 exhaust gas temperature signal on the Isomap transformation. Figure 10 shows the overlay of the engine speed signal on the Isomap transformation. The plots of the two cylinder exhaust temperatures from each bank (Figures 8 and 9) are quite similar to each other, as one would expect. They are also structured similar to the Turbo speed overlays where the data points with low temperature values also have low speed values, while the high temperature points have high speed values.

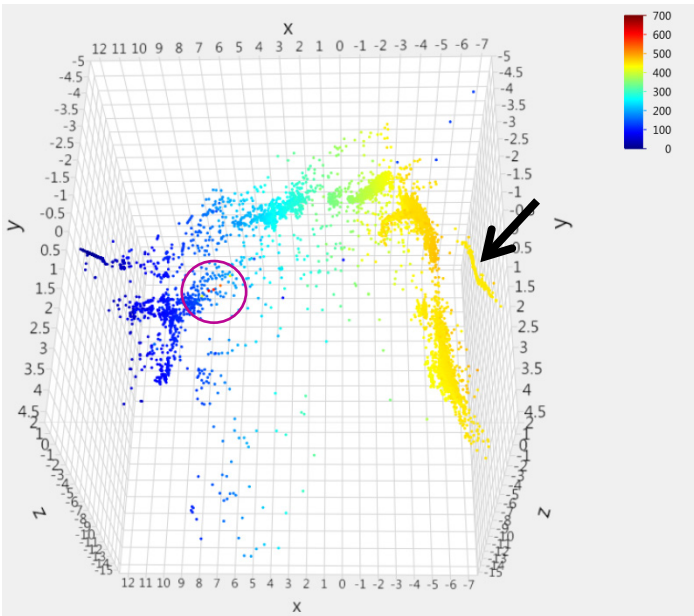


Figure 8: Turbo A Cylinder 1 exhaust gas temperature [°C] overlay onto Isomap transformation

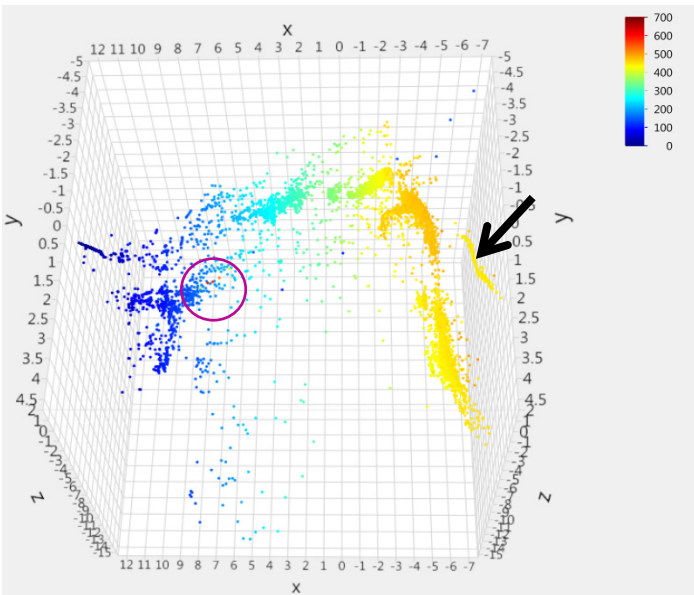


Figure 9: Turbo B Cylinder 2 exhaust gas temperature [°C] overlay onto Isomap transformation

A small cluster of high cylinder temperature indications is found within the low operating range group indicated by the purple circle in Figures 8 and 9. These points were later found to be the points occurring five minutes before the turbocharger seizure. The location of these points in the low speed range is consistent with the findings of the turbocharger incident report (described in Section 2), where an increase in speed was requested but no increase in speed was realized although the exhaust gas temperatures rose to high levels invoking the alarms.

Although engine speed was not a signal used to generate the Isomap transformation, here it is used as a means to support the previously observed trends of low and high operating ranges in the plots. From Figure 10, the engine speed overlay demonstrates a similar distribution of the low and high engine speeds in the data

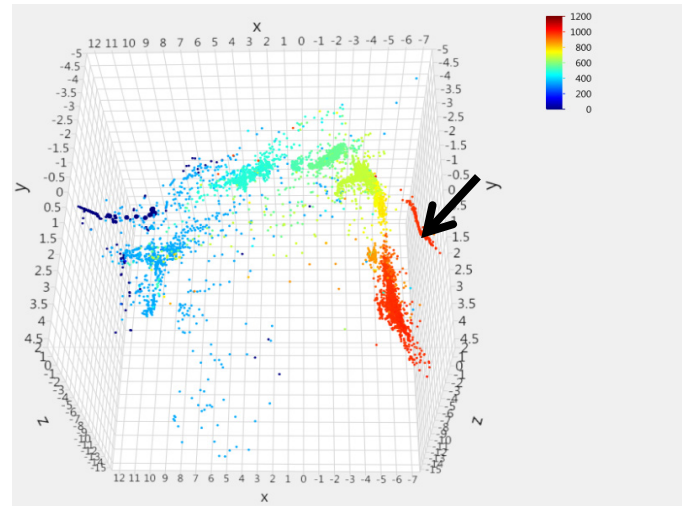


Figure 10: Engine speed [RPM] overlay onto Isomap transformation

structure. Looking again at the data points indicated by the arrow in Figure 10, the trend seen previously with the Turbo speeds and cylinder exhaust temperatures is confirmed.

According to Figure 5, the data points indicated by the arrow in the prior figures are clearly distant from the main set of data and are still designated as ‘healthy’. It was also pointed out that the small cluster of cylinder exhaust gas temperature indications (some of the points inside the region of the purple circle in Figures 8 and 9) represent values that were recorded between five hours and five minutes to the turbocharger seizure. These points embed themselves inside low operating ranges of these signals and are also designated as ‘healthy’ data (Figure 5). In both cases, the aforementioned groups of points display interesting traits through the unique combination of their healthy/failed designation, the value of the overlaid signal that they show, and their distance from other groups of data points. This demonstrates that the failure of the turbocharger system may not always be associated with extreme signal values, as one may assume. These traits help uncover new trends in the data structure, like the abnormality indicated by the arrow or the cluster of points within the purple circle in the preceding figures, which should be investigated. Thus, being able to combine expert domain knowledge of sensors and failures with data mining tools is an invaluable method of extracting and understanding information from sensor measurements.

7. Concluding Remarks

This work describes continued analysis of sensor data for the turbocharger subsystem of a diesel engine system. The engine has hundreds of sensors monitoring both the inputs of the engine operator and the resulting equipment outputs. The objective of the data analysis was to characterize and distinguish the healthy and failed states of the turbocharger seizure as recorded by the diesel engine sensor system. The analysis approach included the mapping of high-dimensional sensor data to a low-dimensional space using a variety of linear and nonlinear techniques in order to highlight and visualize the underlying structure of the information.

Estimates of the intrinsic dimension were obtained to determine the appropriate number of dimensions required by the

low-dimensional transformations and to guide the interpretation of the visualization spaces produced. For this case, three dimensions was an appropriate estimate. Through the unsupervised process of these transformations, the structure of the turbocharger data could be visualized and inspected. The transformation methods included principal components, Sammon mapping, t-Distributed Stochastic Neighbour Embedding, and Isomap. The transformation of the 31-dimension space representing the 31 turbocharger parameters to a low-dimensional 3-D space shows a distinct data structure in the various mapping techniques. The failed points are not mapped to an obvious outlier location in the transformations that can be easily distinguished from the healthy points.

In order to gain more physical insight into the internal data structure of the resulting mappings, the transformation plots were analyzed from the perspective of several engine sensor signals. By overlaying operating ranges of individual sensor signals, certain regions of the mappings could be associated with distinct operational states of the engine. Low and high operating engine regions could be clearly seen in the internal data structure, and several anomalies could be identified which were then associated to various points in the turbocharger seizure. These results are extremely promising and demonstrate how operational knowledge can be easily incorporated with the data analytics tools to enhance the insights that can be gained from the sensor measurements.

In this work, data mining and machine learning techniques are implemented to gather useful insights from the large amounts of sensor data collected for this diesel engine system. By incorporating expert domain knowledge with the low-dimensional representations of the data, a more practical understanding of the data structure presented in the mappings is provided which helps ensure that the results are relevant and accessible to the operator and end-user.

Future efforts are aimed at expanding this analysis to data from other diesel engines and other failures in the engine system. Further work to generalize the analysis to a diesel engine system model instead of a turbocharger-specific model is in progress. Efforts are also underway to better characterize the healthy and failed states, through classification and anomaly detection techniques. The development and implementation of these tools should help enable advance indication of a change in behavior that could be investigated before a major incident.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to acknowledge and thank Defence Research and Development Canada for their support of this research.

References

[1] C. Cheung, J. J. Valdés, A. Lehman Rubio, R. Salas Chavez, C. Bayley, "Low-dimensional spaces for the analysis of sensor network data: identifying behavioural changes in a propulsion system" in 5th International Symposium on Robotics and Intelligent Sensors, Ottawa, Canada, 2017. <https://www.doi.org/10.1109/IRIS.2017.8250133>

[2] A. Widodo, B. Yang, "Support vector machine in machine condition monitoring and fault diagnosis" *Mech. Syst. Signal Pr.*, 21(6), 2560-2574, 2007. <https://www.doi.org/10.1016/j.ymssp.2006.12.007>

[3] S. Li, "Induction motor fault diagnosis based on fuzzy support vector machine" in 3rd International Conference on Electromechanical Control Technology and Transportation, 2018. <https://www.doi.org/10.5220/0006975105880592>

[4] M. Saimurugan, K. Ramachandran, V. Sugumar, N. Sakthivel, "Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine" *Expert Syst. Appl.*, 38(4), 3819-3826, 2011. <https://www.doi.org/10.1016/j.eswa.2010.09.042>

[5] M.C. Garcia, M.A. Sanz-Bobi, J.D. Pico, "SIMAP: intelligent system for predictive maintenance" *Comput. Ind.*, 57(6), 552-568, 2006. <https://www.doi.org/10.1016/j.compind.2006.02.011>

[6] U. Fayyad, G. Piatesky-Shapiro, P. Smyth, "From data mining to knowledge discovery in databases" *AI Mag.*, 17(3), 1996. <https://doi.org/10.1609/aimag.v17i3.1230>

[7] C. Torrano-Gimenez, H.T. Nguyen, G. Alvarez, K. Franke, "Combining expert knowledge with automatic feature extraction for reliable web attack detection" *Secur. Commun. Netw.*, 8, 2750-2767, 2015. <https://www.doi.org/10.1002/sec.603>

[8] F. Kuusisto, I. Dutra, M. Elezaby, E.A. Mendonça, J. Shavlik, E.S. Burnside, "Leveraging expert knowledge to improve machine-learned decision support systems" *AMIA Jt Summits Transl Sci Proc.*, 87-91, 2015.

[9] J. Valdés, C. Y. S. Létourneau, "Data fusion via nonlinear space transformations" in 1st International Conference on Sensor Networks and Applications, San Francisco, USA, 2009.

[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press Professional Inc., 1990.

[11] E. Facco, M. d'Errico, A. Rodriguez, A. Laio, "Estimating the intrinsic dimension of datasets by a minimal neighborhood information" *Sci. Rep.-UK*, 7(1), 2017. <https://www.doi.org/10.1038/s41598-017-11873-y>

[12] D. Granata, V. Carnevale, "Accurate estimation of the intrinsic dimension using graph distances: Unraveling the geometric complexity of datasets" *Sci. Rep.-UK*, 6(1), 2016. <https://doi.org/10.1038/srep31377>

[13] P. Campadelli, E. Casiraghi, C. Ceruti, A. Rozza, "Intrinsic dimension estimation: Relevant techniques and a benchmark framework" *Math. Probl. Eng.*, 2015. <https://doi.org/10.1155/2015/759567>

[14] E. Levina, P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension" in 17th Conference in Neural Information Processing Systems, Vancouver, Canada, 2004.

[15] P. Grassberger, I. Procaccia, "Measuring the strangeness of strange attractors" *Physica D*, 9(1-2), 189-208, 1983. [https://www.doi.org/10.1016/0167-2789\(83\)90298-1](https://www.doi.org/10.1016/0167-2789(83)90298-1)

[16] J.A. Costa, A.O. Hero, "Geodesic entropic graphs for dimension and entropy estimation in manifold learning" *IEEE T. Signal Proces.*, 52(8), 2210-2221, 2004. <https://www.doi.org/10.1109/tsp.2004.831130>

[17] K. Pettis, T. Bailey, A. Jain, R. Dubes, "An intrinsic dimensionality estimator from near-neighbor information" *IEEE T. Pattern Anal.*, 1(1), 25-37, 1979. <https://www.doi.org/10.1109/tpami.1979.4766873>

[18] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.

[19] J. Kruskal, "Nonmetric multidimensional scaling: a numerical method" *Psychometrika*, 29(2), 115-129, 1964. <https://www.doi.org/10.1007/bf02289694>

[20] J. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis" *Psychometrika*, 29(1), 1-27, 1964. <https://www.doi.org/10.1007/bf02289565>

[21] I. Borg, P. Groenen, *Modern multidimensional scaling - theory and applications*, Springer Series in Statistics, 1997.

[22] I. Borg, P.J.F. Groenen, P. Mair, *Applied Multidimensional Scaling*, Springer Verlag, 2013.

[23] J. W. Sammon, "A nonlinear mapping for data structure analysis" *IEEE T. Comput.*, 18(5), 401-409, 1969. <https://www.doi.org/10.1109/tc.1972.5008933>

[24] G. Hinton, S. Roweis, "Stochastic neighbor embedding" *Adv. Neur. Inf. Proc. Systems*, 15, 857-864, 2003.

[25] L. van der Maaten, G. Hinton, "Visualizing data using t-sne" *J. Mach. Learn. Res.*, 9, 2579-2605, 2008.

[26] M. Nguyen, S. Purushotham, H. To, C. Shahabi, "M-tsne: a framework for visualizing high-dimensional multivariate time series", University of Southern California, 2017. <https://arxiv.org/abs/1708.07942>

[27] J. Tenenbaum, V. de Silva, J. Langford, "A global geometric framework for nonlinear dimensionality reduction" *Science*, 290(5500), 2319-2323, 2000. <https://www.doi.org/10.1126/science.290.5500.2319>

[28] M. Bernstein, V. de Silva, J. Langford, J. Tenenbaum, "Graph approximations to geodesics on embedded manifolds" *Stanford University, Tech. Rep.*, 2000.

[29] V. de Silva, J. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction" *Adv. Neur. Inf. Proc. Systems*, 15, 721-728, 2003.

[30] H. Choi, S. Choi, "Robust Kernel Isomap" *Pattern Recogn.*, 40(3), 853-862, 2007. <https://www.doi.org/10.1016/j.patrec.2006.04.025>

Fuzzy Uncertainty Management in Multi-Shift Single-Vehicle Routing Problem

Francesco Nucci*

Department of Engineering for Innovation, University of Salento, Lecce, Italy

ARTICLE INFO

Article history:

Received: 14 August, 2018

Accepted: 19 October, 2018

Online: 01 November, 2018

Keywords:

Vehicle Routing Problem

Fuzzy Uncertainty

Scheduling

ABSTRACT

Our research deals with the single-vehicle routing problem (VRP) with multi-shift and fuzzy uncertainty. In such a problem, a company constantly uses one vehicle to serve demand over a scheduling period of different work shifts. Our issue relies on a routing problem in maintenance jobs, where a crew executes jobs in different sites. The crew runs during several work shifts but repeatedly returns to the depot before the shift ends. The goal is executing all the activities minimizing the makespan. We examine the impact of uncertainty in driving and maintenance processing time on system performance. We realize an Artificial Immune Heuristic to find optimal solutions considering both makespan and overtime avoidance. First, we introduce a framework to assess the uncertainty impact. Then, we produce a numerical company case study to examine the problem. Outcomes present significant improvements are obtained with the proposed approach.

1 Introduction

Vehicle routing problem (VRP) consists in determining a set of routes to visit a fixed set of customers, in order to minimize the total path length. Several versions of the VRP exist. If each customer specifies availability time windows, we deal with a VRP with Time Windows (VRPTW). Basic variants of the VRP consider the route planning for a vehicle fleet in a single period (shift). In this case, vehicles should come back to the depot before the shift ends. This problem is derived from a healthcare routing question. The healthcare company regularly dispatch products to medical sites. In this case, if overtime is allowed, performance could be significantly upgraded [1]–[4]. For instance, if a location scheduled in the next shift is on the current return path to the depot, a limited overtime allows vehicle serving it. This can significantly diminish next shift workload.

Investigation on the connection between health status and work hours documented concern about the influence of working long hours on people fitness [5]. Furthermore, working long hours increases the chance of micro-sleep in car drivers [6]. In these conditions, the company could be accused of vehicle collision due to enormous workload planning [7]. In fact, shift scheduling objective should limit both anxiety and

harmful consequences on healthiness maintaining the work shift as steady as possible.

In an effort of limiting extra time, uncertainty effect on scheduling has to be restricted. Frequently, in optimization problems, data are supposed to be known with certainty. Nevertheless, in practice this is rare. Most commonly, the real data depend on uncertainty because of their irregular essence. Because the optimization problem solution usually reveals a great inclination to the data disturbances, neglecting the data ambiguity may conduct to suboptimal or infeasible solutions for a real case. Stochastic VRP (SVRP) was introduced in [8]. See [9] and [10] for a complete review of SVRP.

Robust optimization is a significant technique to deal with optimization problems subject to uncertainty [11]. The main motive for considering robust optimization is the non-stochastic nature of uncertain parameters. In such a case, a methodology is required to analyze compromise between performance and robustness.

Fuzzy set theory is a useful approach to handle uncertain information, while the stochastic approach is suitable to manage the stochastic parameters in VRP [12]. Some VRP papers utilize fuzzy sets theory for studying the influence of uncertain factors [13]–[23].

*Corresponding Author: F. Nucci, Via per Monteroni, Lecce, Italy, Tel.: +39.0832.297805, Email: francesco.nucci@unisalento.it

In this work, the VRP in a fuzzy random context is analyzed and the fuzzy random theory is adopted to manage such uncertain data. Fuzzy random variables illustrate a well-formalized notion involving data obtained from a random test in which data are supposed to be fuzzy sets.

We examine a multi-shift VRP with no overtime allowed considering fuzzy driving and job processing time. The question we considered is derived from a routing problem in maintenance activities. A maintenance team performs jobs in different sites using a vehicle for movements. A crew works in shifts and should come back to the depot before the shift ends. The goal is completing the maintenance activities in various places reducing the makespan. Since we deal with maintenance jobs, we ignore customer waiting times. We investigate the influence of the uncertainty of driving and job processing time on makespan and overtime avoidance.

The body of this paper is structured as follows. In Section 2, we report the problem formulation. In Section 3, we propose the Artificial Immune Heuristic to solve the problem. We present in Section 4 a case study analyzed with the proposed approach. We give concluding remarks in Section 5.

2 Problem formulation

2.1 Classical problem

Problem notation is reported in Table 1.

Let $N = \{1, 2, \dots, n\}$ be the set of maintenance jobs to be executed at customer places. Parameter $d_{i,j}$ stands for the travel time between the location of jobs i and j . Parameter q_i expresses job i processing time. Parameters e_i and l_i represent job i time window.

Maximum shift duration is L , whereas p is the number of shifts in the scheduling horizon and $P = \{1, 2, \dots, p\}$ is shift set. We generate $p + 1$ depot copies described by nodes $n + 1, \dots, n + p + 1$:

- node $n + 1$ is the origin depot of shift 1,
- node $n + h$ represents the destination depot of shift $h - 1$ together with the origin depot of shift h , with $h \in 2, \dots, p$.
- node $n + p + 1$ stands for the destination depot of last shift p .

The problem can be represented as a directed graph $G = (V, E)$, where $V = N \cup \{n + 1, \dots, n + p + 1\}$. Arcs $(n + h, n + h + 1)$ are used in the graph to illustrate cases in which a vehicle is not exploited during shift h .

Lastly, b_h and c_h represent the begin and end time of shift h , with $b_h = (h - 1)L$ and $c_h = hL$, $h \in P$.

In this work, the next hypotheses are made. Driving times are positive, i.e. $d_{ij} > 0, \forall i, j \in V$. The triangular inequality is valid for d_{ij} , i.e. $d_{ij} \leq d_{ik} + d_{k,j}, \forall i, j, k \in V$. At least any single job can be executed in a shift because we suppose shift length satisfies $L \geq d_{0,i} + q_i +$

$d_{i,n+1}, \forall i \in N$. Time windows are greater than the activity processing time: $l_i - e_i \geq q_i, \forall i \in N$. Consequently, a linear formulation for VRPTW is considered. Moreover, a feasible distance set is defined and all the jobs can be processed within their time window and shift duration.

In the following, decision variables are presented:

- $x_{i,j} = 1$ if the crew drives from node i to j , and 0 otherwise,
- α_i arrival time of the crew at node $i \in V, i \neq n + 1$,
- δ_i departing time of the crew at node $i \in V, i \neq n + p + 1$,
- σ_h real shift h duration,
- $y_h = 1$ if shift h is used by the crew, and 0 otherwise,

The problem is expressed as a mixed integer program (MIP) described in (1)-(13):

- The objective function (1) minimizes the system makespan.
- Constraints (2) and (3) provide that the depot nodes $\{n + 1, \dots, n + p + 1\}$ are always visited. Depot node $n + h + 1$ is visited after $n + h$. Depot nodes separate different shifts. All job nodes visited between node $n + h$ and $n + h + 1$ are served during shift h . If the vehicle is not used in shift h , no job node is visited between node $n + h$ and $n + h + 1$. Such constraints assure that any shift path starts from the origin depot and ends at the destination depot.
- Constraints (4) ensure that each customer is visited once, that is each node $i \in N$ is inspected once.
- Constraints (5) force that for all intermediate nodes (first node $n + 1$ and last node $n + p + 1$ are excluded) the inflow is equal to the outflow.
- Constraints (6) are the sub-tour elimination constraints and ensure coherence of time variables. Parameter M is an upper bound of $\delta_i + d_{ij}, \forall i, j \in V$.
- Constraints (7) represent the connection between arrival and departing times.
- Constraints (8)-(9) define time window constraints for the customer nodes, whilst (10)-(11) represent the time windows constraints for the depot nodes.
- Constraints (12) provide the real shift duration. Considering (10)-(12), parameter L is an upper bound of shift duration $\sigma_h, \forall h = 1, \dots, p$.
- Constraints (13) determine whether shift h is exploited or not. In fact, if $x_{n+h,n+h+1} = 1$, then no demand node is inspected between depot h and $h + 1$ (shift h).

Table 1: Problem notation

Symbol	Description
n	Number of Jobs
$N = \{1, 2, \dots, n\}$	Job set
$i, j \in N$	Job index
$d_{i,j}$	Travel time between job i and job j locations
q_i	Job i processing time
e_i	Begin of job i time window
l_i	End of job i time window
L	Maximum shift duration
p	Number of shifts in the planning horizon
$P = \{1, 2, \dots, p\}$	Shift set
$h \in P$	Shift set index
b_h	Begin time of shift h , equals to $(h - 1)L$
c_h	End time of shift h , equals to hL
$x_{i,j} \in \{0, 1\}$	Crew travels from node i to j , if $x_{i,j} = 1$
δ_i	Departing time of crew at node $i \in V, i \neq n + p + 1$
α_i	Arrival time of crew at node $i \in V, i \neq n + 1$
σ_h	Actual shift h duration
$y_h \in \{0, 1\}$	Crew is active in shift h , if $y_h = 1$
M	Upper bound of $\delta_i + d_{ij}, \forall i, j \in V$.

$$\begin{aligned}
 & \text{Minimize} && \alpha_{n+p+1} && (1) \\
 & \text{subject to:} && && \\
 & \sum_{j \in N \cup \{n+h+1\}} x_{n+h,j} = 1 && \forall h \in P && (2) \\
 & \sum_{i \in N \cup \{n+h\}} x_{i,n+h+1} = 1 && \forall h \in P && (3) \\
 & \sum_{j \in V \setminus \{n+p+1\}} x_{j,i} = 1 && \forall i \in N && (4) \\
 & \sum_{j \in V} x_{j,i} - \sum_{j \in V} x_{i,j} = 0 && \forall i \in V \setminus \{n+1, n+p+1\} && (5) \\
 & \delta_i + d_{ij} \leq \alpha_j + Mx_{i,j} && \forall i, j \in V && (6) \\
 & \alpha_i + q_i \leq \delta_i && \forall i \in N && (7) \\
 & e_i \leq \alpha_i && \forall i \in N && (8) \\
 & \delta_i \leq l_i && \forall i \in N && (9) \\
 & b_h \leq \delta_{n+h} && \forall h \in P && (10) \\
 & \alpha_{n+h+1} \leq c_h && \forall h \in P && (11) \\
 & \alpha_{n+h+1} - \delta_{n+h} \leq \sigma_h && \forall h \in P && (12) \\
 & y_h = 1 - x_{n+h,n+h+1} && \forall h \in P && (13)
 \end{aligned}$$

2.2 Problem with fuzzy uncertainty

We describe the uncertainty on driving times and job processing times with triangular fuzzy numbers, see Fig. 1. We consider the following notation:

- fuzzy travel time: $\tilde{d}_{ij} = (d_{ij}^A, d_{ij}^B, d_{ij}^C) \forall i, j \in V$
- fuzzy job processing time: $\tilde{q}_i = (q_i^A, q_i^B, q_i^C) \forall i \in N$

For this reason, decision variables related to arrival and departing time at node i turn into fuzzy variables $\tilde{\alpha}_i$ and $\tilde{\delta}_i$. Also, actual shift duration variables $\tilde{\sigma}_i$ become fuzzy. Therefore, fuzzy numbers are introduced both in objective function (1) and constraints (6)-(12).

An important issue arises when models with fuzzy parameters are considered: defining the comparison method for objective function values [24]. In order to solve this question, we have to consider the ranking of fuzzy numbers [25, 26, 27]. The fuzzy ranking method is part of the solution approach in order to solve mathematical programs having coefficients of the objective function and coefficients of the constraints represented by fuzzy numbers.

The Expected Existence Measure (EEM) operator can be applied to a temporal fuzzy number \tilde{t} , see (14) and [28, 29].

$$EEM_{\tilde{t}}(t) = \frac{\int_{-\infty}^t \mu_{\tilde{t}}(\tau) d\tau}{\int_{-\infty}^{+\infty} \mu_{\tilde{t}}(\tau) d\tau} = \frac{\int_{t_A}^t \mu_{\tilde{t}}(\tau) d\tau}{\int_{t_A}^{t_C} \mu_{\tilde{t}}(\tau) d\tau} \in [0, 1] \quad (14)$$

Briefly, *EEM* defines the possibility with which a fuzzy event occurred at a certain time. Assigned a temporal value t_0 , the *EEM* stands for the possibility a temporal fuzzy number \tilde{t} is lower than t_0 . We denote this as $\Phi(\tilde{t} \leq t_0) = EEM_{\tilde{t}}(t_0) = \gamma_0$. Consequently, $\Phi(\tilde{t} \geq t_0) = 1 - \gamma_0$. In Fig. 1, the membership function $\mu_{\tilde{t}}(t)$ of the triangular fuzzy number \tilde{t} and the corresponding $EEM_{\tilde{t}}(t)$ are represented. Note that $\Phi(\tilde{t} \leq t^A) = EEM_{\tilde{t}}(t^A) = 0$ and $\Phi(\tilde{t} \leq t^C) = EEM_{\tilde{t}}(t^C) = 1$.

Various studies solve fuzzy mathematical problems by exploiting fuzzy ranking methods. First, fuzzy mathematical programming problems are converted into classical mathematical problems. Second, conventional techniques are applied. In the following section, we propose an innovative approach to consider the fuzzy variables in the *entire* solution process. We present the decision-maker with different optimal solutions based on 2-factor comparison: the objective function value and the possibility degree with which constraints are satisfied.

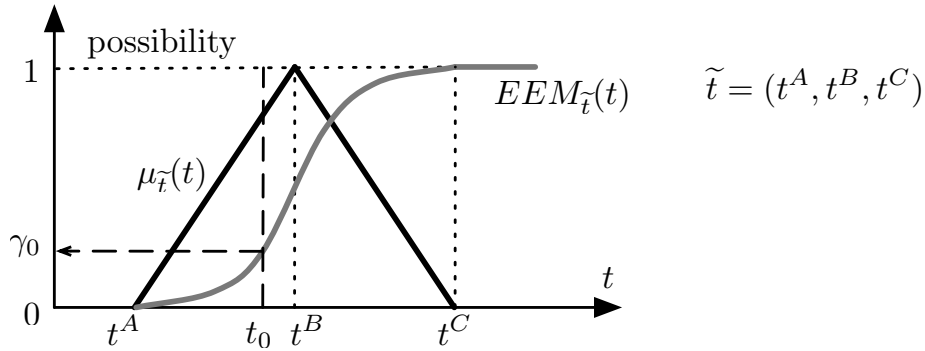


Figure 1: Triangular Fuzzy Number

In the previous work [30] we published on this question, we dealt with a different objective function: number of exploited shifts and maximum shift duration. In this research, we consider a different problem in which makespan is minimized. Furthermore, the solution algorithm has been improved with an appropriate solution ‘affinity’ definition in Section 3.2. Moreover, we extended the experimental campaign in order to investigate our approach performance. For such a reason a mathematical background is introduced in sections 4.2, 4.2.2, and 4.2.3.

3 Solution Method: Artificial Immune Heuristic

The animal immune system is a versatile pattern-recognition system that protects against foreign viruses and bacteria. The immune system is able to recognize and kill germs. The immune system cells, named *antibodies*, are casually diffused in the body. The system reacts to pathogens and improves the process of recognizing and eliminating pathogens by using two principles: clonal selection and affinity maturation. When a pathogen invades the organism, clonal selection creates a number of immune cells that recognize and eliminate the pathogen. While cellular reproduction occurs, the cells experience high rate physical mutations, as well as a selective process. Cells with superior affinity to the invading pathogen spread into memory cells.

Artificial Immune Algorithm (AIA) is a meta-heuristic based on such system [31, 32, 33]. This paper intends proposing a fuzzy artificial immune algorithm to find optimal solutions for the aforementioned problem. AIA notation is reported in Table 2.

3.1 Solution Encoding

A solution is encoded as a string by using a fixed-length integer code and providing the order in which nodes are reached. Solution Ψ , representing a scheduling problem with $n = 5$ jobs on $p = 3$ available shifts, is reported in Fig. 2. Moreover, Fig. 2 shows the corre-

sponding graph path, see section 2. The origin for shift 1 is node $n + 1 = 6$ that is the path starting node. Then, the crew visits nodes 2, 3 and 5. Shift 1 ends at node $n + 2 = 7$. After, shift 2 begins and crew inspects nodes 1 and 4. Shift 2 ends when node $n + 3 = 8$ is reached. Finally, node $n + p + 1 = 9$ is reached (path end) because no job is executed in shift 3.

Because node $(n + 1)$ and $(n + p + 1)$ are respectively the ‘starting’ and ‘end’ node, solution Ψ is just defined by the succession of the residual nodes of the graph, that is $v_\ell \in V \setminus \{n + 1, n + p + 1\}, \forall \ell = 1, \dots, n + p - 1$. Note that when $x_{ij} = 1, i, j \in V, i \neq n + 1, j \neq n + p + 1$, then nodes i and j are directly connected in the graph, consequently $\exists \ell = 1, \dots, n + p - 2: v_\ell = i$ and $v_{\ell+1} = j$, that is nodes i and j are consecutive in the string encoding. In Fig. 2 we have $x_{14} = 1$, consequently $\exists \ell = 5, v_5 = 1$ and $v_6 = 4$.

Considering fuzzy numbers, we have to appropriately evaluate both the solution objective function f and solution feasibility degree γ .

First, considering the solution Ψ and adopting the fuzzy addition operator described in (15), we calculate actual duration of shift 1, $\tilde{\sigma}_1$, and shift 2, $\tilde{\sigma}_2$, see (16)-(17). Because shift 3 is not used by the crew, we have the same arrival time at node 8 and 9: $\tilde{\alpha}_9 = \tilde{\alpha}_8$. We calculate fuzzy makespan as $\tilde{\alpha}_9 = \tilde{\alpha}_8 = L + \tilde{\sigma}_2$. Indeed, makespan is the crew arrival time at the depot at the end of shift 2 (node 8), that is the sum of shift 2 begin time L and actual shift 2 duration $\tilde{\sigma}_2$. Since $\tilde{\sigma}_2$ is a fuzzy number, the crisp objective function f in (18) can be adopted using the modal value σ_2^B of the fuzzy number $\tilde{\sigma}_2$.

Second, we determine the degree with which the solution Ψ is feasible. So we calculate the possibility the shift duration constraints are enforced: $\Phi(\tilde{\sigma}_1 \leq L) = \gamma_1$ and $\Phi(\tilde{\sigma}_2 \leq L) = \gamma_2$. Time windows constraints are considered in the same way. Supposing for solution Ψ , fuzzy shift durations $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ are those reported in Fig. 3, it is possible that shift 1 actual duration is greater than maximum limit L . Indeed, half of fuzzy number $\tilde{\sigma}_1$ stays on the right side of L . The value $\gamma_1 = EEM_{\tilde{\sigma}_1}(L) = 0.5 < 1$ indicates the possibility that shift 1 ends before time L . Since, $\gamma_2 = EEM_{\tilde{\sigma}_2}(L) = 1$,

Table 2: AIA notation

Symbol	Description
$popsiz$	No. population antibodies
ng	No. generations
$pr1$	Rate of Rule1: full random node selection
$pr2$	Rate of Rule2: higher probability for closer node selection, $pr2 = 1 - pr1$
nc	No. clones in each generation
mr	Mutation Rate
nm	No. mutations in each generation
nea	No. exchangeable antibodies.
Ψ	Solution
v_ℓ	Solution encoding, $\forall \ell = 1, \dots, n + p - 1, v_\ell \in V \setminus \{n + 1, n + p + 1\}$
f	Solution makespan
γ	Solution feasibility degree
\mathcal{S}	Solution set, $ \mathcal{S} = popsiz$
Γ	solution Pareto set, $\Gamma \subset \mathcal{S}$
$A^{\mathcal{S}}(f, \gamma)$	Affinity for solution (f, γ) in set \mathcal{S}
exp	No. experiments in the campaign
\mathcal{E}	experiment set, $\mathcal{E} = \{1, \dots, exp\}$
ϵ	Experiment index, $\epsilon \in \mathcal{E}$
$\bar{\Gamma}^{\mathcal{E}}$	Hyper Pareto set: dominant solutions for set $\bigcup_{\epsilon \in \mathcal{E}} \Gamma_\epsilon$
ρ_ϵ	Impact of experiment ϵ

shift 2 duration constraint is certainly enforced. Finally, the solution Ψ feasibility degree γ in (19) is determined $\gamma = 0.5$.

$$\begin{aligned} \tilde{t}_s &= \tilde{t}_1 + \tilde{t}_2 = (t_1^A, t_1^B, t_1^C) + (t_2^A, t_2^B, t_2^C) \\ &= (t_1^A + t_2^A, t_1^B + t_2^B, t_1^C + t_2^C) \end{aligned} \quad (15)$$

$$\tilde{\sigma}_1 = \tilde{d}_{6,2} + \tilde{q}_2 + \tilde{d}_{2,3} + \tilde{q}_3 + \tilde{d}_{3,5} + \tilde{q}_5 + \tilde{d}_{5,7} \quad (16)$$

$$\tilde{\sigma}_2 = \tilde{d}_{7,1} + \tilde{q}_1 + \tilde{d}_{1,4} + \tilde{q}_4 + \tilde{d}_{4,8} \quad (17)$$

$$\begin{aligned} f &= \alpha_{n+p+1}^B \\ &= L + \sigma_2^B \end{aligned} \quad (18)$$

$$\begin{aligned} \gamma &= \min_{\substack{i \in N \\ h \in P}} \{\Phi(\tilde{\alpha}_i \geq e_i), \Phi(\tilde{\delta}_i \leq l_i), \Phi(\tilde{\sigma}_h \leq L)\} \\ &= \min\{0.5, 1\} = 0.5 \end{aligned} \quad (19)$$

Eventually, for a given solution, we associate a crisp objective function value f and a feasibility degree γ . We perform a Pareto comparison on (f, γ) pairs for determining the solution Pareto set Γ : f should be minimized and γ should be maximized.

For example, we consider an alternative solution Ψ' in which all the available 3 shifts are exploited. The alternative solution Ψ' is reported in Fig. 4. Unlike the previous solution Ψ , in Ψ' Job 5 has moved from shift 1 to shift 3. On one hand, we have objective function $f = 2L + \sigma_3^B$, see (18). On the other, for Ψ' , we have γ is equal to 1 because all shift duration constraints are definitely enforced ($\gamma_i = 1, \forall i = 1, \dots, N$), see $\sigma_h^C < 1, \forall h = 1, 2, 3$ in Fig. 4. Comparing the solution Ψ' to the previous one Ψ , we note that f value gets worse ($f^{\Psi'} > f^\Psi$) and γ values is better ($\gamma^{\Psi'} > \gamma^\Psi$). Indeed, with the alternative solution, shifts are completed in time but one additional shift is required. Since we

adopt a 2-factor comparison, both solution Ψ and solution Ψ' are Pareto optimal.

3.2 Affinity

Usually, in meta-heuristic approaches as AIA, an affinity level to be maximized is defined in order to take into account both solution objective value and constraint enforcement. Unfeasible solutions are penalized in terms of affinity. Solutions are ranked on the basis of their affinity values and, eventually, the one with the greatest affinity is defined as 'best solution'. In our approaches, two different factors (f and γ) are both considered in determining the affinity. Solutions having $\gamma = 0$ are certainly unfeasible.

We define the solution set as \mathcal{S} . Solutions in the Pareto set $\Gamma = \{(f, \gamma) | \gamma > 0\} \subset \mathcal{S}$ represent the optimal set, and we assign them the same maximum affinity value as in (20). Note that \bar{f} is the makespan of the solution having feasibility degree 1. Since f stands for makespan to be minimized, whereas affinity is to be maximized, we put \bar{f} in the second term as the makespan upper bound pL occupies the first term.

Using the normalized Manhattan distance from Pareto set (22), the affinity for solutions not in the Pareto set is computed (23). That is, considering a solution (f, γ) not in Γ , the minimum Manhattan distance from Pareto set is calculated and is used as a penalty for affinity. Quantity Δf is used as a weight in Manhattan distance for balancing the effects of makespan and feasibility degree.

3.3 Artificial Immune Algorithm

The proposed AIA is described in the following:

Step 1 - Initialization:

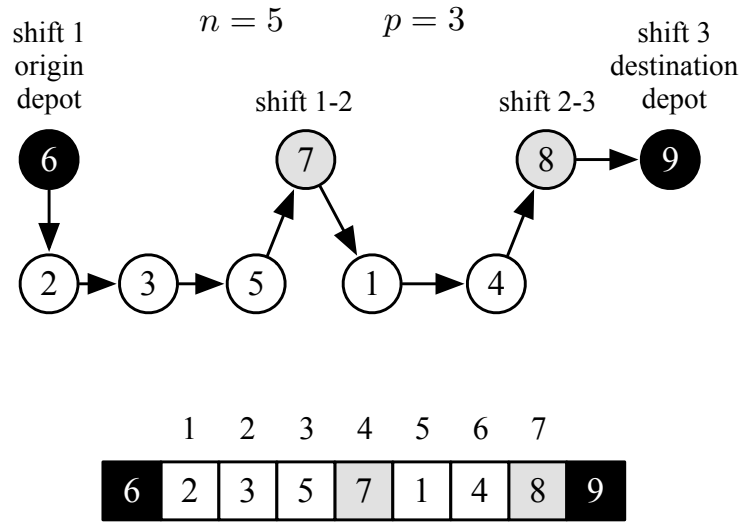


Figure 2: Solution Ψ graph path and encoding

$$\mathcal{A}^S(f, \gamma) = pL - \bar{f} \quad \forall (f, \gamma) \in \Gamma \subset \mathcal{S}; (\bar{f}, 1) \in \Gamma \quad (20)$$

$$\Delta f = \left[\max_{(f, \gamma) \in \Gamma} f \right] - \left[\min_{(f, \gamma) \in \Gamma} f \right] \quad (21)$$

$$\mathcal{D}(f, \gamma, f', \gamma') = |f - f'| / \Delta f + |\gamma - \gamma'| \quad (f, \gamma) \in \mathcal{S} \setminus \Gamma, (f', \gamma') \in \Gamma \quad (22)$$

$$\mathcal{A}^S(f, \gamma) = (pL - \bar{f}) \left[1 - \min_{(f', \gamma') \in \Gamma} \mathcal{D}(f, \gamma, f', \gamma') \right] \quad \forall (f, \gamma) \in \mathcal{S} \setminus \Gamma \quad (23)$$

- (a) *Parameter setting*: fix the initial population *popsiz*e, the number of generations *ng*, the rate *pr1* of Rule1, the rate *pr2* of Rule2, for each generation the number of clones *nc*, the mutation rate *mr*, the number of mutations *nm*, the number of exchangeable antibodies *nea*.
- (b) *Initial population generation*: create *pr1* · *popsiz*e initial solutions by Rule1 and produce *pr2* · *popsiz*e initial solutions by Rule2.
- (b) Include the *nm* extra antibodies to the current generation.
- (c) Exchange *nea* worst antibodies with new ones produced like those in Step 1b.
- (d) Copy the Pareto optimal antibodies to the next generation.

Step 2 - Objective function assessment:

- (a) calculate the objective function *f* in (18) and the feasibility degree γ in (19) for each antibody.
- (b) determine the Pareto set for (f, γ) pairs.
- (c) calculate the affinity for each antibody as reported in (20) and (23).

Step 3 - Clonal selection and expansion:

- (a) Take *nc* antibodies, from the population, with the greatest affinity.
- (b) Produce *nc* copies of the antibodies considered in Step 3a exploiting a binary tournament rule (randomly select two antibodies from *nc* antibodies and determine the best antibody).

Step 4 - Generating the next population:

- (a) Randomly choose *nm* antibodies from *nc* clones and use the mutations to create *nm* extra antibodies. Apply each mutation operator with the probability *mr*.

Step 5 - Conclusion test:

- (a) If the stopping criterion is met, return the Pareto optimal antibodies; otherwise, go to Step 2

At Step 1a, Rule1 is the full random rule: we randomly select one by one the nodes in the set $V \setminus \{n + 1, n + p + 1\}$. While in Rule2 we choose nodes using a probability that is inversely proportional to the distance between the current node and candidate node. Mutation operator randomly selects two solution indexes (Section 3.1) and swaps content. If the mutation makes unfeasible the depot node sequence $n + 2, \dots, n + p$, mutation is canceled.

Considering the ordinary AIA approach, the innovation in the method presented in this paper differs in the steps 2b, 2c and 4d. Indeed, step 2b and 2c are used to determine the new antibody affinity. Whereas, step 4d preserves the entire Pareto set in the next generation.

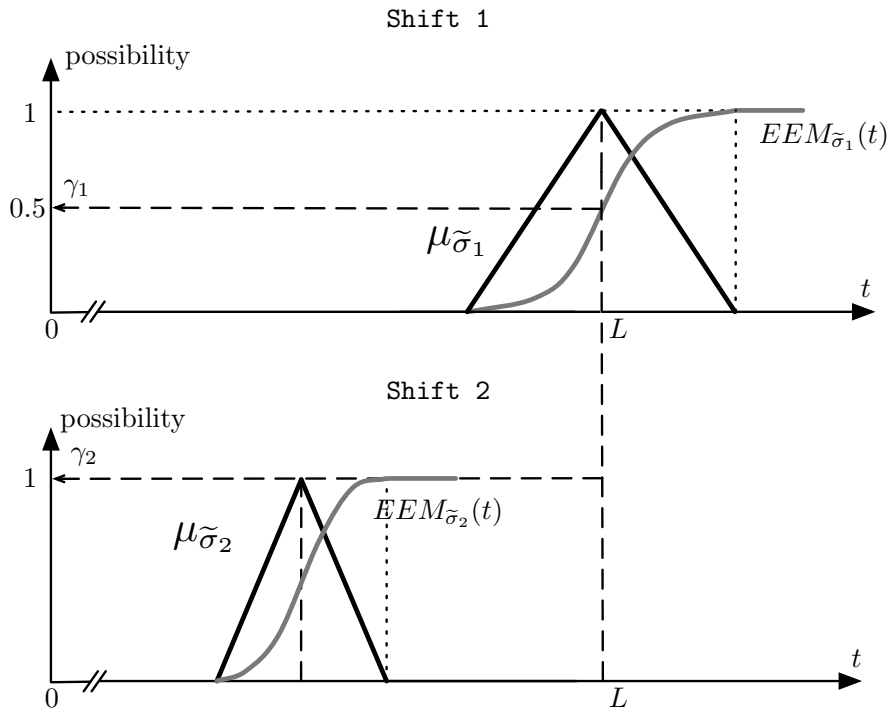


Figure 3: Fuzzy shift durations for solution Ψ

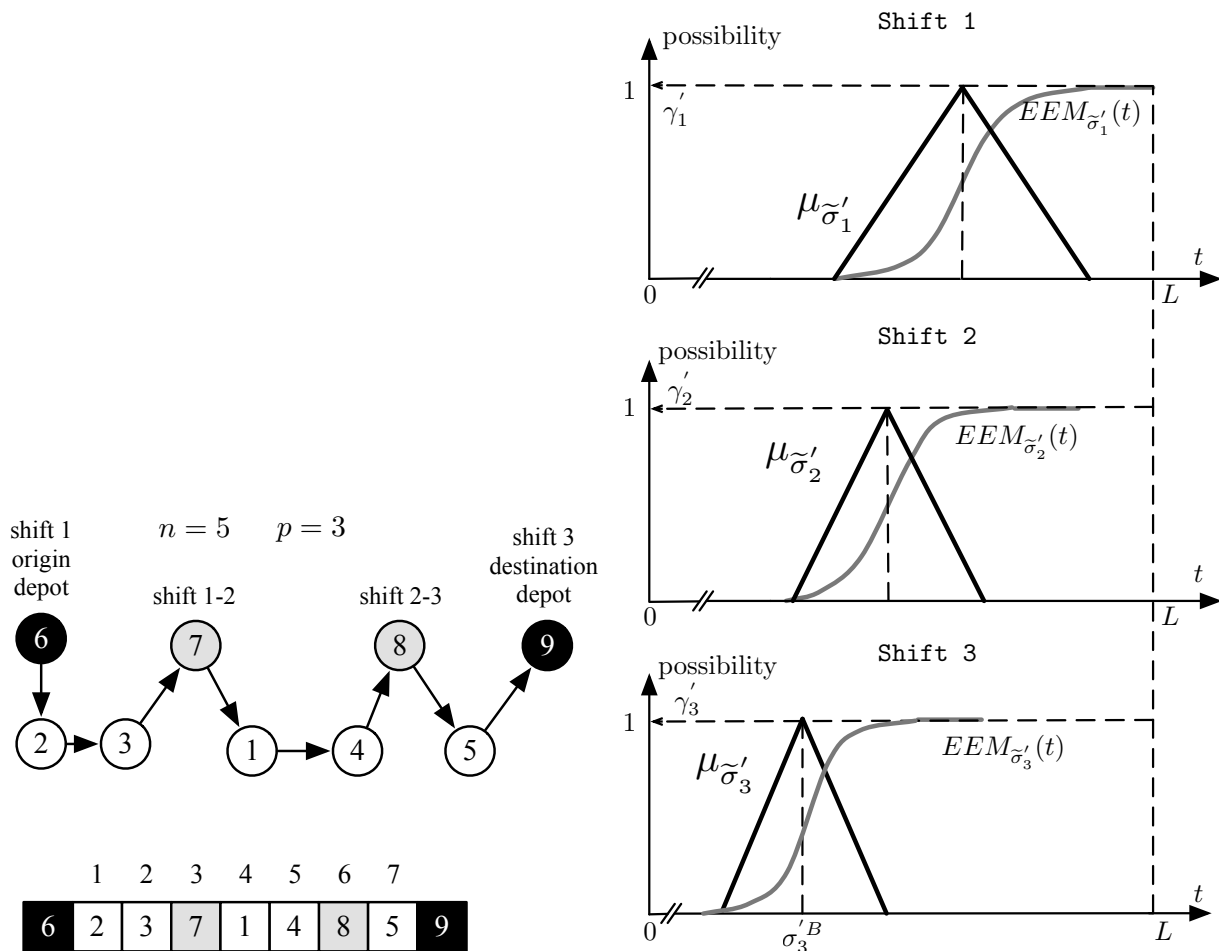


Figure 4: Solution Ψ' graph path, encoding and fuzzy shift durations

4 Numerical results

We assessed the performance of the approach described in Section 3.3 by generating different experiments. First, considering a standard non-fuzzy problem, we confronted the performance of the MIP reported in (1)-(13) in Section 2, with the AIA, that is Section 3.3 without steps 2*b*, 2*c*, and 4*d*. Second, introducing fuzzy uncertainty, we adopted the AIA reported in Section 3.3 considering innovative steps 2*b*, 2*c*, and 4*d*.

We developed an application on an x86 family 2.5 GHz Intel Core i7 processor having 4GB RAM and SSD, exploiting “C#.Net4” language. In the following, the AIA parameters and their criteria are defined. Population size $pop = 200$, number of generations $ng = 10000$, rate for Rule1 and Rule2 $pr1 = pr2 = 0.5$, number of clones $nc = 20$, mutation rate $mr = 0.75$, mutation number per generation $nm = 40$, exchangeable antibodies number $nea = 20$.

Three industrial test cases A, B, and C are adopted. In Table 3 the number of demand nodes n and the number of available shifts p is reported. Maximum shift duration is set to $L = 480$ min. Since industrial records are protected by a non-disclosure agreement, we report only summary data. The considered geographical is the Salento region in southeast Italy. Traveling and processing times are subject to uncertainty and are set considering user-experience values. Job processing times are equally distributed in three groups: small (around 15 min), medium (around 30 min) and large (around 40 min).

4.1 Non-fuzzy version results

Focusing on the non-fuzzy version of case studies A, B, and C, we compared the performance of the MIP model in (1)-(13) with AIA. Crisp values t^B are considered in place of the fuzzy version \tilde{t} . The non-fuzzy MIP model reported in (1)-(13) was solved with software package IBM CPLEX v.12.5 setting a time limit equal to the running time of the AIA application. In Table 4 we described the achieved results. In particular, for each case study and for each approach, we indicated the time (sec) when a new better solution is detected together with the corresponding solution optimal gap (%). In the first case study A, AIA calculates formerly the MIP optimal solution. For case study B, the optimal solution gap of AIA from MIP is 1% but AIA is tenfold faster than MIP. Finally, in the case study C, the optimal gap is 4%.

4.2 Fuzzy version results

Considering the fuzzy version of case studies A, B, and C, we adopted the fuzzy AIA to obtain the solution Pareto set. As reported in Section 3, multiple optimal solutions exist because of the 2-factor comparison. Company manager receives the optimal Pareto set and assesses the best solution. In Section 4.2.1, an example is provided. Assuming different AIA parameter values leads to different Pareto sets as shown in Section 4.2.2.

Our main objective consists in determining the AIA parameters to find as many Pareto set solutions. Consequently, we designed an experimental campaign to determine the best combination of AIA parameter values, see Section 4.2.3. Each experiment ϵ produces a solution Pareto set Γ_ϵ . Considering all the campaign experiments $\mathcal{E} = \{1, \dots, exp\}$ leads merging the corresponding Pareto sets $\bigcup_{\epsilon \in \mathcal{E}} \Gamma_\epsilon$. Union of Pareto sets is not a Pareto set, because a solution of a Pareto set can be dominated by a solution of another Pareto set experiment. We define $\bar{\Gamma}^\mathcal{E}$ as Hyper Pareto set for experiments $1, \dots, exp$ containing only the dominant solutions obtained from the experimental campaign. For each experiment ϵ , we denote the experiment impact ρ_ϵ as the number of solutions belonging both to the experiment Pareto set Γ_ϵ and the global Hyper Pareto set $\bar{\Gamma}^\mathcal{E}$.

Table 3: Case study definition

Case study	n	p
A	21	5
B	33	8
C	45	10

4.2.1 Single experiment results

Considering the fuzzy version of case study A, our software produces the Pareto set reported in Fig. 5. The collected solutions $s1, s2, \dots, s10$, belonging to the optimal Pareto set, are reported in Table 5. For such solutions, jobs are always completed in 3 shifts.

Feasibility degree for solution $s1$ is 1, because $\sigma_i^C < 480, \forall i = 1, 2, 3$. Solution $s1$ makespan is $2 \cdot 480 + 348 = 1308$ min. Since solution $s1$ is very conservative, the manager is unlikely to accept such a high safety margin.

Whereas, solution $s2$ has $f = 1263$ min and $\gamma = 0.968$. Indeed, in Table 5 we have $\sigma_1^C = 486 > 480$ and $\sigma_2^C = 499 > 480$; we note that 96.8% of the area of fuzzy number $\tilde{\sigma}_2 = (349; 424; 499)$ stays on the left side of 480 (see Fig. 1). The manager prefers solution $s2$ to $s1$ as assuming a low risk (3.2%) he/she decreases makespan by 45 minutes. Solutions $s3$ and $s4$ have similar feasibility degrees but $s4$ makespan is lower than $s3$. Whereas, solutions $s5, s6$, and $s7$ have similar makespan but $s5$ feasibility degree is better than others. Feasibility degrees for solutions $s8, s9$, and $s10$ are lower than 0.5: see $\sigma_2^B > 480$ min in last three rows in Table 5.

Referring to Fig. 5, manager declares that solutions $s2, s4$, and $s5$ are the most valuable and selects the best one based on his/her experience.

In Fig. 6, shift tours for solution $s1$ are graphically reported using ‘Google Maps’ website (maps.google.com). For privacy agreement, only the position of Lecce, Salento main city, is indicated. Depot position is red pinned. Job locations are indicated with circles. Basically, shift 1 tour accomplishes jobs in the north side. Shift 2 serves central zone and shift 3 satisfies south side. Each shift serves 7 jobs.

Table 4: Comparison AIA vs. MIP for non-fuzzy version of case studies A, B, C

Case study A				Case study B				Case study C			
AIA		MIP		AIA		MIP		AIA		MIP	
t [sec]	gap [%]	t [sec]	gap [%]	t [sec]	gap [%]	t [sec]	gap [%]	t [sec]	gap [%]	t [sec]	gap [%]
1	130	1	230	1	123	1	450	1	187	1	345
20	95	31	99	31	115	30	129	21	125	36	234
45	43	76	87	51	33	45	97	48	49	54	221
105	22	273	54	120	12	150	65	532	4	169	198
123	0	778	23	246	1	264	51			254	187
		1002	13			778	42			675	85
		1324	6			1042	21			1201	31
		1457	0			1256	14			1312	20
						2001	9			1416	12
						2398	2			2395	7
						2405	0			5998	2
										7194	0

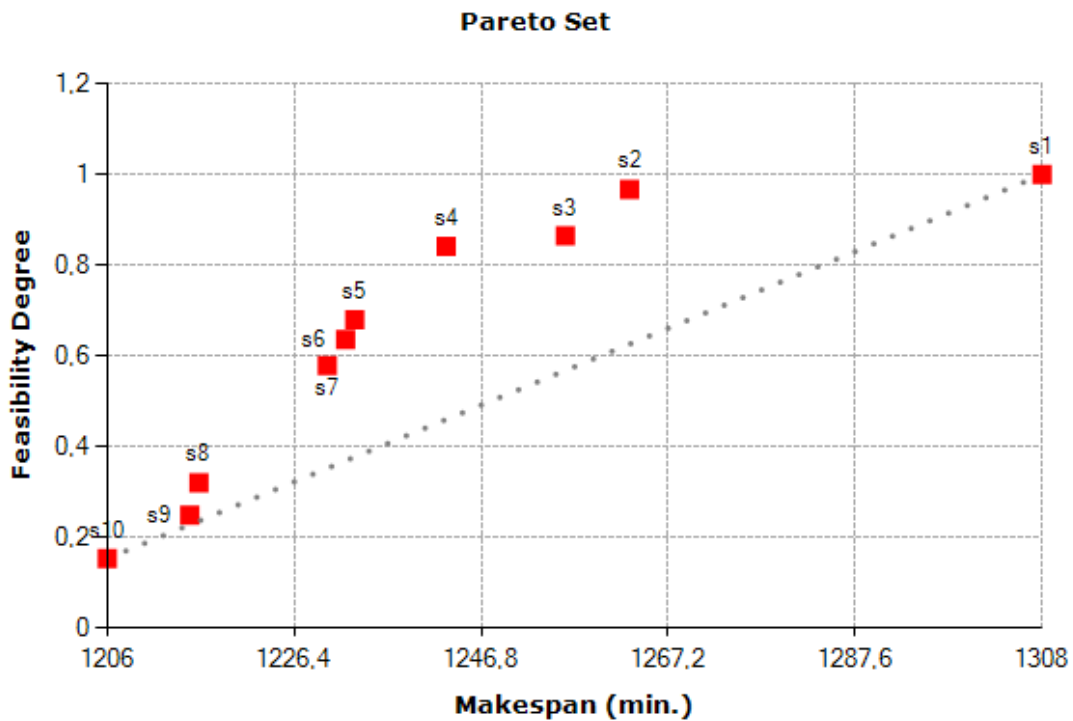


Figure 5: Final solution Pareto set for fuzzy case study A with basic parameters

Table 5: Details of final solutions for fuzzy case study A with basic parameters

Solution	Actual Fuzzy Shift Duration $\bar{\sigma}_i$		
	Shift 1	Shift 2	Shift 3
s1	(357;417;477)	(345;410;475)	(298;348;398)
s2	(366;426;486)	(349;424;499)	(263;303;343)
s3	(372;432;492)	(369;444;519)	(256;296;336)
s4	(390;450;510)	(365;445;525)	(248;283;318)
s5	(402;467;532)	(369;444;519)	(238;273;308)
s6	(402;467;532)	(394;469;544)	(237;272;307)
s7	(400;460;520)	(388;473;558)	(240;270;300)
s8	(400;460;520)	(412;497;582)	(226;256;286)
s9	(418;473;528)	(420;505;590)	(220;255;290)
s10	(371;426;481)	(433;518;603)	(211;246;281)

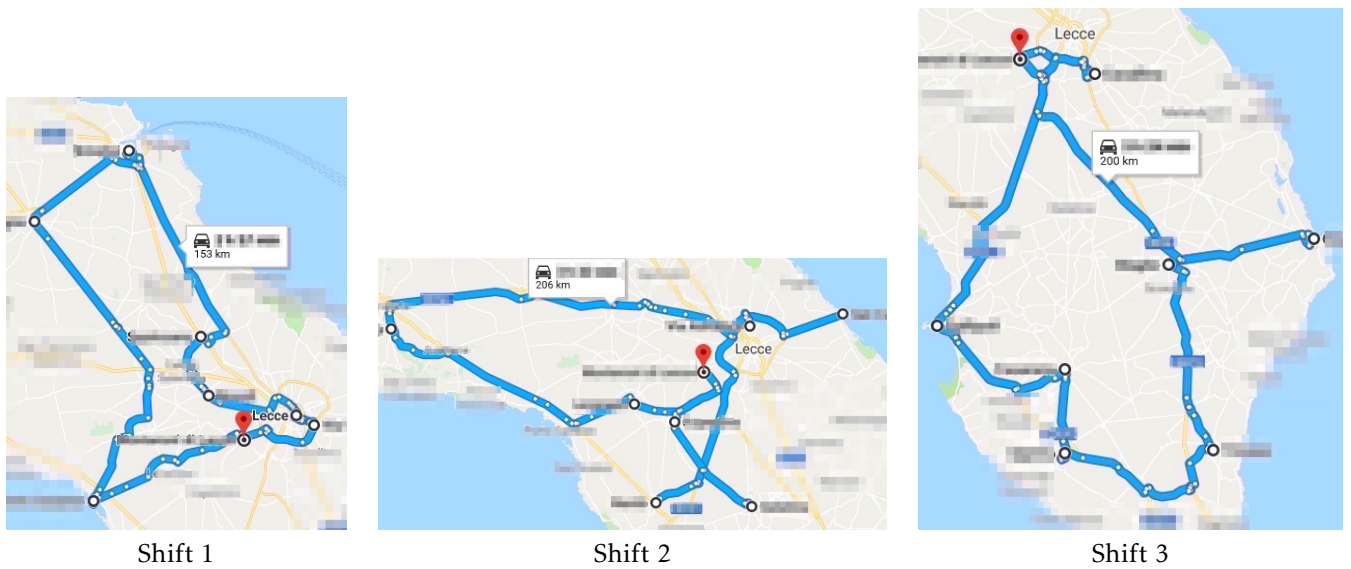


Figure 6: Shift tours for the solution s1

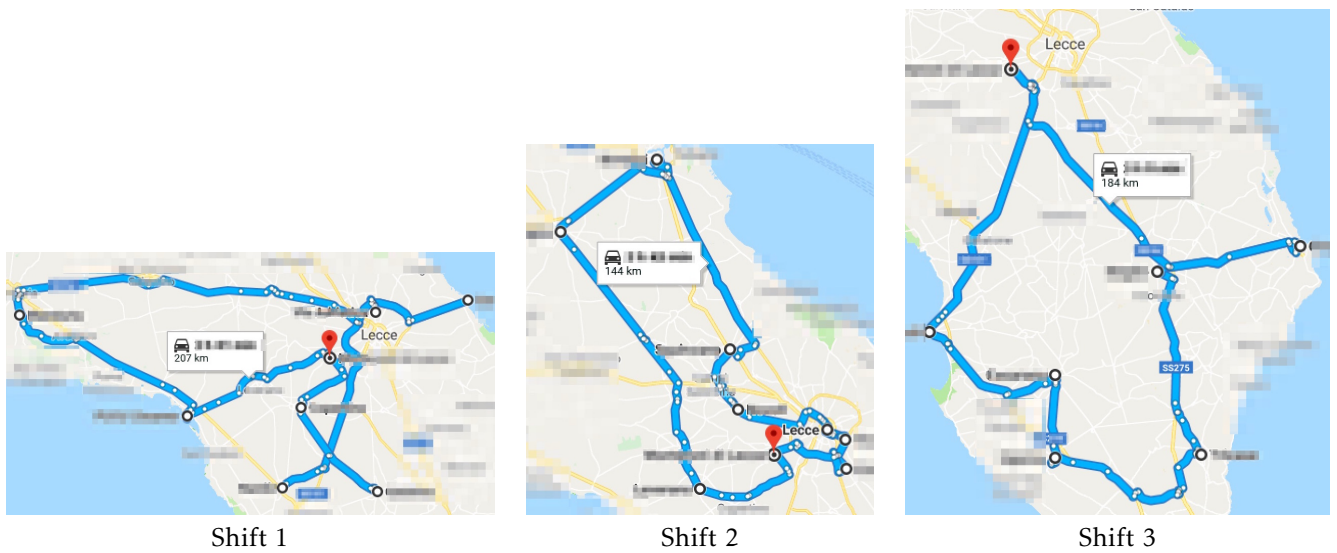


Figure 7: Shift tours for the solution s2

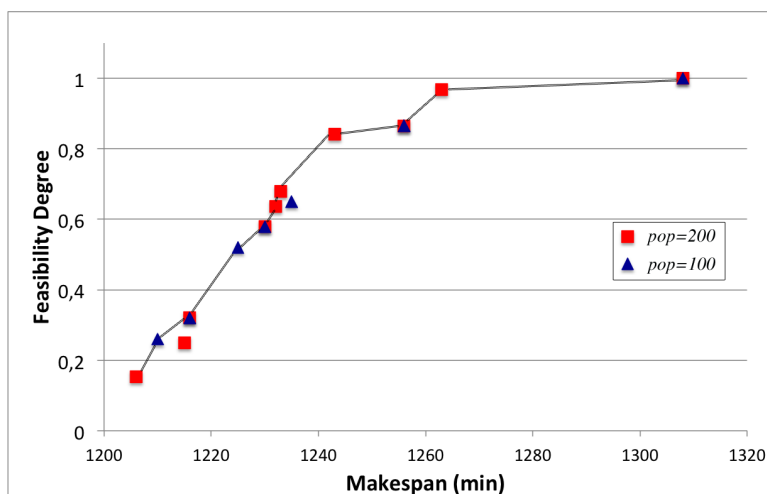


Figure 8: Final solution Pareto set for fuzzy case study A with two experiments

Table 6: Experiment impact results

(a) Case study A					(b) Case study B					(c) Case study C				
<i>exp</i>	<i>ng</i>	<i>pop</i>	<i>pr1</i>	ρ	<i>exp</i>	<i>ng</i>	<i>pop</i>	<i>pr1</i>	ρ	<i>exp</i>	<i>ng</i>	<i>pop</i>	<i>pr1</i>	ρ
1	5000	100	0.25	5	1	5000	100	0.25	7	1	5000	100	0.25	9
2	5000	100	0.50	5	2	5000	100	0.50	7	2	5000	100	0.50	9
3	5000	100	0.75	5	3	5000	100	0.75	7	3	5000	100	0.75	9
4	5000	200	0.25	7	4	5000	200	0.25	8	4	5000	200	0.25	9
5	5000	200	0.50	8	5	5000	200	0.50	11	5	5000	200	0.50	14
6	5000	200	0.75	6	6	5000	200	0.75	8	6	5000	200	0.75	11
7	5000	300	0.25	7	7	5000	300	0.25	9	7	5000	300	0.25	12
8	5000	300	0.50	8	8	5000	300	0.50	11	8	5000	300	0.50	15
9	5000	300	0.75	6	9	5000	300	0.75	8	9	5000	300	0.75	11
10	10000	100	0.25	5	10	10000	100	0.25	7	10	10000	100	0.25	10
11	10000	100	0.50	6	11	10000	100	0.50	7	11	10000	100	0.50	9
12	10000	100	0.75	5	12	10000	100	0.75	7	12	10000	100	0.75	10
13	10000	200	0.25	8	13	10000	200	0.25	10	13	10000	200	0.25	13
14	10000	200	0.50	9	14	10000	200	0.50	12	14	10000	200	0.50	16
15	10000	200	0.75	8	15	10000	200	0.75	11	15	10000	200	0.75	16
16	10000	300	0.25	8	16	10000	300	0.25	11	16	10000	300	0.25	15
17	10000	300	0.50	9	17	10000	300	0.50	12	17	10000	300	0.50	16
18	10000	300	0.75	9	18	10000	300	0.75	12	18	10000	300	0.75	16
19	15000	100	0.25	5	19	15000	100	0.25	7	19	15000	100	0.25	10
20	15000	100	0.50	6	20	15000	100	0.50	8	20	15000	100	0.50	11
21	15000	100	0.75	5	21	15000	100	0.75	7	21	15000	100	0.75	10
22	15000	200	0.25	8	22	15000	200	0.25	11	22	15000	200	0.25	15
23	15000	200	0.50	9	23	15000	200	0.50	13	23	15000	200	0.50	18
24	15000	200	0.75	9	24	15000	200	0.75	12	24	15000	200	0.75	16
25	15000	300	0.25	8	25	15000	300	0.25	11	25	15000	300	0.25	15
26	15000	300	0.50	9	26	15000	300	0.50	13	26	15000	300	0.50	18
27	15000	300	0.75	9	27	15000	300	0.75	14	27	15000	300	0.75	17

In Fig. 7, tours for solution s_2 are reported. The comparison of solution s_2 with s_1 is described in the following (see Fig. 6 and 7). Shift 1 tour lies in the central area whereas shift 2 concerns the north zone. In s_2 , to reduce makespan, one job is removed from shift 3 and added to shift 2. A job swap is performed between shift 1 and 2 in order to limit the shift 2 duration.

4.2.2 Two-experiment results

We compare the experiment performed in previous section 4.2.1 ($\epsilon = 1$), with a new experiment ($\epsilon = 2$) in which pop parameter is decreased from 200 to 100. Consequently, considering the experimental campaign set $\mathcal{E} = \{1, 2\}$, we report the corresponding solution Pareto sets Γ_1 ($pop = 200$) and Γ_2 ($pop = 100$) in Fig. 8. Set Γ_1 , in red, is reported in Fig. 5 too. Set Γ_2 contains 7 solutions, in particular: 4 solutions belong to Γ_1 too, 2 solutions are dominant solutions not included in Γ_1 and one solution is dominated by another in Γ_1 . Instead, one solution in Γ_1 is dominated by a solution in Γ_2 . Since, the new experiment ($\epsilon = 2$) produces two additional solutions in Pareto set and removes one, we have $|\bar{\Gamma}^{\mathcal{E}}| = 11$. In Fig. 8, the line represents Pareto front. Calculating $\rho_1 = 9$ and $\rho_2 = 6$, we can assess that first experiment has a greater impact than second.

4.2.3 Experimental campaign results

Considering the case study described in Section 4, we designed an experimental campaign. We examined 3 parameters ng , pop , and $pr1$ on 3 different levels: $ng \in \{5000, 10000, 15000\}$, $pop \in \{100, 200, 300\}$, and $pr1 \in \{0.25, 0.50, 0.75\}$. Consequently, for each case study, we have 27 experiments, $\mathcal{E} = \{1, \dots, 27\}$. For each experiment $\epsilon \in \mathcal{E}$, we calculate the corresponding impact ρ_{ϵ} . In Table 6a, experiment impact values are reported for case study A. As parameter pop increases, we obtain a larger number of Pareto set solutions. Increasing the parameter ng from 5000 to 10000 succeeds in increasing ρ , the same cannot be said when passing from 10000 to 15000. Parameter $pr1$ has a positive influence when $pr1 = 0.50$.

In Table 6b and Table 6c, experiment impact values are reported for case study B and C. The results confirm the findings of case study A. In conclusion, our solution Pareto set approach significantly depends on population parameter pop . Indeed, in each generation, we preserve the 'entire' solution Pareto set, so it is important having a larger pop value than the non-fuzzy approach. Parameter $pr1 = pr2 = 0.50$ represents a good compromise between generating Rule1 full random solutions and Rule2 solutions (Section 3.3).

5 Conclusion

This study presents some real insights into the single-vehicle routing problem with multi-shift and fuzzy uncertainty. Our objective consists of minimizing both the system makespan and shift overtime occurrence. In particular, we investigated the effect of uncertainty in driving and job processing time. We provide optimal solutions for the decision-maker considering a 2-factor comparison: the objective function value (makespan) and the degree with which overtime is avoided. Our approach is currently used by the case study company.

References

- [1] S. Zangeneh-Khamooshi, Z. B. Zabinsky, and J. A. Heim, "A multi-shift vehicle routing problem with windows and cycle times," *Optimization Letters*, vol. 7, no. 6, pp. 1215–1225, aug 2013.
- [2] Y. Ren, M. Dessouky, and F. Ordóñez, "The multi-shift vehicle routing problem with overtime," *Computers & Operations Research*, vol. 37, no. 11, pp. 1987–1998, 2010.
- [3] G. Onder, I. Kara, and T. Derya, "New integer programming formulation for multiple traveling repairmen problem," *Transportation Research Procedia*, vol. 22, pp. 355–361, 2017. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S2352146517301771>
- [4] S. Nucamendi-Guillén, I. Martínez-Salazar, F. Angel-Bello, and J. M. Moreno-Vega, "A mixed integer formulation and an efficient metaheuristic procedure for the k-Travelling Repairmen Problem," *Journal of the Operational Research Society*, vol. 67, no. 8, pp. 1121–1134, aug 2016. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1057/jors.2015.113>
- [5] K. Sparks, C. L. Cooper, Y. Fried, and A. Shirom, "The Effects of Working Hours on Health: A Meta-Analytic Review," in *From Stress to Wellbeing Volume 1*. London: Palgrave Macmillan UK, 2013, pp. 292–314.
- [6] C. C. Caruso, "Negative Impacts of Shiftwork and Long Work Hours," *Rehabilitation Nursing*, vol. 39, no. 1, pp. 16–25, jan 2014.
- [7] G. Costa, "Shift Work and Health: Current Problems and Preventive Actions," *Safety and Health at Work*, vol. 1, no. 2, pp. 112–123, 2010.
- [8] T. M. Cook and R. A. Russell, "A Simulation and Statistical Analysis of Stochastic Vehicle Routing with Timing Constraints," *Decision Sciences*, vol. 9, no. 4, pp. 673–687, oct 1978.
- [9] M. Gendreau, G. Laporte, and R. Séguin, "Stochastic vehicle routing," *European Journal of Operational Research*, vol. 88, no. 1, pp. 3–12, jan 1996.
- [10] G. Yaohuang, X. Binglei, and G. Qiang, "Overview of Stochastic Vehicle Routing Problems," *Journal of Southwest Jiaotong University*, vol. 10, no. 2, pp. 113–121, 2002.
- [11] C. Lee, K. Lee, and S. Park, "Robust vehicle routing problem with deadlines and travel time/demand uncertainty," *Journal of the Operational Research Society*, vol. 63, no. 9, pp. 1294–1306, sep 2012.
- [12] J. Xu, F. Yan, and S. Li, "Vehicle routing optimization with soft time windows in a fuzzy random environment," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 6, pp. 1075–1091, 2011.
- [13] H. Ewbank, P. Wanke, and A. Hadi-Vencheh, "An unsupervised fuzzy clustering approach to the capacitated vehicle routing problem," *Neural Computing and Applications*, vol. 27, no. 4, pp. 857–867, may 2016.
- [14] Z. Zhu, J. Xiao, S. He, Z. Ji, and Y. Sun, "A multi-objective memetic algorithm based on locality-sensitive hashing for one-to-many-to-one dynamic pickup-and-delivery problem," *Information Sciences*, vol. 329, no. C, pp. 73–89, feb 2016.
- [15] A. G. Novaes, E. T. Bez, P. J. Burin, and D. P. Aragão, "Dynamic milk-run OEM operations in over-congested traffic conditions," *Computers & Industrial Engineering*, vol. 88, no. C, pp. 326–340, oct 2015.
- [16] H. Zhao, W. A. Xu, and R. Jiang, "The Memetic algorithm for the optimization of urban transit network," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3760–3773, may 2015.
- [17] D. Muñoz-Carpintero, D. Sáez, C. E. Cortés, and A. Núñez, "A Methodology Based on Evolutionary Algorithms to Solve a Dynamic Pickup and Delivery Problem Under a Hybrid Predictive Control Approach," *Transportation Science*, vol. 49, no. 2, pp. 239–253, may 2015.
- [18] S. F. Ghannadpour, S. Noori, R. Tavakkoli-Moghaddam, and K. Ghoseiri, "A multi-objective dynamic vehicle routing problem with fuzzy time windows: Model, solution and application," *Applied Soft Computing*, vol. 14, pp. 504–527, jan 2014.
- [19] D. Sáez, C. E. Cortés, and A. Núñez, "Hybrid adaptive predictive control for the multi-vehicle dynamic pick-up and delivery problem based on genetic algorithms and fuzzy clustering," *Computers & Operations Research*, vol. 35, no. 11, pp. 3412–3438, nov 2008.
- [20] L. de C.T. Gomes and F. J. Von Zuben, "Multiple criteria optimization based on unsupervised learning and fuzzy inference applied to the vehicle routing problem," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 13, no. 2-4, pp. 143–154, 2002.
- [21] Y. He and J. Xu, "A class of random fuzzy programming model and its application to vehicle routing problem," *World Journal of Modelling and Simulation*, 2005.
- [22] D. Teodorović and G. Pavković, "The fuzzy set theory approach to the vehicle routing problem when demand at nodes is uncertain," *Fuzzy Sets and Systems*, vol. 82, no. 3, pp. 307–317, sep 1996.
- [23] K. Tan and K. Tang, "Vehicle dispatching system based on Taguchi-tuned fuzzy rules," *European Journal of Operational Research*, vol. 128, no. 3, pp. 545–557, 2001.
- [24] M. Jiménez, M. Arenas, A. Bilbao, and M. V. Rodríguez, "Linear programming with fuzzy parameters: An interactive method resolution," *European Journal of Operational Research*, vol. 177, no. 3, pp. 1599–1609, mar 2007.
- [25] H.-J. Zimmermann, M. A. E. Kassem, N. M. El-Badry, and A. Bilbao, "Fuzzy mathematical programming," *Computers & Operations Research*, vol. 10, no. 4, pp. 291–298, jan 1983.
- [26] M. Jiménez, "Ranking Fuzzy Numbers Through The Comparison Of Its Expected Intervals," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 04, no. 04, pp. 379–388, aug 1996.
- [27] H. J. J. Kals, International Design Seminar 1999 Enschede, and International Institution for Production Engineering Research, *Integration of process knowledge into design support systems : proceedings of the 1999 CIRP International Design Seminar, University of Twente, Enschede, The Netherlands, 24 - 26 March 1999*. Kluwer, 1999.
- [28] Baoding Liu and Yian-Kui Liu, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 4, pp. 445–450, aug 2002.
- [29] Quan Nguyen and Tu Van Le, "Fuzzy discrete-event simulation with FTipLog," in *1996 Australian New Zealand Conference on Intelligent Information Systems. Proceedings. ANZIIS 96*. IEEE, pp. 195–198.
- [30] F. Nucci, "The multi-shift single-vehicle routing problem under fuzzy uncertainty," in *2017 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, sep 2017, pp. 156–161. [Online]. Available: <http://ieeexplore.ieee.org/document/8120987/>
- [31] M. Mobini, Z. Mobini, and M. Rabbani, "An Artificial Immune Algorithm for the project scheduling problem under resource constraints," *Applied Soft Computing*, vol. 11, no. 2, pp. 1975–1982, mar 2011.

- [32] A. Bagheri, M. Zandieh, I. Mahdavi, and M. Yazdani, "An artificial immune algorithm for the flexible job-shop scheduling problem," *Future Generation Computer Systems*, vol. 26, no. 4, pp. 533–541, apr 2010.
- [33] J. Al-Enezi, M. Abbod, and S. Alsharhan, "Artificial Immune Systems - models, algorithms and applications," *International Journal of Research and Reviews in Applied Sciences*, vol. 3, no. 2, pp. 118–131, 2010.

Holistic Access Control and Privacy Infrastructure in Distributed Environment

Uche Magnus Mbanaso^{1,*}, Gloria A Chukwudebe²

¹Centre for Cyberspace Studies, Nasarawa State University, Keffi, Nigeria

²Department of Electrical & Electronic Eng., Federal University of Technology Owerri, Nigeria

ARTICLE INFO

Article history:

Received: 18 August, 2018

Accepted: 27 October, 2018

Online: 01 November, 2018

Keywords:

Internet

Internet of Things

Cyber- physical system

Digital trust,

Confidentiality,

Privacy

Access Control

Distributed Environment

ABSTRACT

This article discusses IoT security in situations whereby devices do not share the same security domains, which raises security, privacy and safety concerns. It then presents an Access Control and Privacy infrastructure for addressing these concerns in the context of distributed environments. IoT deployments allow billions of connected physical devices to collect, process and share data; collaborate and cooperate in automating tasks in an unrivaled fashion. However, security and safety are still top major fears that demand holistic approach, particularly when devices do not share the same digital trust. This is not a surprise, as a revolutionary system, IoT comes with inherent vulnerabilities, threats and risks like most other computing and data processing systems. Conversely, when security breaches or compromises occur, it is most likely to have a far-reaching and upsetting consequences that extends traditional concerns. The fact that IoT can be deployed in plethora of application scenarios; means that end-to-end security should be treated contextually and in a dynamic manner. Consequently, these concerns; trust, confidentiality, and privacy at the IoT application stack need to be addressed robustly. Thus, in this article, a novel distributed access control infrastructure based on configurable policy constructs is presented. The infrastructure provides a mechanism for gradual negotiated release of provable attributes to dynamically build trust before protected resources are made available. In this configuration, IoT transaction parties can express their Capabilities (competences, features, etc.) and Requirements (rules and provable attributes required to access the capabilities) as the basis for sharing data or collaboration in solving business problems.

1. Introduction

This paper is an extension of work originally presented during the 13th International Conference on Electronics, Computer and Computation (ICECCO) in 2017 [1]. The deployment of Internet of Things (IoT) is developing in many areas and contexts. Its deployment spans across diverse spaces and is anticipated to continue to extend beyond present expectations [2]. In some deployments, the range of IoT devices that may work together or share data are unlikely to belong to a single (or the same autonomous) security domain [3]. By security domain, we basically mean a collection of connected entities or applications that are part of a specific digital trust infrastructure (or administered by common cryptographic policy), i.e. Public Key Infrastructure (PKI) security arrangement for authentication, authorization and session management. Invariably, when devices,

which are members of different security domains want to collaborate in providing business solutions, it raises trust, confidentiality and privacy challenges in a variety of application contexts. This demands fresh security requirements as threats posed by cyber, physical and human factors span beyond traditional risk landscapes. However, trust, confidentiality and privacy have received substantial attention in the literature in different contexts [4][5][6][3][7].

Notwithstanding, IoT has different set of physical and virtual (or logical) fresh crop of security issues that varies, and are contextually multifaceted. That is, IoT security, privacy and safety may be seen from a variety of surfaces including those specific to the device, cloud computing, mobile apps, network interfaces, software, physical and access control. While a number of these security areas, can be addressed specifically by vendors and/or manufacturers, yet some have to be addressed in application context and real-time, and cannot be on the basis of 'one solution fits all'. Arguably, it requires security infrastructure

*Uche Mbanaso, Centre for Cyberspace Studies, Nasarawa State University, Keffi, Email: ozara.oru@gmail.com

www.astesj.com

<https://dx.doi.org/10.25046/aj030604>

that is adaptable and flexible, taking into consideration the business or solution environments. It therefore aptly suggests that IoT operational environments can be highly contestable with several attack opportunities for intruders. This raises strong digital trustworthiness as germane for treating identity and access management in emerging smart environments. The assurance of how data is collected, processed and shared, entails an obligation to mutually respect contractual data access agreements that met security principles and privacy.

A typical example can buttress our point. A smart vehicle arriving a city would like to request certain city-based data from available connected city IoT systems [2]. But a secure conversation will demand that these city devices cannot wittingly disclose data without first ascertaining the trustworthiness of such third party entity. It is assumed here that the smart vehicle belongs to another security domain and has no existing digital trust affiliation with the city's security domain[8]. For security and safety purpose, both parties should exchange information based on the ability to trust each other[9][10]. It implies that access and data security as well as safety of these connected entities must be reciprocally assured. Inversely, the smart vehicle may equally not be ready to disclose its profile to the city systems straight away, and the city systems cannot assume that the smart vehicle's mission is harmless, and thus share data with the vehicle or conduct operations together. In either direction, both parties have security, privacy and safety issues of concerns. In this regard, critical challenges can be inferred as follows:

- Unauthorised activities of hostile entities to compromise security and safety of IoT;
- Activities of friendly parties to disregard mutually contractual agreements to violate security and safety of devices, resources or underpinning infrastructure.

To this extent, IoT security and safety landscapes are still evolving issues constrained by encumbrances that have both socio-economic and security impetus. Furthermore, privacy and trust are subjective to cultural perceptions with unpredictable degree of individual capacity and expectations [11]. Unlike trust that is by no way regulated, privacy rights are subjective to some sort of regulations, especially by legislation, security principles, procedures, ethics, etc. in a variety of countries [12][13]. Aside, it is expected that a blanket access cannot be allowed to typical IoT systems, particularly for safety reasons; besides the need to secure sensitive resources and/or attributes. This, uniquely makes a further strong case for a symmetric security infrastructure at application stack that supports fine-grained policy in decision-making and authorization.

Therefore, it is obvious that IoTs are likely to operate in a variety of security domains and without pre-established digital trust but may have to share data, and where possible work together to address common business issues but in a secure and trusted manner. Thus, it is incumbent that identity and access management at application stack are critical requirement for treating security, privacy and safety. In this light, we present a bilateral symmetric and configurable policy-based infrastructure to address this critical application layer security in IoT distributed systems. This infrastructure uses Obligation of Trust (OoT) protocols[3] that allows reciprocated interexchange of policy constructs described as *Requirements* and *Capabilities* to gradually establish dynamic trust before making available

protected information or performing some mutual tasks. Traditional solutions assume a form of digital trust based on simple use of username/password pair, which is highly susceptible to a variety of threats [14][15][16]. This is not ideal in many of IoT solution space, particularly in distributed environments.

Our solution is novel in many respects. First, it offers a real-time mutual treatment of security and privacy using configurable policy constructs that permits both parties in transaction to reciprocally take access decision based on their individual security requirements and capabilities. Second, it is a departure from one-way protection perception whereby only the security concerns of the party providing services is considered. Third, it is a highly scalable access control mechanism that has the capability to deal with present and future threats through robust and extensible rule constraints. Fourth, by addressing trustworthiness in privacy protection in a unified fashion, the infrastructure provides mechanisms for accountability, trust-based digital evidence as basis for dispute resolution, which is a critical requirement for IoT security and safety.

The rest of the paper is organized as follows: Section II reviews related works while Section III presents threat analysis and challenges in the context of privacy, trust and confidentiality. Section IV describes the novel Distributed Access Control Infrastructure while Section V presents discussions on the novel infrastructure. Section VI concludes the paper.

2. Review of Related Works

IoT requires a holistic approach to solve security, privacy and safety concerns in a particular security layer. This may entail combining technical, procedural and legal controls to minimize the severity of risks associated with access and availability of protected data as well as intellectual or proprietary property [1][3]. IoT operations take place at application stack where systems collect, store, analyze and share data. In some cases, sensitive attributes of service requesting parties are required to perform authorization. Undisputedly, privacy is most often considered from simple legal statements without automation of enforceable technical measures. Meanwhile IoT application level security is similar to those faced by other computing application space, existing identity and access control models can be adapted to suit IoT environments. However, the complexity is that an autonomous security domain may have hundreds or even thousands of connected objects with sensors and actuators to manage. This is the differentiator, which makes IoT risk landscape differ significantly. To this extent, the challenge before us is how to extend and adapt existing models and controls to address numerous IoT security issues. In the section that follows, security models and standards that influenced our solution are reviewed.

2.1. IoT Reference Model

The increasingly broad adoption of IoT devices span wide area of use cases across multiple business domains including smart cities, smart manufacturing, smart agro, smart parks, smart hospitals, smart patient supported living solutions, etc.[17].

In smart cities, for instance, IoT sensors can focus on sensing the environment on some crowded areas. For instance, sensors can be used to ascertain air quality among others in an effort to monitor particular densely inhabited city zones periodically [18]. However,

sensor data can be spoofed or can become attack vector to facilitate a particular threat. In this context, understanding the tenets of security, privacy and safety issues is incumbent to the different IoT security layers. Thus, the IoT reference models create a common understanding of operational layers, features and functionality of IoT, which can help in insightful conceptualisation security architecture [18]. More importantly, it is instructive to note that no single security layer is a complete solution. However, there are plethora of IoT reference architectures, which helped to conceptualise IoT security[18][14][19].

2.2. Federated Identity Management (FIM)

Simply, the Federated Identity Management (FIM) is an infrastructure model used to associate identity information and attributes of entities across trusted several security domains [20]. The approach provides a mechanism for “single sign-on” in a fashion that allows transaction parties to obtain trusted access tokens from their local Identity Provider in order to be allowed access to outside services in a confederated manner [10][21]. An example of the FIM is the OpenID [22]. Usually, FIM is a classical transient trust built by using username/password pair to authenticate to a party’s local Identity Provider (IdP) while this IdP issues and communicates signed access control assertions or tokens to the service providing party.

In some deployments, Attribute Authority (AA) is an integral part of FIM developed to provide a much more resilient access control engine as opposed to simple authentication provisions [23][24]. Typically, AA is simply, a trusted repository for secure storage of attributes/properties of parties commonly used in Attribute-Based Access Control (ABAC) infrastructure [20]. In some use cases, FIM attempts to distinguish authentication operations from authorization process on the basis of separation of security duties.

Although FIM offers user convenience and efficiency in managing identity provisions, users and relying parties, the use of username/password pair makes it defenseless against numerous threats. More so, issuance of access tokens is not on itself sufficient to guarantee the behaviour of a transaction entity. However, managing vast IoT identities has been raised as also a contending issue due to the anticipated volume of IoTs. Thus, FIM is potentially well suited for managing IoT identities, and as an integral part of a distributed access control infrastructure.

2.3. eXtensible Access Control Markup Language (XACML)

The XACML is a standard access control policy construct developed by the Organization for the Advancement of Structured Information Standards (OASIS), that provides collective framework for specifying a range of access control rules [25]. It has its foundation from eXtensible Markup Language (XML), and presents an extensive access control structure and encoding schemes to describe fine-grained access control rules as well as message level request-response construct that allow constituent part to work together in distributed access control operations. It exemplifies a modular infrastructure that is loosely coupled based on functionality and application domain, in a manner that allows them to be hosted independently.

2.4. Security Assertion Markup Language (SAML)

SAML is a very powerful and extensible language based on XML scheme specifically developed for the exchange of access control information from one transaction party to another. Usually, an identity provider (a SAML issuer or SAML authority)

www.astesj.com

makes one or more assertion statements about a principal or entity in an opaque string, which is communicated to a consuming party, typically a service provider to grant access to the subject described on the digitally signed assertion [26][27]. The relying party decides to trust the SAML issuer based on some pre-existing trust relationship provided by digital certificate, which asserts that the subject is trustworthy.

From purely technical perspective, SAML assertion is the primary standard used by most single sign-on (SSO) schemes, even the FIM. The XML structure has an Issuer element that describes the SAML authority; the *Signature* section that holds the *signature block*, which encapsulate the PKI information of the issuer, algorithms and transforms as well as the resulting digital signature, etc. The *Subject* element encapsulates the identity of the subject; the *Condition* element describes obligatory conditions as an additional rule constraints. The *Assertion Statement* specifies the assertion context including authentication, attribute, authorization decision, or other user-defined constructs that can facilitate access control.

2.5. Obligation of Trust (OoT) Protocol

The Obligation of Trust Protocol (OoT) provides an innovative symmetric access control protocol as described in [7][28], which illustrates a bilateral and symmetric method that combines digital trust negotiation and access control operations for the treatment of security and privacy protections based on enforcement of mutual policy rules between two or more parties in distributed application environments. Ideally, the OoT protocol allows two or more transaction parties to interexchange policy constructs contained in *Requirements* and *Capabilities* in real-time. The OoT SAML request message described as a *Notification of Obligation* (NoB), first notifies the services requesting party the conditions for accessing its resources expressed as *Requirements* and its available services or features in *Capabilities*. The response message after execution of *Matching Algorithms* is the assurances that describes the fulfillment of each other’s conditions contained in the *Requirements* policy element. The response message is characteristically the *Signed Acceptance of Obligations* (SAO). The details of OoT access control protocol that demonstrates how parties in conversation can use SAML Obligation of Trust Assertion can be found in [3][7].

3. Threat Analysis and Challenges

Fundamentally, to understand operational IoT security and safety issues, all layers of IoT must be considered and thoroughly assessed. Thus, the five security goals i.e. confidentiality, integrity, availability, authenticity, and non-repudiation should form the basis to assess threats. Consequently, in assessing these threats, three classic IoT system threats are described as follows:

• A Target of an Attack

Conventionally, an IoT device is potentially exposed to many threats faced by a typical computing system, particularly at network and application layers. It implies that IoT can suffer data breach or the device can be degraded, which can result to violation of confidentiality (or privacy) and integrity, as well as denial of service (availability). Most IoT systems have in-built application servers that equally face the same security challenges as traditional web servers[29]. OWASP[30], described ten top categories of IoT vulnerabilities that can be exploited by a hostile

party. Thus, threats can materialize through evading authentication provisions due to weak configurations and associations to the extent that it is too difficult to repudiate (non-repudiation) the nefarious actions. The extent to which this can happen depends on the mission, capability and the motivation of a hostile party.

• **A Tool for an Attack**

The composition of an IoT device includes sensors and actuators, implying the potential to be manipulated intelligently to distribute nefarious programmes or become an integral part of a malicious network that can take part in Distributed Denial of Service (DDoS) attack to cause unavailability. Likewise, operationally, an IoT can simplify a variety of attacks as sensors and actuators can conveniently become attack vectors. A malicious party can leverage unpretentious IoT operations for illegitimate purposes. In this context, such attacks may include undercover use of IoT engine to perpetuate cybercrime, financial fraud or cyberwarfare.

• **Incidental to an Attack**

This type of threat becomes possible when an IoT ecosystem is indirectly involved in an attack (i.e. stealthily supports criminal activities such as when in itself it is used to store data for criminal activities). It infers that possibly, an IoT can expedite an attack to occur much quicker by leveraging its power or functionality, or operational processes, which can make an attack more challenging to detect and attribute thereby causing non-repudiation attack.

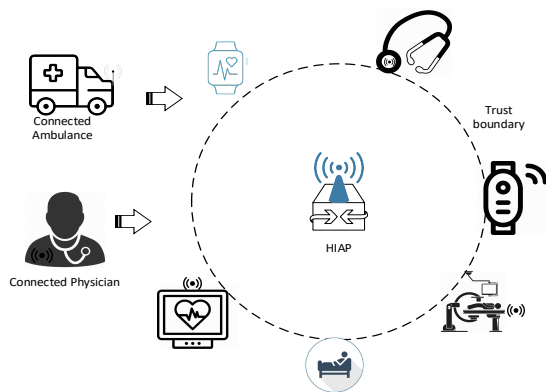


Figure 1: Typical IoT Use Case.

Already, cybercriminals have leveraged the inherent vulnerabilities in IoT engines to cause major disruptions, especially the Distributed Denial of Service (DDoS) attacks, which exploited Domain Name System (DNS) requests [31]. With the estimated 20 billion connected physical objects by 2020, and the explosion of industrial internet of things, recruiting thousands of connected devices to cause DDoS may be trivial. The foregoing typically suggests that a trusted party authorized to gain access to an IoT engine, can misuse it by employing the device to carry out other functions than originally programmed. For instance, GDPR[32] prescribes that personal data be used only for the initially stated purposes [13]. This provision makes privacy a contractual responsibility that must be respected by transacting IoT entities. Similarly, trust relationship is the requirement that attempts to guarantee the expected behaviour i.e.

the hope that an IoT entity will behave reciprocally and responsibly without impairment to the other party. Thus, in practice, this mutuality may be too difficult to achieve.

Potentially, IoT poses different set of threats and risks in diverse environments and contexts, which should be addressed dynamically and in perspective [4]. For instance, in healthcare scenario, a Physician may have the need to work in several hospitals, of which her digital trust is not provided by the same or single security domain. It implies that this kind of use case requires that a high level of trust be established, privacy to be guaranteed, and confidentiality to be kept as well as the assurance of accountability and non-repudiation [33]. In healthcare environment, diagnosis, monitoring and assessment of patients may require significant number of devices interconnected by Heterogeneous IoT Access Point (HIAP) to collaborate and cooperate to solve patient’s problems. In the same vein, the Physician’s IoTs without previous trust relationship, may be required to interact with other IoTs within the environment. In this scenario, for convenience sake, the Physician IoTs should discover and connect automatically to the same HIAP or gateway without recourse to manual configuration. Another simple real-world use case is a connected ambulance that brings a patient to a smart hospital environment, under the characteristics of the mission (or emergency), the ambulance should be able to discover and automatically connect to relevant devices to accomplish its mission without manual configurations.

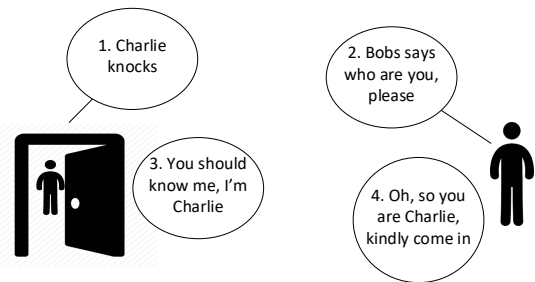


Figure 2: Trust Dialogue between Bob and Charlie.

Figure 1 depicts a typical smart healthcare environment. In many IoT systems, the manufacturers can provide a sort of security abstraction, which other security features can be derived, such features cannot by default solve application layer security that is usually contextual.

Like other computing devices, inbuilt security at abstraction layer cannot address application layer security and privacy out of the box, especially in distributed environments. In practical sense, security, privacy, and trust are not static security requirements. The implication is that these issues must be treated in context and instantaneously too. Moreover, IoT systems are probably going to expose services through Application Programmable Interface (API), this further reinforces the requirement for dynamic security, privacy and safety solutions that should be configurable[2]. To further provide insight to underlying concepts, we examine trust, privacy and confidentiality individually, and in perspective.

3.1. Trust Context

Building digital trust in typical IoT in distributed environment raises fresh security issues. In digital space, building trustworthiness is vital, and can then be built between physical

objects, physical objects and people, physical objects and systems as well as systems. Thus, trust is a critical factor in distributed IoT environments that must be examined holistically. Theoretically, a simple trust dialogue between Charlie and Bob can be used to demonstrate generally, the subtlety of trust as a concept, shown in Figure 2:

- Step 1: Charlie arrives at Bob’s door and knocks;
- Step 2: Bob says ‘who are you, please?’
- Step 3: Charlie answered, ‘you should know me, I’m Charlie’;
- Step 4: Bob says “oh, so you are Charlie, kindly come in. In this case Bob allowed Charlie because he seems to recognize the voice of Charlie and anticipated to see him.

Examining this simple trust dialogue, there is a potential that Bob can open the door and see an imposter (who imitated Charlie’s voice) instead of Charlie. This simplistic example, can be the basis to further discuss three important variables associated with trust namely: behaviour, reputation and expectations. Bob has merely trusted the statement based on known reputation and behaviour of Charlie with the anticipation that he will remain trustworthy. This modest example suggests there is inherent risk factors in the general concept of trust, thereby buttressing the point that current trust models provided by Public Key Infrastructure (PKI) are sufficient to guarantee trust in IoT environments. It further underscores the fact that providing security and safety features in IoT distributed systems require a sort of arrangement that gives the communicating parties to gradually establish more trust based on other attributes beyond PKI provisions.

In literature [34][23][35], digital trust is well researched, and provides the mechanism to verify and validate trust relationships, privileges, claims, identity attributes and information, etc.; giving the identity consuming party the opportunity whether to rely on the real-time assertions of proving party or not, based on the extended properties defined in the rule constraints.

3.2. Direct vs Indirect Trust

Traditionally, digital trust is simply based on either direct or indirect (or transitive) trust relationships. In a typical IoT system, access to resources can be granted based on verification and validation of pre-existing trust relationships that authenticates a party requesting a service. Generally, this can be referred to as a direct trust, a form of shared secret, such as username/password pair or digital certificates, etc.; which is usually created offline between parties prior to communications as depicted in Figure 3(a). Figure 3(b) illustrates the concept of indirect trust whereby a service provider requires to verify and validate the assertions made by a party requesting service but there is no existing digital trust relationship between them. Thus, for a secure conversation, a trusted intermediary must prevail to vouch for the requesting party in a manner that a relying party can trust its assertions. Such examples as practiced today include signing into other third party online applications using Facebook or Twitter accounts.

In highly sensitive and safety critical IoT applications, simple trust models as described above is flawed substantially, which can easily be fooled by malicious parties. Based on the analysis already presented, different application contexts in distributed IoT environments, will require full-proof digital trust built gradually by reciprocated negotiation of verifiable attributes to assure security, privacy and safety.

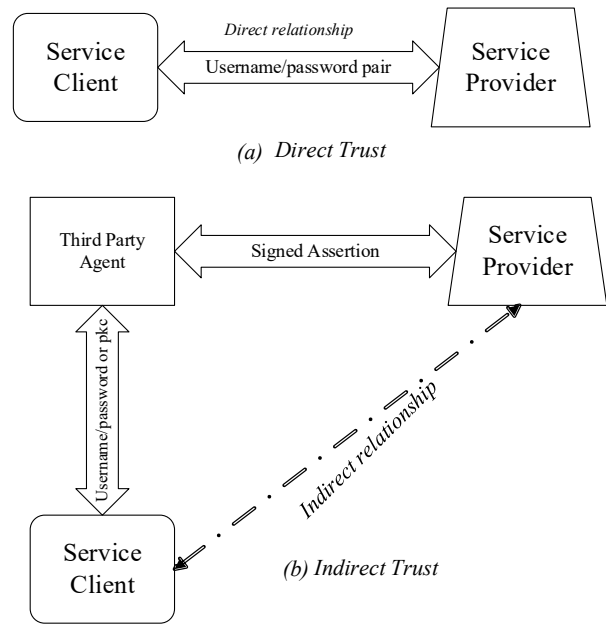


Figure 3: Classical Trust Models

3.3. Privacy and Confidentiality Context

In many instances, privacy and confidentiality still remain a misunderstood concepts. It is not surprising as the terms are closely related. Conversely, in distributed application environments, where two or more actors are involved, privacy and confidentiality operationally are difficult to guarantee. For example, during access control phase, data of privacy value may be shared between these actors i.e. from requesting party’s domain to the relying party, yet this privacy data may be disclosed by an intermediary party of another security domain. Data is fluid, and once shared with a third party, exercising full control thereafter becomes uncertain. While confidentiality can obviously be addressed by access restriction and/or encryption, privacy is subjective to trust expectations. So, it can be inferred that confidentiality is a means to ensure privacy protection especially when data is at rest but once the same data is passed unto a third party, then privacy may be eroded.

It is certain that personally identifiable information shared during transactions can be stored by either parties, to which the collector purportedly proclaims controls on behalf of the data owner. This raises privacy concerns, the data subject may lose track of the parties holding its data, and has no option but to rely on the facts of promise statements that the information will be given adequate privacy safeguards and protection. However, the new European Union General Data Protection Regulation (GDPR) has altered data privacy protection[13][36]. In the same wise, it may seem obviously that GDPR legal rules may have put stringent proscriptions but monitoring compliance and conformance is still operationally, a challenging task. Consequently, it suggests that in real-time, managing identity and access management, requires interacting parties to ensure that vouching for trustworthiness is cryptographically signed.

Above, entails that trust, privacy and confidentiality are strongly related and require a homogenous infrastructure to address them concurrently. To this extent, important questions can be raised to stimulate design assessment as follows:

- (i) How can transaction entities account for their actions when privacy attributes are compromised or breached?
- (ii) What are the technical mechanisms that can monitor how privacy data are accessed, shared and processed?
- (iii) What is the assurance that a party can keep privacy promises made to another party, support and safeguard proportionately by suitable operational means?
- (iv) Is there a technical mechanism to guarantee that transfer and processing of privacy information conforms to relevant standards and regulations, and its subsequent processing by second level third party?
- (iv) What are the mechanisms to handle conflicts and risks? Is there a valid channel to handle and resolve conflicts that supports strong digital evidence?
- (v) How can the liable parties be determined in multifaceted data breach involving several actors?
- (vi) Is there any difficult-to-repudiate digital evidence that is admissible in courts of law to support assertions in an event of disagreement?

Imperatively, these questions can form open issues that challenges the research community and the need to find optimal solutions to address complex security, privacy and safety posture of IoT threat and risk landscapes, especially in the wake of increasing value of data [36][32]. Furthermore, it is an acknowledged fact that technology alone cannot answer all of the questions hypothesized above. As a consequence, it is remarkable to state that suitable governance, regulation and compliance, conflict resolution and assurance mechanisms, are vital inputs, which buttress the point that there is a strong interplay between technology, policy and law in solving privacy equations.

Notwithstanding, robust technical infrastructure has significant role in responding to the above named issues. Technically speaking therefore, dealing with real-time security, privacy and safety of connected devices operationally, require a flexible and distributed infrastructure that supports configurable policy constructs to manage IoT risks based on informed and preferred decisions.

4. Distributed Access Control Infrastructure

To design applicable access control infrastructure for IoT in distributed systems requires thorough examination of the various actors in a typical IoT conversations. As illustrated in Figure 4, there are likely to be multiple actors from different security domains that can interoperate in classical IoT service deployments. This is assumed on the ground that one security provider may be unsuitable for identity and access control to authenticate and validate security assertions that can be trusted across some high profile IoT distributed environments [12].

For instance, access to IoT systems in classical medical environment may require personal attribute of Medical Consultants from the Medical Council as well as a referrer attributes from a City Council as the basis to share or disclose resources. Equally, in a smart city, a rule requirement may entail that for vehicles to interact with city cameras for example, the vehicle license plate number as well as insurance certificate may need to be authenticated before access is allowed to protected resources. In this typical case, the attribute providers may unlikely be part of a single security domain. It implies that in distributed application scenarios, IoT access requires a robust and scalable infrastructure to treat security, privacy, and safety dynamically and in trusted fashion.

Figure 4 clearly illustrates conceptual view of access control entities in distributed environments. It shows entities and the various responsibilities as well as data flows. It is widely acknowledged that IoT is resource constrained for now; buttressing the fact that process consuming access control operations such as evaluation of access control policies may not be executed within IoT system. As such, light weight access control operations e.g. such as enforcement of decisions can be carried out in IoT systems, while other functions be delegated to trusted external parties.

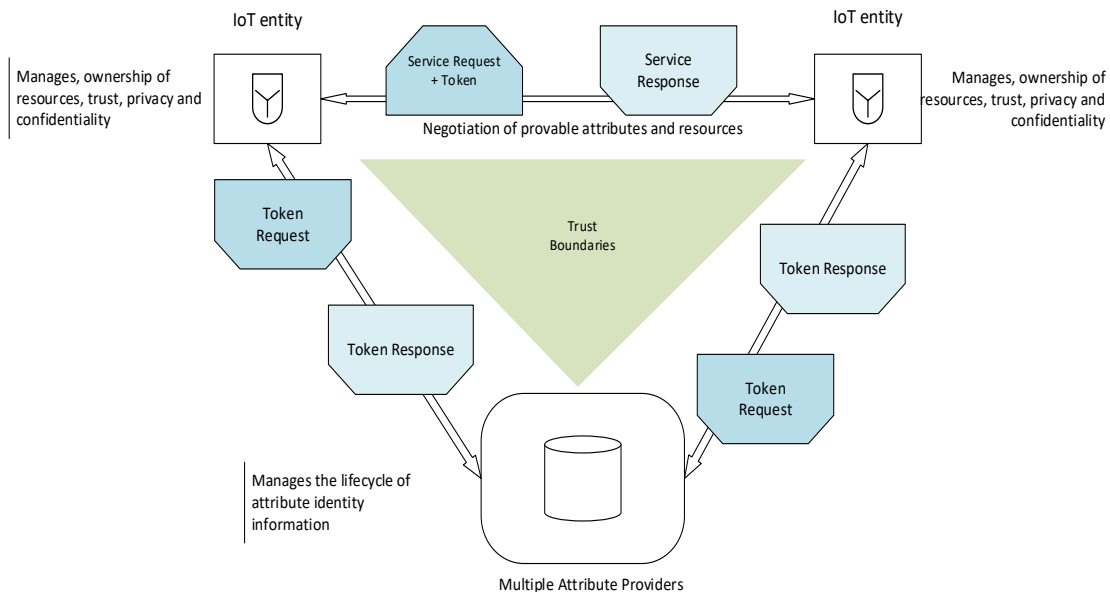


Figure 4: Conceptual View of Access Control Entities in a Distributed Environment.

4.1. Distributed Access Control Infrastructure for IoT

Figure 5 depicts Distributed Access Control Infrastructure that integrates components of FIM, Identity/Attribute Authority (IAA), and Obligation of Trust (OoT). The infrastructure components are loosely coupled in a distributed manner to allow interoperability and flexibility in deployment due to resource constrained IoT environment. The infrastructure consists of three logical subunits grouped according to areas or separation of concerns. The gatekeeper is tightly coupled to IoT application stack, which comprises Context Handler (CH) and Policy Enforcement Point (PEP) components of XACML. These components can programmatically be part of IoT system through its web service interface. The CH formulates or interprets specific application context data in a required format during conversations. Similarly, the PEP engine is responsible for the enforcement of access control decisions arriving at the gatekeeper after interpretation by CH. The Policy Decision Point (PDP) and Policy Information Point (PIP), still component of XACML provides decision point where serious policy *Matching Algorithms* are implemented. The Identity/Attribute Provider (I/AAP), which is derived from the concept of FIM supplies trusted attributes of entities to facilitate

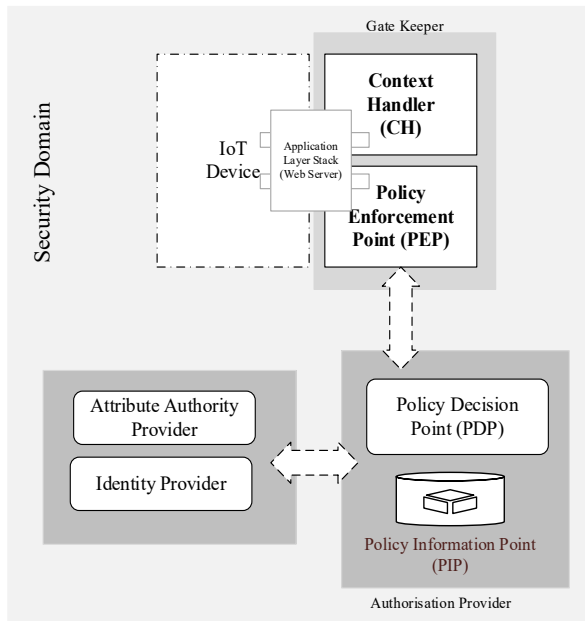


Figure 5: Distributed Access Control Infrastructure.

decision making by PDP. These subunits can then be hosted anywhere in the cloud to provide identity and access control functions. This infrastructural arrangement clearly shows the importance of IoT device belonging to a security domain where there is prevailing trust relationship with the authorization service for the solution to be feasible.

4.2. SMAL OoT Protocol

Figure 6 illustrates a protocol sketch between IoT entities in distributed systems, a way to mutually interexchange SAML OoT Assertion messages in order to decide whether resources can be shared either way [3].

The sequence of interactions are explained in the following steps:

- 1) A classical smart vehicle (SV) arriving a city sends a service request to Smart City Systems (SCS).
- 2) The request is intercepted by SCS Security gate keeper (CH/PEP), which constructs and sends SAML OoT containing Notification of Obligation (NoB) context.
- 3) The SV gate keeper intercepts, constructs and responds with its SAML OoT that contains its NoB.
- 4) The SCS CH constructs another SAML OoT that contains both NoBs and sends to its PDP for processing and decision.
- 5) The SCS's authorization engine based on the policy attributes sends SMAL OoT message to SV's IdP/AA requesting verification of identity/attributes of SV.
- 6) The SV's IdP/AA sends a corresponding SAML OoT *Response* message containing signed identity/attributes requested or makes a fresh request to the sending party (5 & 6 can iterate number of times depending on the trust negotiation configuration).
- 7) The SCS's authorization engine using the policy sets (NoBs) and based on 6 response, performs the *Matching Algorithm* to determine access decision.
- 8) The SCS sends SAML OoT *Response Message* that contains Signed Acceptance of Obligations (SAO) based on 7.
- 9) The SV's sends corresponding SAML OoT message that contains its SAO to SCS.
- 10) Then, SCS sends the requested resources to SV.

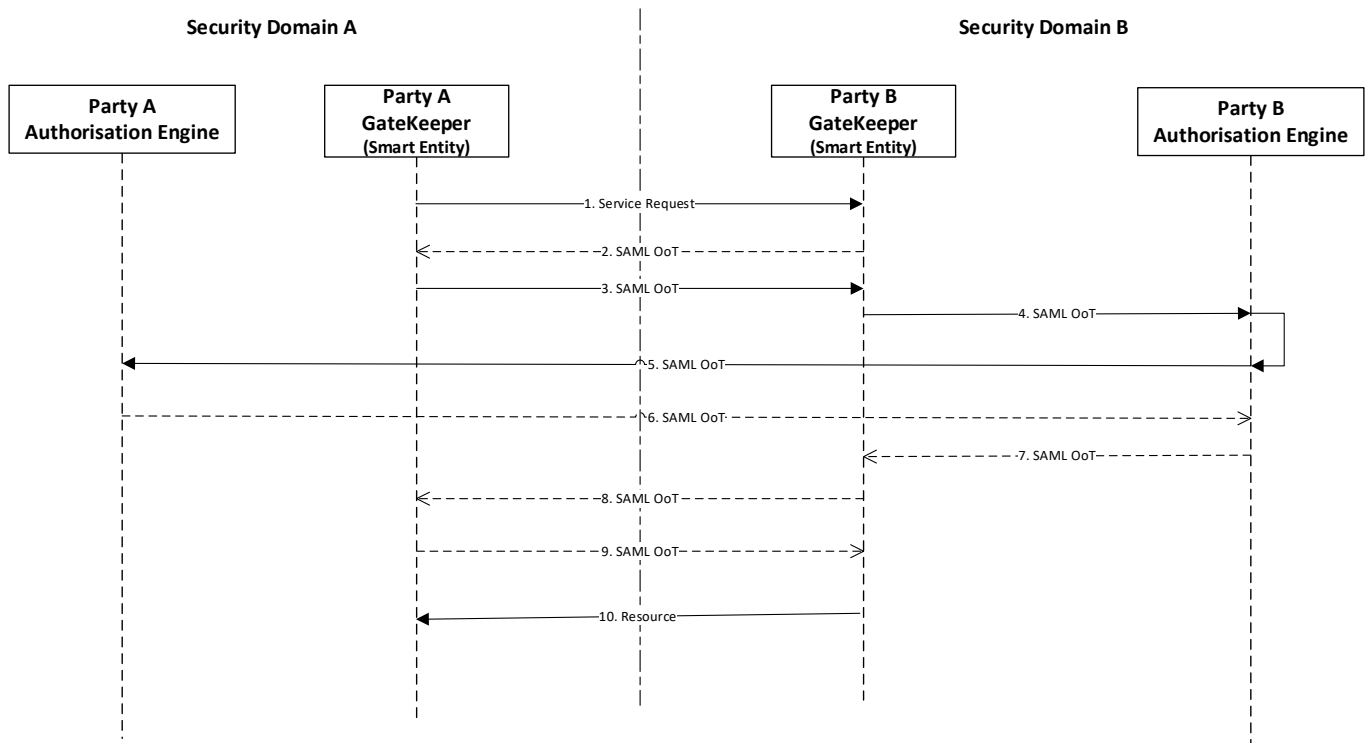
Note: Usually, direct trust can be the basis to initiate the negotiation, which usually, is a form of simple authentication; but not illustrated in the diagram.

4.3. Obligation of Trust Policy Architecture

Technically speaking, IoT in distributed setting is operationally complex and sophisticated, especially in interconnected and integrated application environments where applications talk to applications. To mutually treat trust, confidentiality and privacy, requires a configurable policy set that is robust and scalable. In this context, a policy construct that provides *Requirements* element and *Capabilities* element that permit each party to expressively and granularly describe its obligations and expectations is presented. Conversely, for IoT transaction parties to collaborate together in real-time, they can mutually express and interexchange the policy constructs containing *Requirements* and *Capabilities* in order to treat trust, confidentiality and privacy concurrently as illustrated in Figure 7. As demonstrated, party A's *Requirements* must match with party B's *Capabilities*, and similarly, party B's *Requirements* must match party A's *Capabilities* in a typical access request evaluation. This mutual evaluation gives each party, using granular expressive rules, the preference to decide and balance the sharing of resources in comparison to their mutual benefits. This construct when combined with *Digital Signature* solves confidentiality, integrity, authenticity and non-repudiation, thereby meeting basic security goals in a typical secure transaction.

In summary:

- i. *Requirements* element is used to express a party's obligations (or commitments), it would expect another party requesting for a resource to fulfil before such a resource can be made available;



Note: SAML OoT encapsulates either NoB or SAO Messages

Figure 6: Access Control Conversations between Entities in Distributed Systems.

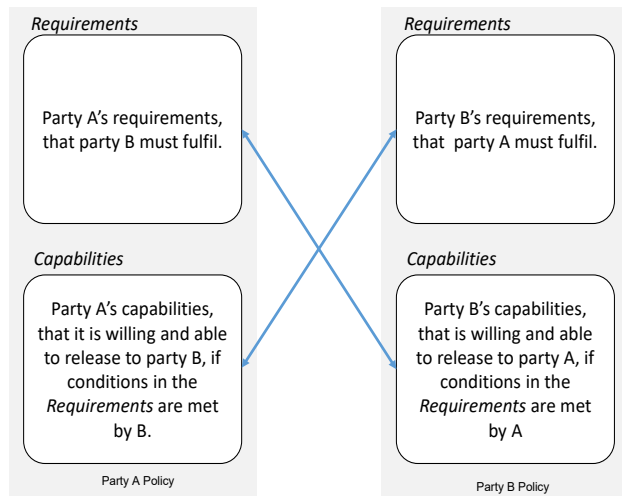


Figure 7: Obligation of Trust Policy Architecture

ii. *Capabilities* is used to express the competences (or services or features) a party is prepared to make available to another party, provided conditions expressed in its *Requirements* section are fulfilled. Thus, *Requirements* and *Capabilities* represent a policy architecture that two or more cooperating parties can leverage to assure trustworthiness, privacy and confidentiality concurrently.

The significant advantages of this policy construct include as follows:

- i. It is a derivative of XACML and SAML standards, making it not too difficult to implement the *Matching Algorithms* and *Messaging* constructs;
- ii. It is flexible to fit into any application context and has the ability to scale proportionately;
- iii. It is extensible.

5. Discussions

The infrastructure presented here uses industry standards such as XACML, SAML, and frameworks including FIM, OoT for distributed access control in IoT environments. The infrastructural subunits are modular and distributed in manner adaptable to use cases where IoT computing resource are constrained. By design, it is expected that for an IoT entity to wade off application layer access intrusions, the gate keeper deny all access request by default. Consequently, to gain access to IoT resources depends on the evaluation of trusted assertion from a corresponding authorization service based on the outcome of *Matching Algorithms* by the PDP using the Policy Sets.

In implementation, there are two ways SAO can be constructed: firstly, the SAO can encapsulate digitally signed *Requirements* and *Capabilities* of a party. This signifies that this asserting party is willing and capable of providing the *Capabilities* if and only if the relying party meets the rules described in its *Requirements*. Secondly, in alternative, the SAO can comprise digitally signed *Capabilities* of the asserting party and the *Capabilities* of the relying party. In this case, it shows that the asserting party agrees to release the *Capabilities* provided the

relying party can reciprocate by releasing its own *Capabilities*.

In scenarios where incremental building of trustworthiness is required, more than one attribute would be required in the rule expressions in such manner that one of the *Subject Descriptors* of the policy indicates the initial attribute to start the first degree trust negotiation as may be required by the parties. Additionally, it can be assumed that parties may not be willing to disclose attributes of privacy value at the first round of the negotiation. In this regard, the policy specification should be such attributes required to build trust are arranged in order of less sensitive to high sensitive attributes. Alternatively, *direct trust* between an IoT entity and its local IdP/AA, can be the basis for starting trust negotiation. The assumption here is that initial information provided by a party is insufficient to breach its privacy or undermine the confidentiality of the protected resources. This initial phase, in theory, is sufficient to counter any attempt by a malicious party to conduct probing attacks[23], usually related with trust negotiations. In this, it is further supposed that if the conversation parties decide to withdraw at the initial stage of the trust negotiation phase, their risks exposure can be significantly reduced. Moreover, if any of the parties is a hostile party, then this early interaction should filter out the access request, and terminate the conversation.

Whereas the first degree of the trustworthiness as described above is inadequate to gain access to IoT services, the parties may provide other levels of trust, which can be specified in the policy construct to help each other reach their various goals. To make this negotiation phase privacy aware, an entity can simply send its SAML OoT containing its security *Requirements* and *Capabilities* across to the other entity. The later party, uncertain whether the other party will conform to its security settings, cannot disclose sensitive information, but correspondingly respond with another SAML OoT that describes its competences and security requirements. This iterative process operationally initializes privacy trust building and interexchange of applicable attribute information in intuitive way, which can result to a number of iterations until both parties are willing to work together.

It is obvious that the prevailing scenarios above is no way a guarantee or assurance that the parties will conform to each other's privacy, so the SAO offers a strong practical protocol that ensures conversation parties generate and interexchange digitally signed difficult-to-repudiate documents containing contextual information that can be admissible in the courts of law.

6. Conclusion

The infrastructure presented here introduces a powerful approach to identity and access management in distributed IoT environments in trust negotiation fashion. It has shown how malicious party's effort to intrude into an IoT system can be thwarted in real-time by gradual and bilateral negotiation to establish trust first before disclosure of protected resources in either direction. Privacy protection and trustworthiness, are behavioural, and possess obligatory expectations, it then implies that privacy and digital trust require a degree of assurance more than traditional security measures can provide. Our infrastructure has addressed security, privacy and safety in situations whereby IoT entities have to solve problems across multiple domains in more trustworthy, adaptive and secure manner. Equally, our approach allows both parties in conversation to mutually address

their security, privacy and safety concerns as opposed to one-way unilateral protection mostly used by the party providing services.

Moreover, we have presented a novel infrastructure with distributed access control components in a fashion that access control Policy Decision Points (PDP), Identity/Attribute Authority (IAA) providers can be delegated to external trusted parties while the constrained IoT system handles context and Policy Enforcement Point (PEP).

Furthermore, by allowing parties to express their access rules and services in *Capabilities* and *Requirements* policy elements, a fine-grained access decisions can improve security and safety. Besides, addressing trust, privacy and confidentiality in a mutual way, our infrastructure provides accountability and conflict resolution approach, which are vital factors for typical IoT distributed deployments.

Conflict of Interest

The authors hereby declare no conflict of interest.

References

- [1] U.M Mbanaso, G.A Chukwudebe, B. Bamidele "Holistic Security Architecture for IoT Technologies," in *13th International Conference on Electronics, Computer and Computation (ICECCO)*, 2017, pp. 11–16.
- [2] U.M Mbanaso, G.A Chukwudebe "Requirement Analysis of IoT Security in Distributed Systems," 2017.
- [3] U. M. Mbanaso, G. S. Cooper, D. Chadwick, and A. Anderson, "Obligations of trust for privacy and confidentiality in distributed transactions," *Internet Res.*, vol. 19, no. 2, pp. 153–173, 2009.
- [4] U. M. Mbanaso, G. S. Cooper, D. W. Chadwick, and S. Proctor, "Privacy preserving trust authorization framework using XACML," *Proc. - WoWMoM 2006 2006 Int. Symp. a World Wireless, Mob. Multimed. Networks*, vol. 2006, pp. 673–678, 2006.
- [5] R. Ross, M. McEvelley, and J. Carrier Oren, "Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems," vol. 1, 2016.
- [6] E. Number and H. Manufacturing, "Privacy and data protection."
- [7] U. M. Mbanaso, G. S. Cooper, D. Chadwick, and A. Anderson, "Obligations for Privacy and Confidentiality in Distributed Transactions," *Ifip Int. Fed. Inf. Process.*, pp. 69–81, 2007.
- [8] T. Ryutov, L. Zhou, C. Neuman, T. Leithead, and K. E. Seamons, "Adaptive trust negotiation and access control," *Symp. Access Control Model. Technol.*, p. 139, 2005.
- [9] H. Gao, J. Yan, and Y. Mu, "Dynamic Trust Model for Federated Identity Management," *Netw. Syst. Secur. (NSS), 2010 4th Int. Conf.*, vol. 2010, pp. 55–61, 2010.
- [10] A. Bhargav-Spantzel, A. Squicciarini, and E. Bertino, "Integrating Federated Digital Identity Management and Trust Negotiations," *CERIAS Tech Rep. 2005-46*, pp. 1–15, 2005.
- [11] J. G. Alessandro Acquisti, "Privacy and Rationality in Individual Decision Making," *IEEE Secur. Priv.*, vol. 3, pp. 26–33, 2005.
- [12] Information Commissioner's Office, "Preparing for the General Data Protection Regulation (GDPR) 12 steps to take now," *Iso*, p. 11, 2016.
- [13] Information Commissioner's Office, "Overview of the General Data Protection Regulation (GDPR)," p. 43, 2017.
- [14] C. Weyrich, Michael und Ebert, "Reference Architectures for the Internet of Things," 2016.
- [15] P. Fremantle, "A reference architecture for the internet of things," *WSO2 White Pap.*, vol. 0, p. 21, 2014.
- [16] Symantec, "An Internet of Things Reference Architecture," *Symantec White Pap.*, pp. 1–22, 2016.
- [17] K. E. Skouby and P. Lynggaard, "Smart home and smart city solutions enabled by 5G, IoT, AAI and CoT services," *Proc. 2014 Int. Conf. Contemp. Comput. Informatics, IC3I 2014*, pp. 874–878, 2014.
- [18] A. Torkaman and M. A. Seyyedi, "Analyzing IoT Reference Architecture Models," *Int. J. Comput. Sci. Softw. Eng. ISSN*, vol. 5, no. 8, pp. 2409–4285, 2016.
- [19] S. V. Nath, "IoT architecture," *Internet Things Data Anal. Handb.*, pp. 239–249, 2017.

- [20] D. Chadwick, G. Inman, and N. Klingenstein, "Authorisation using Attributes from Multiple Authorities – A Study of Requirements," p. 4.
- [21] B. E. Bhargav-Spantzel Abhilasha, Squicciarini Anna Cinzia, "Identity Management Concepts Technologies Systems," *IEEE Secur. Priv.*, vol. 5, no. 2, pp. 55–63, 207AD.
- [22] OpenID, "OpenID Decentralized Authentication," 2017. [Online]. Available: <https://openid.net/>.
- [23] K. E. Seamons Winslett, M. & Yu, T., "Limiting the Disclosure of Access Control Policies during Automated Trust Negotiation," *New York Distrib. Syst. Secur. Symp.*, pp. 1–11, 2001.
- [24] D. Chadwick, G. Zhao, S. Otenko, R. Laborde, L. Su, "Building a Modular Authorization Infrastructure," *Kent Acad. Repos.*, pp. 5–15, 2010.
- [25] B. Parducci and H. Lockhart, "eXtensible Access Control Markup Language (XACML) Version 3.0," *OASIS Stand.*, no. January, pp. 1–154, 2013.
- [26] L. S. V et al., "Security and Privacy Considerations for the OASIS Security Assertion Markup," *Management*, no. August, pp. 1–33, 2004.
- [27] V. Felmetzger, "Security Assertion Markup Language (SAML) SAML as OASIS Standard," 2006.
- [28] U. M. Mbanaso, "Design of Obligation of Trust Protocol," no. May, pp. 19–20, 2008.
- [29] A. S. Elmaghraby and M. M. Losavio, "Cyber security challenges in smart cities: Safety, security and privacy," *J. Adv. Res.*, vol. 5, no. 4, 2014.
- [30] OWASP, "Internet of Things Top Ten," 2017.
- [31] Gartner, "Leading the IoT, Gartner Insights on How to Lead in a Connected World," *Gart. Res.*, pp. 1–29, 2017.
- [32] Information Commissioner's Office, "Guide to the General Data Protection Regulation (GDPR)," p. 153, 2017.
- [33] U. M. Mbanaso, "Privacy Preservation Architecture for Authorization Infrastructure."
- [34] K. E. S. Tatyana Ryutov, Li Zhou, Clifford Neuman, Travis Leithead, "Adaptive trust negotiation and access control," 2005, pp. 139–146.
- [35] U. M. Mbanaso, G. S. Cooper, D. Chadwick, and A. Anderson, "Obligations of trust for privacy and confidentiality in distributed transactions) "Obligations of trust for privacy and confidentiality in distributed transactions Obligations of trust for privacy and confidentiality in distributed transactions," *Obligations Trust Priv. confidentiality Distrib. Trans.*, vol. 19, no. 2, pp. 153–173, 2009.
- [36] U.M. Mbanaso, Centre for Cyberspace, Nasarawa State University, "Personal Data Privacy and Security - Who , What , When , Why , Where and How?," in *DIRISA National Research Data Workshop, Pretoria South Africa*, 2018, no. June, pp. 1–8.

A Holistic User Centric Acute Myocardial Infarction Prediction System With Model Evaluation Using Data Mining Techniques

Procheta Nag^{*1}, Saikat Mondal¹, Arun More²

¹ Computer Science And Engineering, Khulna University, Khulna-9208, Bangladesh

² Department of Cardiology, Ter Institute of Rural Health and Research, Murud-413510, India

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 14 October, 2018

Online: 02 November, 2018

Keywords:

Data Mining

Coronary Heart Disease (CHD)

Acute Myocardial Infarction

Random Forest

Mobile Application

ABSTRACT

Acute Myocardial Infarction (Heart Attack), a Coronary Heart Disease (CHD) is one of the major killers worldwide. Around one thousand data has been collected from AMI patients, people are at risk of maybe a heart attack and individuals with the significant features closely related to heart attack. The sophistication in mobile technology, health care applications offers remarkable opportunities to improve our health, safety and in some sense preparedness to common illnesses. The excess delay time between the onset of a heart attack and seeking treatment is a major issue which may lead to permanent blockage or even die often. So, a proficient mobile application approach is projected in this paper that may predict the possibilities of a attack once an individual is bearing the noticeable symptoms of chest pain. Random forest predicts the result of the user input features and the automated result is shown on the smart-phone application. The application categorizes the prediction of the user's input as a heart attack, maybe heart attack and no heart attack. The experimental results showed that the accuracy of the proposed technique is 92%, whereas the precision is 95%, 92%, 87% respectively for heart attack, maybe heart attack and no heart attack prediction. Our research target is to raise heart attack awareness on time in an innovative way through available and accessible medium to mass people.

1 Introduction

Acute myocardial infarction (AMI) commonly referred to as Heart Attack is the most common cause of sudden deaths in city and village areas [1]. Coronary Heart disease (CHD), cancer, chronic respiratory disease, and diabetes are becoming fatal at an alarming rate [2]. Acute myocardial infarction occurs when there is a sudden, complete blockage of a coronary artery that supplies blood to an area of the heart also known as Heart Attack [3]. A blockage can develop due to a buildup of plaque, a substance mostly made of fat, cholesterol and cellular waste products [4]. Due to the insufficient blood supply, some of the heart muscles begin to die.

According to the World Health Organization

^{*}Corresponding Author: Procheta Nag, Computer Science And Engineering, Khulna University, Khulna-9208, Bangladesh, +8801747120243, Email:prothoma07p@gmail.com

(WHO) report published in May 2014 coronary heart disease deaths (CHDD) in Bangladesh reached 6.96% of total deaths [2],[5]. Detecting heart attack on time is of paramount importance as delay in detecting may lead to severe damage to heart muscle, called myocardium leading to morbidities and mortalities.

The fast integration of mobile devices into clinical observation has been driven by the rising availability and quality of medical software package applications like mobile applications. There are many Mobile Health (mHealth) tools available to the consumer in the prevention of CHD such as self-monitoring mobile applications. Current science shows the evidence on the use of the vast array of mobile devices such as the use of mobile phones for communication and feedback, smart-phone applications [6]. The visiting fees of doc-

tors are costly; however, medical applications aim to vary that considerably within the close future. Essentially, it is currently attainable for a smart-phone to interchange the association in nursing in-person doctor consultation, and the virtual appointment with a doctor is apparently less costly than some real-life doctor visits. The mobile applications within the health care business can basically facilitate patients to schedule appointments, monitor the aspect effects of a medicine, prompt them to require bills, analyze health reports, and do a plethora of works. These advance portable wellbeing health applications can alter the way they approach patients and doctors have interaction whereas remodeling the long run traditional way of the medical sector.

The health sector is enriched with data but the major issues with therapeutic information mining are their volume and complexity, destitute numerical categorization and canonical form [7]. Whereas aggregating clinical records and discoveries on paper are not well composed, consequently, numerous information captured in unstructured is troublesome to total and analyze in any steady way. In any case, since it appears to have such an enormous effect, the information has to be exact, comprehensive and convenient. Therapeutic information is time growing and ever-changing, making it outlandish to expect what unused information or modern necessities will see like and how they can fit into a demonstrate. Classification of medical data for idealize supposition is an on the rising field of significance and explore in records expulsion. Whereas creating standard forms such as a portable application that moves forward quality is one of the objectives in wellbeing care.

When observing patients in the clinic, the queries they are inquired and treatment they are given are all noted down in therapeutic records. Collecting and utilizing data from patients reflect the reality of day to day health condition of AMI patients present time. One of the greatest challenges is to free information from the silos in which it usually remains, a difficulty that influences each patient care and medical analysis. Although this information is private, once anonymized this information holds unlimited potential for public benefit, so long because it is utilized securely and viable. Furthermore, information security is a crucial truth because hackers have made wellbeing care information a significant target nowadays. So, as for individuals to feel comfortable sharing their information on the mobile application and encourage them to use the applications are important.

Individuals who are busy in their homes or offices with their regular works and rural people having no knowledge on the symptoms of heart attack may neglect the chest discomfort. Reducing the delay time between the onset of a heart attack and seeking treatment is a major issue [1]. As the medical diagnosis of heart attack is an important but complicated and costly task, we proposed an automated system for medical diagnosis that would enhance medical care and reduce cost. Our aim is to provide an ubiquitous service that

is both feasible, sustainable and which also help people to assess their risk for heart attack at that point of time or later.

In order to provide the automated result of the user data random forest is used to predict the result, as it is an ensemble method to enhance the accuracy of the large and multi-class dataset [8]. Upon the preprocessed data, the individual decision trees are generated using a random selection of attributes at each node to determine the splits. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [9]–[11]. During classification, each tree votes and the most popular class is returned as the predicted class which is the result of the user input, finally, the developed REST API with Python and Flask shows the automated result on smart-phone application by JSON parsing.

2 Related Works

In [12], the author proposed a system for detecting heart attack by analyzing the number of beats per minute (BPM). They have used the sensor to detect the heart attack and intimate the occurrence of heart attack to the helpline in wireless GSM module. In the system, the heartbeat sensor detects the heartbeat rate from the finger of the user and LCD display is used to display that. Then the output is sent to the microcontroller where the microcontroller runs the heart attack detection algorithm, eventually, display the number of pulses in the LCD display. The pulse rate except for the range (60-90) occurs it is regarded as the indication of abnormal or heart attack. Once if abnormal is detected, then the microcontroller will activate the output to the GSM module and GSM module will send the alert message to mobile numbers already coded in a microcontroller.

In [13], the author proposed a system that can automatically predict heart disease. How data be turned into useful information that can enable healthcare practitioners to make effective clinical decisions, was the main objective of this research. They used Cleveland heart disease dataset which is available in the UCI machine learning repository. They have designed a system using Decision Tree (C4.5) as a method that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. The rules can be prioritized based on the users requirement. They had used KEEL (Knowledge Extraction based on Evolutionary Learning) tool for prediction.

In [14], the author proposed a system that they had developed a prototype Heart Attack Prediction System (HAPS) using data mining techniques, namely, Decision Trees, Naïve Bayes, and Neural Network. The main objective of this research is to develop a Web Application using data mining modeling technique calls Naive Bayes. The scope of the project is that integra-

tion of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation and improve patient outcome. The system takes some dataset, generate questions depending on them and creates decision calculating them by Naive Bayes classifier.

In [15], the researcher proposed a mobile phone application that can help victims to identify whether they are having a heart attack or not without going to a specialist in person. They have used a mobile phone application with some questions to analyze, a wearable electrocardiogram (ECG) sensor, blood pressure measurement device. If a person is in danger (cardiac arrest, fall) and unable to call an ambulance, the mobile phone will automatically determine the current location of the person using WiFi, GSM Cell-id or GPS and sends automated voice and text messages to their cardiologist.

In [16], the author analyzed the Cardiovascular Disease (CVD) rate in Singareni coal mining regions in Andhra Pradesh state, India. This study analyzed the Cardiovascular Disease (CVD) rate in Singareni coal mining regions in Andhra Pradesh state, India. They have used Decision Trees, Naïve Bayes and Neural Network as their method and UCI repository dataset for their analysis with 15 attributes to predict the morbidity. They have used data mining techniques: Decision Trees, Naive Bayes and Neural Network as their method and UCI repository dataset for their analysis. Bayesian model (BN) achieved a classification accuracy of 0.82 with a sensitivity of 0.87, The decision trees (C4.5) achieved a classification accuracy of 0.825 with a sensitivity of 0.8717. However, the neural network model (MLP) performed the best of the four models evaluated. MLP achieved a classification accuracy of 0.897 with a sensitivity of 0.9017.

In [6], the group of scientist reviewed how mobile health can play important role in cardiovascular disease prevention. The studies reviewed in this statement targeted the behaviors (ie, smoking, physical activity, healthful eating, and maintaining a healthful weight) and cardiovascular health indicators (ie, blood glucose, lipids, BP, body mass index) as the primary outcomes in the clinical trials testing mobile health (mHealth) interventions. They showed many statistics and gave some idea how mobile health (mHealth) can help to prevent cardiovascular disease.

In [17], the authors described an attempt was taken to find out interesting patterns from data of heart patients by these three algorithms namely, Decision Tree, Neural Network and Naive Bayes in two different scenarios. An on-line available dataset of heart patients with Weka data mining software was used for implementation. The experiments consist of two scenarios, one scenario with all 14 attributes and the other scenario with 8 selected attributes. The Naive Bayes classifier algorithm with all attributes shows the highest accuracy i.e. 82.914% and Naive Bayes with selected attributes is nearest to it with 82.077% accuracy. On the other hand, C4.5 decision tree (un-pruned) with

all attributes score the lowest accuracy i.e. 77.219%. So, data mining can be used to predict heart disease efficiently and effectively.

Though the approached systems of these papers possess some advantages, the drawbacks of those systems can not be overlooked. The advantages and drawbacks of these papers are mentioned below;

Table 1: Merits and demerits

Paper title	Advantages	Drawbacks
Current science on consumer use of mobile health for cardiovascular disease prevention [5]	Demonstrates the great potential that mobile technologies can have to aid in health care	Lack of evidence to some extent
Embedded based automatic heart attack detector and intimator [12]	Automated message system	Sends alert message based on a fixed pulse rate range without ensuring heart attack
Efficient heart disease prediction system using decision tree [13]	The system has great potential in predicting the heart disease risk level	Can predict the risk of heart diseases but can not predict the risk heart attack. Used only 10 attributes
Heart Attack Prediction System Using Data Mining Techniques [14]	Signifies the computer based patients records for clinical supports	Uses categorical data. For some diagnosis the use of continuous data may be necessary
A Self-test to Detect a Heart Attack Using a Mobile Phone and Wearable Sensors [15]	Automatic messaging system from the patients mobile to the emergency number	Can be costly for mass people as there are few external devices and difficult to use by people who has less tech knowledge

Table 2: Merits and demerits

Paper title	Advantages	Drawbacks
Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques [16]	Shows good analytical results on different algorithms to signify early detection of heart attack	Studied on a particular area (coal mining region). Used only 15 attributes to predict heart attack
Data mining in health care for heart diseases [17]	Uses attribute selection method to increase the classification accuracy and decrease the time and complexity	Though dataset had 76 raw attributes but only 14 of them are actually most important to the related topic

3 Methodology

The total system is organized in three modules: at first, preprocessed the raw data, then applied Random Forest algorithm upon the data to predict the possibilities of heart attack and finally the result is showed through the mobile application.

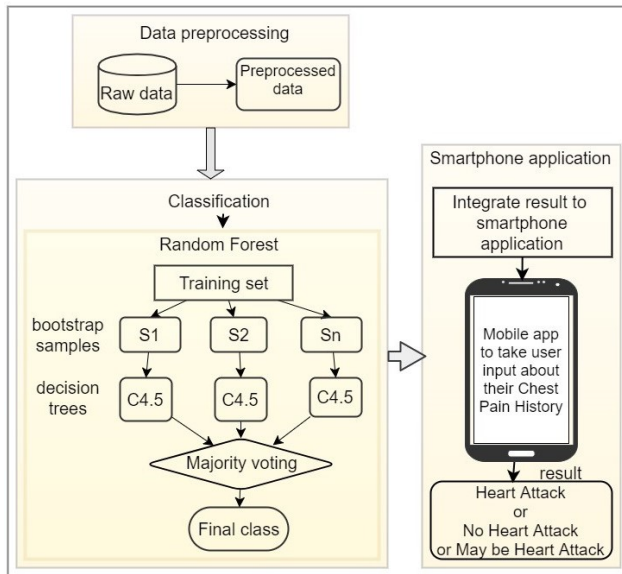


Figure 1: System architecture

The total system is automated, the data mining algorithm is predicting the class of unknown user features on runtime in the back end of the mobile application and predicted class is showed on the application screen.

3.1 Data Preprocessing

Managing the data related to each of the classes and turning it into something usable across a system is one of the major and time-consuming segments for the health data since a total number of 1000 data has been collected from the real world. The datasets consist of different formats (e.g. text, numeric) and sometimes the same data exists in the different dataset and in different formats as they are collected from different sources. Afterward Aggregating this data into a single, central system, inaccurate, incomplete, and inconsistent data may appear as they are commonplace properties of large real-world databases. So, it is obvious to pre-process them in the required format for analysis, hence, it improves the accuracy and efficiency of mining algorithms. Three significant steps of processing dataset are as follow-

1. Trimmed mean is used to handle numerical missing values in the dataset. By deleting a significant portion of outliers, mean is computed using the remaining values, consequently, missing values are filled with the value of the trimmed mean [8].
2. Eliminating redundant features to reduce data size is an emergent step to boost the efficiency of

the training dataset. So attribute subset selection is used to reduces the data set size by removing irrelevant or redundant attributes [8][18].

3. Data are transformed so that the resulting mining process can be more efficient, and the patterns found may be easier to understand. For the shake of correlation, some relevant attributes are added to the original dataset to improve the mining process [8].

3.2 Prediction

After preprocessing, the dataset is trained to the learning model to generate the result that either the chest pain is for heart attack or not or maybe for heart attack. Random Forest, a data mining algorithm [19] [20] is used to predict the classes. There are two stages in the Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage. The whole process is shown below-

At first the Random Forest creation:

1. Randomly selected two-third bootstrap sample features from total features where bootstrap sample < total features.
2. Among the sample features, calculated the node d using the best split point.
3. Splited the node into daughter nodes using the best split.
4. Repeated the steps one to three until l number of nodes had been reached.
5. Built forest by repeating steps one to four for maximum number (such as 200 times) times to create a maximum number (such as 200) of trees.

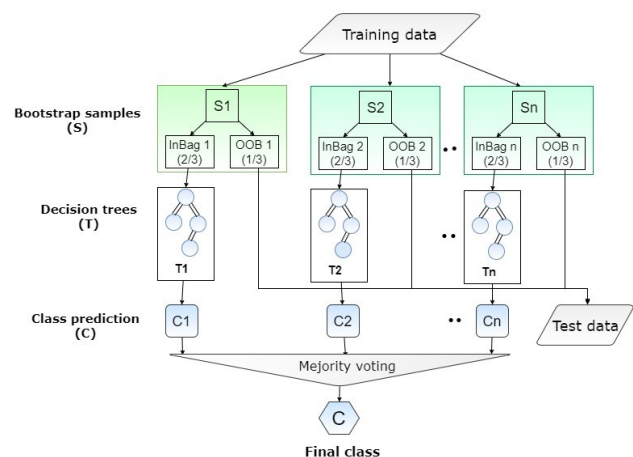


Figure 2: Random Forest algorithm flow chart

In the next stage, while the random forest classifier created, the prediction is made. Prediction procedure by random forest is shown below:

Training data: The collected one thousand data is used as training data of the random forest model.

Test data: User input data from the mobile application is considered as test data which are not classified and unknown because they do not exist in the training dataset. The target is to predict the output class of the test data.

1. Took the test features and use the rules of every randomly created decision tree to predict the outcome and stores the predicted outcome.
2. Calculated the votes for each predicted class.
3. Considered the predicted class which got the majority votes as the final prediction from the random forest algorithm.

Random forest is used for predicting the result because-

Over-fitting reduction: There are considerably lower hazards of over-fitting by averaging some trees.

Handle vast dataset: It has the facility to handle the large information set with higher spatial property and maintain the accuracy of an over sized proportion of knowledge.

Less variance: By utilizing various trees, it reduces the chance of staggering across a classifier that will not perform well owing to the connection between the train and test data.

Feature importance: The Random Forest provides the distinctive vital features from the dataset, in alternative words, feature importance.

3.3 Smart-phone Application

A user input form with heart attack prediction questionnaire which is mostly related to the symptoms of heart attack is made to take input from the users. Whenever a user gives input to the input boxes, the data is sent to the server and generate an automated result from the learning model on the back-end. New data will be saved to the server continuously in order to keep the user records. The resulting process is an off-line and automated system and predicts heart attack on the basis of input that will be provided by the user.

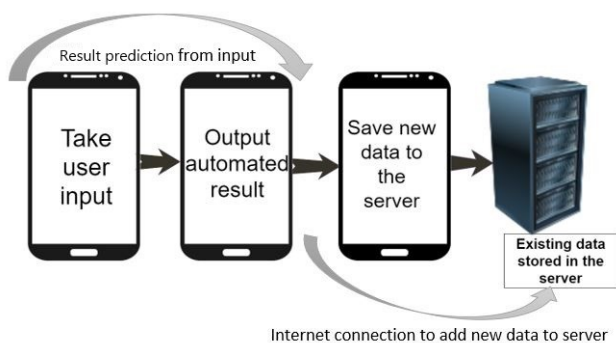


Figure 3: Data flow on smartphone application

4 Implementation and Results Analysis

4.1 Dataset

The model is trained on total 1000 data including 550 AMI patients admitted in three different cardiology specialized hospitals, 280 individuals data who are in risk of heart attack and 170 general people data have been collected and analyzed with 22 attributes which are closely connected to the heart attack symptoms. Age, gender, hypertension, diabetics, cholesterol, smoking, family history, chest pain, without chest pain, chest pain time, chest pain location, chest pain type, chest pain mark, chest pain going, chest pain association, chest pain duration, chest pain subsided, chest pain relieve, past similar pain, doing while it started-are the major features of the model. As a consequence, users input data is also referred to add with the trained dataset.

4.2 Integration

Developed REST APIs with Python and Flask to show the generated result on smart-phone application by JSON parsing.

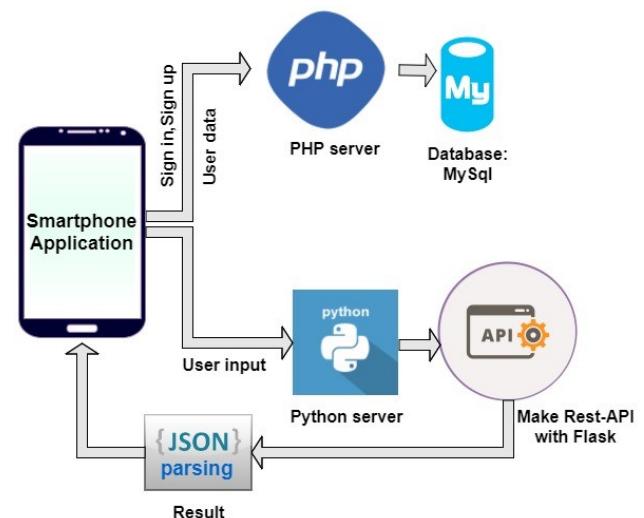


Figure 4: JSON parsing and API creation with Flask

Figure 4.6 illustrates the architecture upon which the mobile application prediction system is built. It shows the flow of how the Mobile application connects to the server and how the server sends particular results according to the request sent by the application. When a user fills up the heart attack prediction questionnaire the inputs are converted to a comma separated string with preprocessing and sent to the server. In the server, the python code takes it as an input (test data) and returns a JSON formatted result. To make this REST API we have used the Python Flask. The JSON data are then parsed inside the application to show the final result (Heart Attack, No Hear Attack,

Maybe Heart Attack). Similarly, for user login and registration, data are sent to the server, executed there and the server sent JSON response which is used by the application to perform specific tasks (Sign in, Sign up). New data will be saved to the server each time after submitting the user form.

4.2.1 Smart-phone Application Functionality

Since it is a cross-platform application it can run any type of smart system as well as smart-phone. The general pages with functionalities of the application are given below

1. A new user has to Sign Up for creating an account whereas an existing user can log in to existing account.
2. This is the Registration page for creating account.
3. After login to check the history of previous analysis result with time click on History. And in order to fill up the questionnaire click on New Entry.
4. A user starts answering the required questions for prediction.
5. After answering all the questions the user may press Calculate Risk button to see what his selected symptoms mean. At the same time, input data will be saved to the server dataset on the presence of Internet connection.
6. In the result, screen user can see the predicting result of with suggestion related to that. Calculating the input data the random forest algorithm predicted the result as Heart Attack along with the necessary suggestion. However, the result may differ according to the user input data.

If the result shows Maybe heart attack which means his chest pain related symptoms can be for heart attack or lead to heart attack, so he must be aware of that as well as may consult with a cardiologist.

On the contrary, if the result shows no heart attack that means his chest pain related symptoms are not for heart attack.
7. The user can see his previously checked result history.

After implementing the whole process the accuracy has come with 92%. Based on symptoms of the user input the result has been calculated within three types: heart attack may be a heart attack or no heart attack. Along with accuracy other result measurement criteria are as follow.

4.3 Confusion matrix

The values in the diagonal would always be the true positives (TP) which are indicated by blue color in the table.

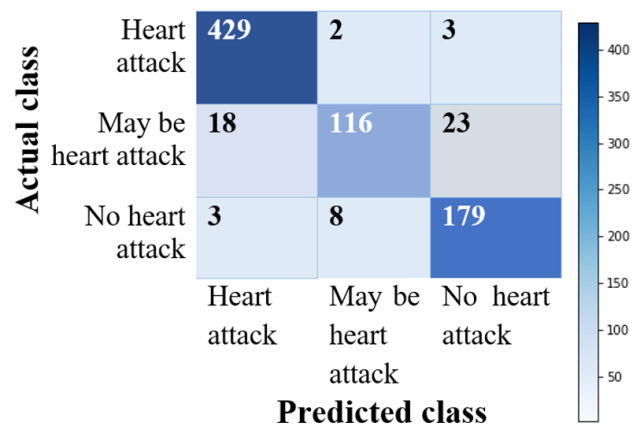


Figure 5: Confusion matrix (without normalization)

429 users data are correctly predicted as heart attack who are actually suffering from AMI and 179 users data are correctly predicted who has no possibility if heart attack apparently. Alongside, 116 data are referred as may be heart attack which actually possesses the symptoms of heart attack.

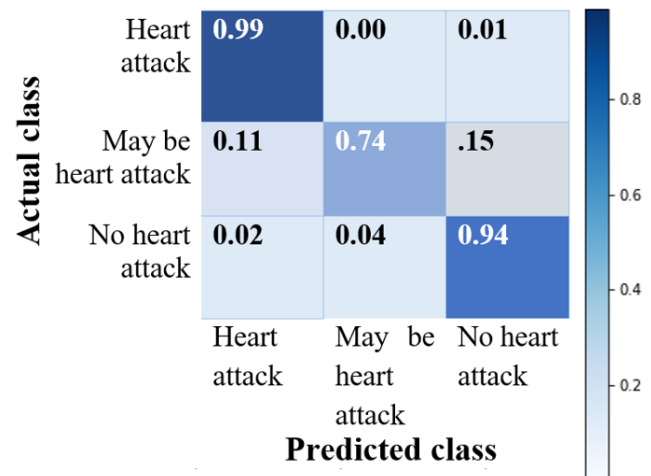


Figure 6: Normalized confusion matrix

Normalizing the confusion matrix it is depicted that the Heart attack prediction rate is very efficient (0.99 out of 1) by the model and also the people who are not suffering from heart attack are almost correctly identified. However a tiny fraction of data has exhibited incorrect prediction.

Accuracy: Accuracy refers that how much a classifier correctly classify the test set that is given to the classifier [8][18].

$$accuracy = (TP + TN)/(P + N) \tag{1}$$

Table 3: Confusion matrix analysis

Class	Precision	Recall	F1-score	Support
Heart at-tack	.95	.99	.97	434
May be heart at-tack	.92	.74	.82	157
No heart attack	.87	.94	.91	190
Avg/Total	.93	.93	.92	781

Table 4: Feature importance

Rank	Serial no	Feature name	Value
1	17	Chestpain_location	0.145674
2	0	Age	0.138248
3	36	Persistence	0.11042
4	20	Chestpain_time	0.094065
5	35	Subsided	0.079373
6	19	Pain_type	0.071126
7	34	Pain_relieved	0.047263
8	33	Doing_while_started	0.039261
9	18	Chestpain_mark	0.035455

The accuracy measure is not appropriate always because it does not check the possibility of tuples belonging to more than one class [8],[18].

Precision: Precision refers to the percentage of correct positive tuples that are labeled as positive [8],[18].

$$precision = (TP / (TP + FP)) \tag{2}$$

Recall: Recall refers to the percentage of positive tuples which are actually positive. Recall is also known as sensitivity or the true positive rate [8],[18].

$$recall = (TP / (TP + FN)) = TP / P \tag{3}$$

F1 score: F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account [14]. F1 is usually more useful than accuracy because accuracy works best if false positives and false negatives have a similar cost. If the cost of false positives and false negatives are very different, it is better to look at both

Precision and Recall.

$$F1\ Score = 2 * (Recall * Precision) / (Recall + Precision) \tag{4}$$

Around 781 data showed true prediction among the total dataset as support is the number of occurrences of each label in predictive true values. Precision, recall, f1-score, and support have been calculated from the confusion matrix. From precision 93% correct tuples that are predicted as positive, similarly from recall 93% data are predicted as heart attack, maybe a heart attack and no heart attack which are actually that.

4.4 Feature Importance

The major benefit of using ensembles of decision tree methods like Random Forest is that they can automatically provide estimates of feature importance from a trained predictive model. It is a technique to evaluate the importance of features of the model. The result of feature importance with ranking the features on their feature importance of our dataset is given below-

Table 5: Feature importance (cont...)

Rank	Serial no	Feature name	Value
10	30	Association_(Sweating)	0.027388
11	3	Hypertension_medicine_years	0.025257
12	21	Pain_going_(Left arm)	0.01841
13	8	Smoking	0.016303
14	5	Diabetes_medicine_years	0.015461
15	1	Gender	0.015165
16	10	Chestpain	0.014167
17	4	Diabetes	0.014047
18	9	Family_history	0.013066
19	37	Similar_pain/chest_discomfort_in_past	0.01179
20	23	Pain_going_(Back)	0.009163
21	29	Association_(Palpitation)	0.008682
22	2	Hypertension	0.00836
23	28	Association_(Nausea)	0.008006
24	14	Symptoms_Except_Chestpain_(Sweating)	0.007676
25	6	Cholesterol	0.005911
26	32	Association_(Vomiting)	0.005026
27	7	Cholesterol_medicine_years	0.004365
28	22	Pain_going_(Right arm)	0.00368
29	12	Symptoms_Except_Chestpain_(Giddiness)	0.002808
30	26	Association_(Dyspnea)	0.001804
31	11	Symptoms_Except_Chestpain_(Dyspnea)	0.001221
32	24	Pain_going_(Upper jaw)	0.001037

The graph exhibits the sequence of important features to calculate the result in ascending order.

5 Conclusion

Our research was focused on the use of data mining techniques interpreting with a mobile application in health care specifically in identifying acute myocardial

infarction. We used training data of heart patients and individuals by collecting from the real-world as well as hospitals whereas test data is received through the mobile application from the users. The mobile application is used to take user input and show the result which is predicted by the random forest algorithm. Also provides the sequence of the importance of the features to check which one has the higher impact of a heart attack. To evaluate the performance of the model prediction different performance metrics were considered where the precision is 93% and recall is 93% in total for heart attack, maybe a heart attack and no heart attack prediction. Finally, the user can check their condition from the application result and be aware of taking the necessary steps.

Table 6: Feature importance (cont...)

Rank	Serial no	Feature name	Value
33	16	Symptoms_Except_Chestpain_(Vomiting)	0.000243
34	15	Symptoms_Except_Chestpain_(Syncope)	0.00008
35	27	Association_(Giddiness)	0
36	13	Symptoms_Except_Chestpain_(Nausea)	0
37	31	Association_(Syncope)	0
38	25	Pain_going_(No_movement)	0

It may get rid of problems connected with human fatigue and habituation raise awareness of abnormalities identifications and alter fast prediction in real time. Development of a mobile application to predict the possibilities heart attack risk would profit lots of individuals. Having a leading framework to suspect the pain as alarming an attack or not an attack might facilitate several such those who tend to neglect the pain and later finally end up within the catastrophe of heart attacks. In this circumstance, we hope that our mobile application PredictAttack will serve the people for useful purposes to taking care of heart through its flexible interactive approach.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment Computer Science and Engineering Discipline, Khulna University, Khulna for providing technological and Rural Health Progress Trust (RHPT), Murud, Latur, India for providing clinical support.

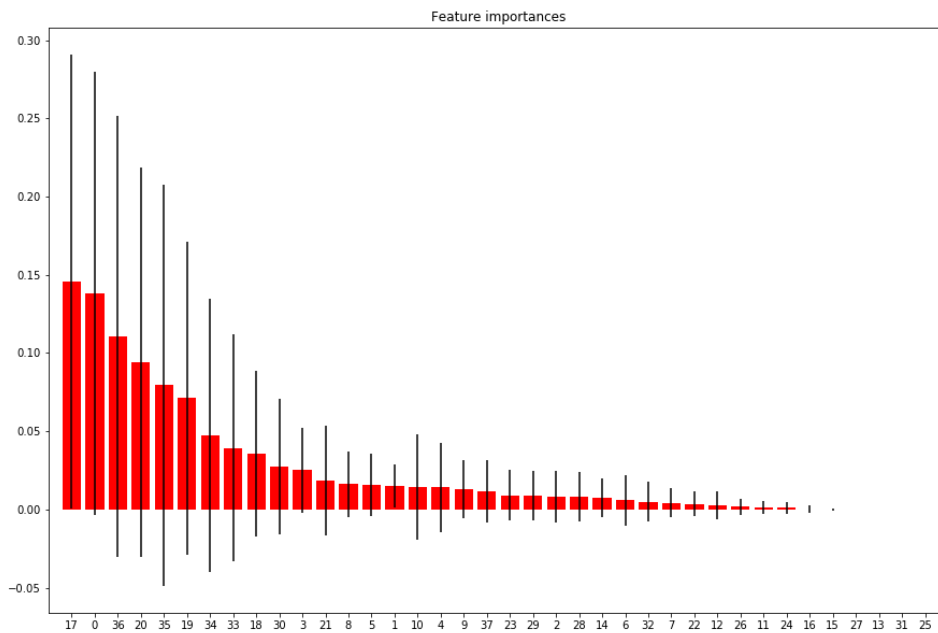


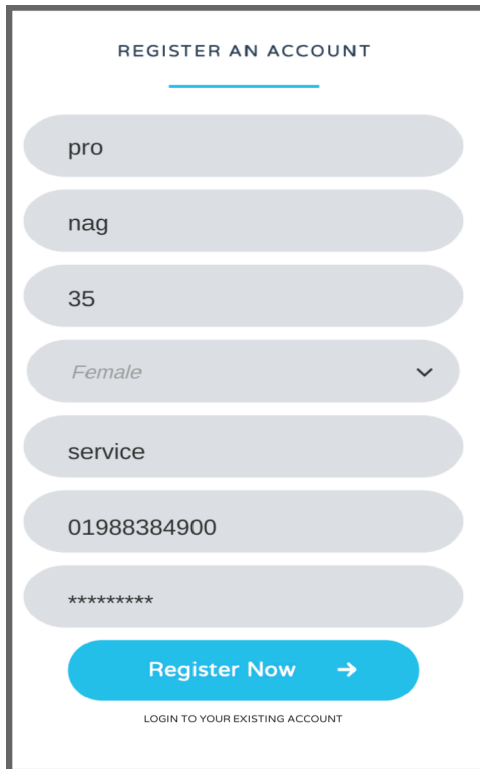
Figure 7: Feature importance

References

- [1] P. Nag, S. Mondal, F. Ahmed, A. More, and M. Raihan, "A simple acute myocardial infarction (heart attack) prediction system using clinical data and data mining techniques," in *Computer and Information Technology (ICCIT)*, 2017 20th International Conference of. IEEE, 2017, pp. 1–6. <https://doi.org/10.1109/iccitechn.2017.8281809>
- [2] Islam, AKM Monwarul, and A. A. S. Majumder. "Coronary artery disease in Bangladesh: a review." *Indian heart journal* 65, no. 4 (2013): 424–435. <https://doi.org/10.1016/j.ihj.2013.06.004>
- [3] White, Harvey D., and Derek P. Chew. "Acute myocardial infarction." *The Lancet* 372, no. 9638 (2008): 570–584. https://doi.org/10.1007/springerreference_44106
- [4] Country statistics and global health estimates by WHO and UN partners website: World Health Organization, "Bangladesh: country profiles", 2015. Available: http://www.who.int/gho/countries/bgd/country_profiles/en/, retrieved on 26 July, 2018.
- [5] <http://www.worldlifeexpectancy.com/bangladesh-coronary-heart-disease>, retrieved on 26 July, 2018.
- [6] L. E. Burke, J. Ma, K. M. Azar, G. G. Bennett, E. D. Peterson, Y. Zheng, W. Riley, J. Stephens, S. H. Shah, B. Suffoletto et al., "Current science on consumer use of mobile health for cardiovascular disease prevention," *Circulation*, vol. 132, no. 12, pp. 1157–1213, 2015. <https://doi.org/10.1161/cir.0000000000000232>
- [7] C. Alexander and L. Wang, "Big data analytics in heart attack prediction," *J Nurs Care*, vol. 6, no. 393, pp. 2167–1168, 2017. <https://doi.org/10.4172/2167-1168.1000393>
- [8] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*, Elsevier, 2011.
- [9] A. Liaw, M. Wiener et al., "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [10] Khalilia, Mohammed, Sounak Chakraborty, and Mihail Popescu. "Predicting disease risks from highly imbalanced data using random forest." *BMC medical informatics and decision making* 11, no. 1 (2011): 51. <https://doi.org/10.1186/1472-6947-11-51>
- [11] Livingston, Frederick. "Implementation of Breimans random forest machine learning algorithm." *ECE591Q Machine Learning Journal Paper*, 2005.
- [12] D. Selvathi, V. V. Sankar, and H. Venkatasubramani, "Embedded based automatic heart attack detector and intimator," in *Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017 International Conference on. IEEE, 2017, pp. 1–6. <https://doi.org/10.1109/iciiecs.2017.8275839>
- [13] Saxena, Kanak, and Richa Sharma. "Efficient heart disease prediction system using decision tree." In *Computing, Communication & Automation (ICCCA)*, 2015 International Conference on, pp. 72–77. IEEE, 2015. <https://doi.org/10.1109/ccaa.2015.7148346>
- [14] S. A. Pattekari and M. A. Yadav, "Heart attack prediction system using data mining techniques." *International Journal of Ethics in Engineering & Management Education*, vol. 1, no. 1, pp. 34–37, Jan. 2014.
- [15] P. Leijdekkers and V. Gay, "A self-test to detect a heart attack using a mobile phone and wearable sensors," in *Computer-Based Medical Systems*, 2008. CBMS08. 21st IEEE International Symposium on. IEEE, 2008, pp. 93–98. <https://doi.org/10.1109/cbms.2008.59>
- [16] K. Srinivas, G. R. Rao, and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," in *Computer Science and Education (ICCSE)*, 2010 5th International Conference on. IEEE, 2010, pp. 1344–1349. <https://doi.org/10.1109/iccse.2010.5593711>
- [17] Shafique, Umair, Fiaz Majeed, Haseeb Qaiser, and Irfan Ul Mustafa. "Data mining in healthcare for heart diseases." *International Journal of Innovation and Applied Studies* 10, no. 4 (2015): 1312.
- [18] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [19] Cutler, Adele, D. Richard Cutler, and John R. Stevens. "Random forests." In *Ensemble machine learning*, pp. 157–175. Springer US, 2012.
- [20] Sazonau, Viachaslau. "Implementation and Evaluation of a Random Forest Machine Learning Algorithm." *University of Manchester, UK*, 2012.

Appendices

Smart-phone application process screen-shot:



REGISTER AN ACCOUNT

pro

nag

35

Female

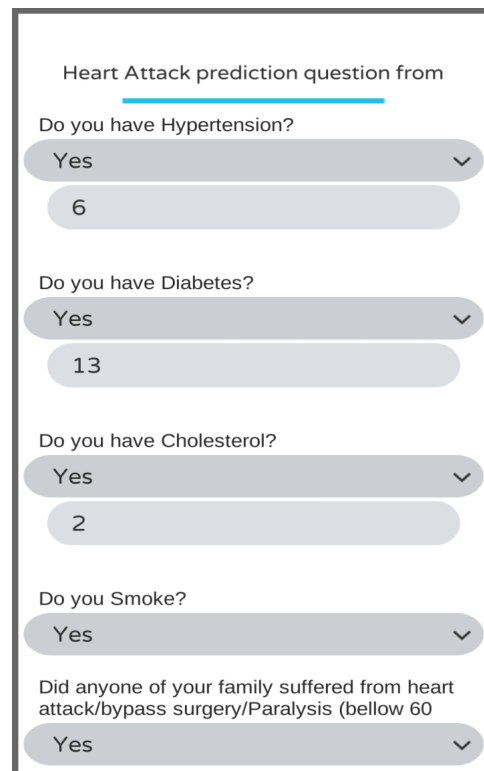
service

01988384900

Register Now →

LOGIN TO YOUR EXISTING ACCOUNT

Figure 9: Registration page.



Heart Attack prediction question from

Do you have Hypertension?

Yes

6

Do you have Diabetes?

Yes

13

Do you have Cholesterol?

Yes

2

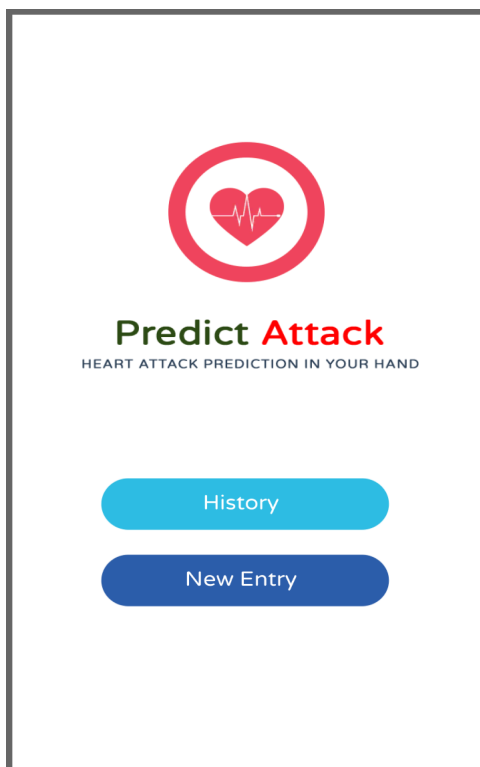
Do you Smoke?


Yes

Did anyone of your family suffered from heart attack/bypass surgery/Paralysis (bellow 60

Yes

Figure 11: Heart attack prediction question set.





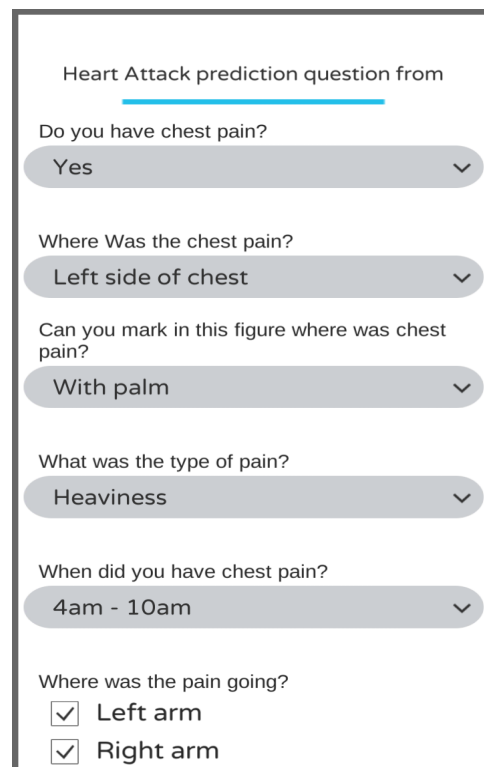
Predict Attack

HEART ATTACK PREDICTION IN YOUR HAND

History

New Entry

Figure 10: Application screenshot.



Heart Attack prediction question from

Do you have chest pain?

Yes

Where Was the chest pain?

Left side of chest

Can you mark in this figure where was chest pain?

With palm

What was the type of pain?

Heaviness

When did you have chest pain?


4am - 10am

Where was the pain going?

Left arm

Right arm

Figure 12: Required questions(continue)

Heart Attack prediction question from 

Back

Upper jaw

No movement

What was it associated with

Dyspnea

Giddiness

Nausea


Palpitation

Sweating

Syncope

Vomiting

What you were doing while it started?

Resting 

Did the pain relived?



No 

Figure 13: Required questions(continue)

RESULT




Heart Attack

You are at high risk of heart attack. It may happen any time if you don't concern with specialist doctor. Take necessary steps As soon as possible and Save a life.

Ok

Figure 15: Result screen.

Heart Attack prediction question from 


Palpitation

Sweating


Syncope

Vomiting


What you were doing while it started?

Resting 


Did the pain relived?

No 

How long it persisted?

1-5 min 

Did you have similar pain / some chest discomfort before some days/ weeks ago?

No 


Calculate Risk 

Figure 14: Submission screen.

USER HISTORY

Date	Result
2018-07-20	Heart Attack
2018-06-20	No heart attack
2018-05-20	Heart Attack
2018-04-20	No heart attack

Go Back

Figure 16: Checked previous result history

Similarity-based Resource Selection for Scientific Workflows in Cloud Computing

Takahiro Koita^{1,*}, Yu Manabe²

¹*Doshisha University, Faculty of Science and Engineering, Japan*

²*NAIST, Department, Division of Information Science, Japan*

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 22 October, 2018

Online: 01 November, 2018

Keywords:

Cloud computing

Amazon web services

Scientific workflows

Resource selection

ABSTRACT

There are high expectations for commercial cloud services as an economical computation resource when executing scientific computing workflows, for which the computation is increasing on a daily basis. However, no method has been developed for determining whether a scientific computing workflow can be executed at a low usage cost, and thus scientists have difficulty in selecting from the diverse range of computational resources. The aim of this study is to provide clear criteria for selecting a computational resource while executing a scientific computing workflow. This study focuses on the performance of application execution for one such commercial cloud service, Amazon EC2, and proposes a method for selecting the optimal resource showing high similarity to a target application in execution time and usage cost. The novelty of this study is its approach of employing application similarity in resource selection, which enables us to apply our method to unknown applications. The contributions of this work include (1) formularizing performance values of computational resources, as well as similarity values of applications, and (2) demonstrating the effectiveness of using these values for resource selection.

1. Introduction

This paper is an extension of work originally presented in ICBDA2018 [1]. Scientific computing workflows [2] are applications that performs a sequence of processes by dividing applications handling scientific computing into small tasks and, by executing these tasks in stages. Figure 1 shows an overview of scientific computing workflows of Epigenomics and Montage. Characteristically, scientific computing workflows can deal with large amount of data, and the work quantity differs for each task. Here, as examples, we introduce three types of scientific computing workflows. Montage is an application developed by NASA that processes celestial images. A feature of Montage is that, since it handles large-sized images, it requires high levels of I/O performance. Broadband is an application that generates a vibration record diagram from multiple earthquake simulations, and requires high levels of memory performance. Epigenomics is an application dealing with DNA, and requires high levels of processing ability based on CPU performance. Thus far, scientific research has mainly consisted of experiments and theories. However, with developments in hardware, advanced calculation

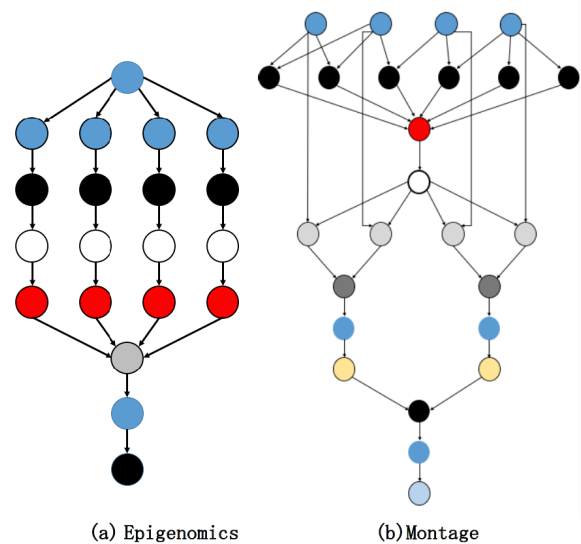


Figure 1. Scientific Workflow

*Takahiro KOITA, Email: tkoita@gmail.com

Table 1. List of Instances

(a) Role list

role	instance name	
General purpose	t2	m4
CPU optimized	c4	
Memory optimized	r3	x1
Storage optimized	i2	d2
GPU instance	g2	
GPU computing	p2	

(b) Performance list

performance			
nano	micro	small	medium
large	xlarge	2xlarge	4xlarge
10xlarge	16xlarge	32xlarge	

Table 2. Instance Performance

	vCPU	ECU	memory (GB)	storage (GB)	cost (\$/h)
t2.micro	1	variable	1	8	0.013
c4.large	2	8	3.75	8	0.105
m4.large	2	6.5	8	8	0.12
m4.xlarge	4	13	16	8	0.239
r3.large	2	6.5	15.25	32	0.166
i2.xlarge	4	14	30.5	800	0.853

has become possible, and computer-based simulations have become an essential new research method. The importance of scientific computing workflows are only expected to grow for future science. Many scientific computing workflows have been developed based on distributed processing using high performance computers (HPC) with grids, PC clusters, and supercomputers, and scientists have used their own PC clusters and grid computing, such as Open Science Grid [3], when executing scientific computing workflows. However, through the development of hardware, the data quantity that can be operated is increasing on a daily basis, and the processing capabilities of computation sources and the storage capacity required are expanding in the same way.

When executing scientific computing workflows, for which the computational data is increasing on a daily basis, the use of commercial cloud services as a computation resource, in place of PC clusters and Open Science Grid, has attracted attention. The commercial cloud service is a service in which scientists can use servers on the network by paying usage fees. Features of such services are that computational sources and storage can be swiftly added, and computational sources of various performance (instances) are prepared. Scientific computing workflows are designed based on distributed processing, and it is possible to execute these in the cloud in which distributed processing is performed through distributed computing. Additionally, as a wide variety of performance instances are prepared, tasks with different processing can be performed in respectively optimized environments. It also has the characteristics of being a measured rate system in which you only pay for the time you use, the fact that maintenance costs are not required, and initial investment for constructing facilities is not necessary. It promises to be applicable to scientific computing workflows, and is to be used as a highly-economical computation source. In this study, we use the commercial cloud service Amazon Elastic Compute Cloud (EC2) [4] used in the preceding research [5]. EC2 is a web service provided by Amazon. The users can select virtual machines, called

instances, according to various purposes. With EC2, five types of roles and multiple respective processing resources are prepared. A list of the instances is shown in Table 1.

One of the important problems with using EC2 is that it is difficult to select the instance and the application to execute with the instance performance table. If the instance performance does not satisfy the performance requirements of the application, execution will be impossible, or the execution time will increase, leading to an increase in usage costs. On the other hand, if the instance performance is higher than necessary, the cost per unit time will be higher, and even if the execution time is shorter, the costs would increase. Currently, when selecting the instance to execute the application, specialized knowledge about applications and cloud or user experience are required. This situation makes it difficult to select a suitable computational resource from a large number of computation resources when considering execution time and usage costs, and this is the problem for scientists using commercial cloud services.

To solve the problem, this study aims to provide a clear criterion for selecting instances for executing scientific computing workflows. Using the provided selection criteria, it will be possible for scientists to casually engage in cloud services, and to perform experiments using advanced computing resources for low research fees. The novelty of this study is its approach of employing application similarity in resource selection, which enables us to apply our method to unknown applications. The contributions of this work include (1) formularizing performance values of computational resources, as well as similarity values of applications, and (2) demonstrating the effectiveness of using these values for resource selection.

2. Current Issues

There are four main issues in executing scientific computation workflow using the cloud, as follows:

- 1) Virtualization overhead
- 2) Low throughput in shared/parallel file systems
- 3) Low network performance
- 4) Unclear usage costs

Issue 1 appears in a significant way when CPU performance is required. Additionally, 2 and 3 clearly have an impact on applications requiring I/O performance [6]. Based on the above features, the current commercial cloud services cannot achieve HPC-equivalent performance. It is expected that issues 1 to 3 shall be resolved on the hardware front, through the development of commercial cloud services. However, this will not solve 4, which occurs when scientists use commercial cloud services. Scientists need to select the computation resources satisfying the processing ability required for the application based on uncertain factors such as their own knowledge and experience. A cause of issue 4 not being satisfied is that there are no criteria for selecting computational resources that consider the necessity of applications [7].

This study focuses on the issue of usage cost. The issue is how to select instances that will satisfy computing performance requirements and have the lowest user cost when executing a scientific workflow on a commercial cloud service. In EC2, the performance of each instance is published, and users can determine the computation resources based on the performance table values. Table 2 shows part of the published performance table. vCPU expresses the number of virtual Server cores, and ECU (EC2 Computing Unit) is a numerical representation of the total processing performance for the instances. In case of an instance where ECU is 8 and vCPU is 2, the CPU processing ability per core is 4. EBS (Elastic Block Storage) is the block unit storage provided by Amazon, and a total of four types are prepared, comprising two types of SSD and two types of HDD [8]. For all instances in this study, the versatile SSD type found in the default settings is used.

Currently, the user selects the instance using their experience, based on the performance capability required for the executed applications and the values of the instance performance table, and it is possible that the instance with the shortest execution time or the lowest usage costs may not be selected.

Table 3. Instance Performance Values

	ECU	EMU	EFU
m4.large	6.50	9.30	9.50
m4.xlarge	13.0	9.62	9.25
c4.large	8.00	8.00	8.00
c4.xlarge	16.0	8.04	9.08
r3.large	6.50	8.48	9.05
r3.xlarge	13.0	8.45	9.05

We explain this situation using the example of a prime number calculation application. Prime number calculation applications are applications that mainly require CPU processing ability. For this reason, it is predicted that the user will select the c4 interface,

which has enhanced CPU performance. However, at that time, they need to decide whether to choose the c4.large with 8 ECU, or the c4.xlarge with 16 ECU. The result of actually executing this was that the execution time was shorter for c4.xlarge, but the usage costs were lower with c4.large. Due to this, until we actually execute the application, it is unclear which instance has the shortest execution time or which has a lower usage costs. Additionally, the processing ability used for prime number calculation examples is virtually CPU only, but with the actual application, memory and I/O processing ability are required at the same time. When selecting the instance, it is important to have a proper understanding of the processing ability required by the application.

To execute the application with a short execution time or a low usage cost, it is necessary to quantitatively grasp the performance ability required by the application and the instance performance and clarify these relationships. Therefore, in the next section, we will quantitatively demonstrate the instance performance and the processing ability required by the application and perform preliminary experiments to provide clear selection criteria.

Very few studies have been made to quantitatively grasp the performance required by the application or the instance. Tovar et al. [9] classified tasks in scientific workflows and proposed an estimation method for the tasks. They showed that the execution time can be estimated and that CPU, memory and I/O performance indexes are important for this estimation. Sfiligoi et al. [10] showed the characteristics of scientific workflows statistically, and the results were effective for their experiment’s applications. However, these studies are useful only for known applications whose behavior information can be given well in advance of instance selection. Consequently, if such information is insufficient, these studies cannot be applied. Thus, the current study employs several values to achieve resource selection for unknown applications. Furthermore, previous studies assumed that their target instance was a single type and thus did not consider the various types of instances in commercial cloud services.

3. Preliminary Experiment

We perform preliminary experiments to quantitatively show the instance performance and processing ability required by the application.

3.1. Instance Performance Value

We describe the instance performance as numerical values. We focus on instance performance in terms of CPU, memory, and I/O. This is because CPU, memory, and storage are enhanced respectively in EC2, and because the instances are mainly prepared in relation to these, it is assumed that these will have the greatest impact on execution time and usage cost. The CPU processing ability uses ECU, published by Amazon. In this study, memory and I/O processing ability are defined respectively as EMU and EFU, and these are measured and expressed numerically in these preliminary experiments. In these preliminary experiments, the versatile instances m4.large and m4.xlarge, the CPU optimization instances c4.large and c4.xlarge, and the memory optimization instances r3.large and r3.xlarge are used. The performance of each of these is shown, respectively, in Table 3.

Table 4. Execution Performance Values
(partial result only for two applications in [11])

	execution time [sec]	CPU	memory	I/O
Memory bound apl.	291.6	44.9	31.4	30.7
I/O bound apl.	25.6	3.94	2.75	2.70

Measurements are performed using a program prepared for the purpose of measuring performance evaluations. Memory performance evaluations involve reading and writing memory multiple times, whereas I/O performance is assessed by reading and writing a text file multiple times. The respective execution times are measured, with the ratio with the c4.large value and 8, which is the same as ECU, to discover the EMU and EFU of each instance. The respective instance performance is shown in Table 4.

3.2. Execution Performance Value

In this experiment, the processing ability required by the application expressed as a numerical value is used as the execution performance value. The execution performance value shows the effect of the CPU, memory, and IO required for the application on the execution time. As an example, we shall show a formula for obtaining the execution performance value based on CPU performance. The performance values based on memory or I/O performance can be calculated by making the respective ECU values the EMU and EFU values according to the following formula.

Here, as an example, we shall seek the two application execution performance values of memory performance evaluation and I/O performance evaluation used when evaluating instance performance. The execution performance values need to actually be executed with the instances. In this preliminary experiment, measurement was performed using the versatile instance m4.large. The execution time and execution performance values based on the CPU, memory, and I/O performance are as follows.

From the results, we can see that the size of the execution performance values changes depending on the execution time, and that differences appear in the execution performance value ratios based on the processing content. The two applications used in this preliminary inspection have high execution performance values based on CPU performance and, as with the instance performance evaluation, the execution time for both applications was shorter with the c4 instance optimized for CPU performance, by referencing the execution performance values, it is possible to grasp the processing capability required by the application.

4. Proposed Method

We propose a method for selecting resources that uses an application with similar execution performance to select the instance that can run an application in the lowest time or with lowest usage cost. In the past, the run time and usage cost were unknown before actually running an application, and there was a risk of costs increasing when the application was run several times. The proposed method enables a resource to be selected, running

the application a minimum number of times, by considering the application execution performance values and instance performance.

We expect that if the performance required by two applications is the same, the computing resources required for the shortest execution time or lowest usage cost will be the same for them as well. The proposed method selects an application with similar execution performance values as the application in question from among several that have been run in the past, and selects the computing resource able to execute the application in question in the shortest time or at lowest cost. The method is comprised of the following four steps.

- 1) Standardization Step: Measure the execution performance of the sample applications
- 2) Measurement Step: Measure the execution performance of the application in question
- 3) Comparison Step: Select a sample application with similar execution performance values
- 4) Selection Step: Select the instance with shortest execution time or with lowest usage cost

Details of each step are described below.

Standardization Step: The sample applications are executed on each instance, and the instances producing the shortest run time and lowest usage cost are selected. The execution performance is also computed using an arbitrarily selected instance. Several applications performing different processes are used as sample applications for the proposed method.

Measurement Step: This step deals with the application for which a computing resource is being selected. The application is executed on an instance selected in the standardization step to measure its execution performance values.

Comparison Step: In this step, the execution performance values of all sample applications measured in the standardization step are compared with the execution performance values of the application in question, as measured in the measurement step. For this method, a similarity level is used for this comparison. The similarity level is expressed as a distance between the execution performance values of the two applications. The normalized execution performance values of CPU, memory and I/O of application A are denoted E_{Ae} , E_{Am} , and E_{Ai} , respectively. These execution performance values, are obtained by the preliminary experiment described in Section 3. Similarly, the execution performance values for application B are denoted E_{Be} , E_{Bm} , and E_{Bi} . The similarity, D_{AB} , is given by the following equation (1). This equation uses Euclidean Distance between applications A and B . If the similarity value is sufficiently high, they are considered similar applications. The highest similarity value is thus used to select the most similar application. To select a resource, the target application A is fixed while application B varies. That is, the distances to application A from all other applications are calculated. The distance is determined by the values of execution time, memory, and I/O described in the previous section. More details can be seen in an earlier work [11].

$$D_{AB} = \sqrt{(E_{Ae} - E_{Be})^2 + (E_{Am} - E_{Bm})^2 + (E_{Ai} - E_{Bi})^2} \quad (1)$$

Selection Step: In the selection step, the similarity of the application with each of the sample applications is computed. The sample application with the smallest similarity level is selected as the one that is most similar to the application in question. The instance able to execute this most-similar sample application with the shortest execution time or lowest cost is then selected as the instance to run the application in question.

5. Evaluation

To show that instances can be selected based on similarity of execution performance values, we conducted experiments to evaluate the effect of similarity on the execution time and usage cost.

5.1. Experiment Overview

Using the same instances as in the preliminary experiments, the execution time, usage cost and similarity were measured for multiple applications. If the instances running the applications with the lowest execution time and usage cost are the same for applications that are similar, the proposed method will select resources correctly. The instances used were the same as those used in the preliminary experiments. The test procedure was as follows:

- 1) Run applications
- 2) Calculate execution performance values
- 3) Calculate similarities
- 4) Select similar applications.

Each of these steps is described in more detail below.

Run applications: Multiple applications were run on each instance, and the execution time was measured. From the execution times for each instance, the instances producing the shortest execution time and lowest usage cost were selected for each.

Calculate execution performance values: Each application was run on an arbitrarily selected instance and the execution performance values were measured. For these tests, we used the m4.large general-purpose instance.

Calculate similarities: Here, the computed execution performance values were normalized, and the similarity to all of the other applications was computed for each application.

Select similar applications: For each application, the one with the lowest similarity level was selected as the most similar application. Each application was compared with the other application most similar to it, and we checked whether the instances producing the shortest execution time and usage cost were the same.

5.2. Applications

The applications used here include Sysbench [12] and UnixBench [13], which are comprehensive benchmark applications, and Hadoop [14], which is a distributed framework. These are described in more detail below.

Sysbench is a general benchmark application for Linux/Unix operating systems. Sysbench has six types of evaluation (e.g. CPU or memory) and enables each of them to execute with adjusting application parameters such as the number of CPUs or the file size. For our experiments, we performed CPU, memory and I/O tests. Prime numbers are computed for the CPU test, reading and writing to memory is done for the memory test, and reading and writing files to storage is done for the I/O test. Each test was done repeatedly and the execution times were measured. The term of test means a specific execution to perform one type of evaluation.

UnixBench is a benchmark application used with Unix-type operating systems. The test covers a variety of tasks from integer arithmetic through to OS system calls. In these experiments, benchmarks for integer computation (Dhrystone), floating-point computation (Whetstone), and file copying (fsdisk) were performed. Results are given in terms of processing capability per unit time. These were converted to results in terms of a time required to complete a fixed-size process for these experiments.

Hadoop is a distributed framework that enables multiple computers to be treated as a single computer with improved performance. Hadoop can be used in any of three modes: stand-alone mode, which runs on a single CPU, pseudo-distributed mode, which virtualizes use of two machines on a single machine, and fully-distributed mode, which uses multiple computers. For these experiments, we used pseudo-distributed mode, measuring execution time for standard sample processes including computing pi, counting words, and sorting files.

5.3. Results

Each application was executed on each of the instances. For some of the applications, such as Dhrystone in UnixBench, the results are given in number of loops per second rather than total execution time. In such cases, the results were converted to a time required to perform a set number of loops. Execution times and usage costs are given in Tables 4 and 5, and execution performance values, as discussed previously, are given in Table 6.

Instance c4.xlarge had the shortest execution time, and instance c4.large had the lowest usage cost in most cases. Execution performance values were calculated using the execution times and the ECU, EMU, and EFU performance values for m4.large, and these were then normalized.

Similarities were then computed using these values. Below, we give an example of computing the similarity, D_{pw} , is given by equation (2) using the execution performance values from the prime number computation in Sysbench and the word count process on Hadoop.

$$D_{pw} = \sqrt{(0.038 - 0.020)^2 + (0.047 - 0.025)^2 + (0.047 - 0.025)^2} \quad (2)$$

Similarities were computed for all process pairs, and that with the smallest similarity value was designated as the similar application for each application. Table 8 shows whether the instances producing the shortest execution time and lowest usage cost were the same for these similar applications.

The process most similar to the prime number computation on Sysbench was word count on Hadoop. The instance with the shortest execution time for the prime number process in Sysbench was c4.xlarge, which was the same as for the Hadoop word count,

Table 5. Execution Times [sec]

application		execution time					
		m4.large	m4.xlarge	c4.large	c4.xlarge	r3.large	r3.xlarge
Sysbench	prime number	157.51	78.99	128.68	64.32	153.31	76.47
	memory read	42.55	38.61	35.79	33.72	40.62	36.39
	memory write	52.32	41.82	43.63	36.96	49.90	40.09
	random read/write	2.07	0.42	1.71	0.78	1.51	1.00
	sequential read/write	34.42	22.90	34.36	22.90	38.14	24.60
hadoop	pi	3328.29	1692.45	2828.35	1438.95	3339.85	1698.05
	word count	83.14	54.62	77.71	50.90	88.76	56.86
	file sort	40.86	36.68	40.00	36.36	42.82	36.55
UnixBench	Dhrystone	0.76	0.85	0.64	0.32	0.74	0.37
	Whetstone	0.34	0.73	0.35	0.17	0.41	0.20
	fsdisk	0.71	0.76	0.68	0.83	0.95	1.14

Table 6. Usage Costs [\$]

application		usage cost					
		m4.large	m4.xlarge	c4.large	c4.xlarge	r3.large	r3.xlarge
Sysbench	prime number	21.89	21.96	16.21	16.21	30.66	30.51
	memory read	5.91	10.73	4.51	8.50	8.12	14.52
	memory write	7.27	11.63	5.50	9.31	9.98	15.99
	random read/write	0.29	0.12	0.22	0.20	0.30	0.40
	sequential read/write	4.78	6.37	4.33	5.77	7.63	9.81
Hadoop	pi	462.63	470.50	356.37	362.62	667.97	677.52
	word count	11.56	15.18	9.79	12.83	17.75	22.69
	file sort	5.68	10.20	5.04	9.16	8.56	14.58
UnixBench	Dhrystone	0.11	0.24	0.08	0.08	0.15	0.15
	Whetstone	0.05	0.20	0.04	0.04	0.08	0.08
	fsdisk	0.10	0.21	0.09	0.21	0.19	0.46

Table 7. Execution Performance Value

application		performance value		
		CPU	memory	I/O
Sysbench	prime number	0.03837	0.04725	0.04725
	memory read	0.01030	0.01270	0.01270
	memory write	0.01269	0.01564	0.01564
	random read/write	0.00042	0.00054	0.00054
	sequential read/write	0.00832	0.01026	0.01026
Hadoop	pi	0.81248	1.00000	1.00000
	word count	0.02021	0.02490	0.02490
	file sort	0.00989	0.01220	0.01220
UnixBench	Dhrystone	0.00010	0.00015	0.00015
	Whetstone	0.00000	0.00002	0.00002
	fsdisk	0.00009	0.00013	0.00013

Table 8. Matching Instance and Result of Similarity-based method

application		best time	best cost	similarity application	best time instance	best cost instance
Sysbench	prime number	c4.xlarge	c4.xlarge	word count	Same	Different
	memory read	c4.xlarge	c4.large	file sort	Same	Same
	memory write	c4.xlarge	c4.large	memory read	Same	Same
	random read/write	m4.xlarge	m4.xlarge	Dhrystone	Different	Different
	sequential read/write	m4.xlarge	c4.large	file sort	Different	Same
Hadoop	pi	c4.xlarge	c4.large	prime number	Same	Different
	word count	c4.xlarge	c4.large	memory write	Same	Same
	file sort	c4.xlarge	c4.large	memory read	Same	Same
UnixBench	Dhrystone	c4.xlarge	c4.xlarge	fsdisk	Different	Different
	Whetstone	c4.xlarge	c4.large	fsdisk	Different	Same
	fsdisk	c4.large	c4.large	Dhrystone	Different	Different

so the instances with the shortest execution time matched. On the other hand, the instance able to run the Sysbench prime number process at lowest cost was c4.xlarge, while for Hadoop word count, it was c4.large, so the lowest cost instances did not match. This result appears on the first row of data in Table 7. From left to right, it indicates that for the prime number process in Sysbench, the similar application was Hadoop word count, that the instances with shortest execution time matched, and that the instances with lowest usage cost did not match.

6. Discussion

We now discuss the results of these evaluation experiments. Instances for which execution with short execution time and low usage charges are possible tended to be c4.xlarge and c4.large, respectively. The cause of this is that many of the applications used in this experiment were CPU-bound, and this is considered to have had a major impact on the match rate. In particular, the prime number calculation by Sysbench and the performance required for Dhrystone in UnixBench is biased toward the CPU. As the processing for these had the shortest execution time and lowest usage costs in c4.xlarge, which is optimized for the CPU, this is a result compatible with the published ECU values. For the scientific computing workflow, processing differs depending on the task, and there are tasks that require a lot of non-CPU processing. For that reason, in the reading and writing of memory for Sysbench, which is an application that requires not only CPU, but also memory and I/O processing performance, and file sorting by Hadoop, execution time became shorter due to CPU performance. CPU performance is important even for applications requiring memory and I/O processing performance; therefore, creating a calculation formula that is weighted in consideration of the impact of each on execution time and usage cost is effective when selecting resources.

To apply this method to the scientific computing workflows carrying out a variety of processing, it is necessary to increase the number of sample applications handled and support a more diverse range of processing. Additionally, as the sample applications used in this test have a low computational volume, there is a concern that it cannot support scientific computing workflows handling huge volumes of processing. A greater diversity of sample applications is required to realize this method and enable the selection of computational resources in scientific computing workflows.

7. Conclusion

The objective of this experiment is to achieve a method of selecting resources based on execution time and usage costs when using commercial cloud services for scientific computing workflows. By using instance performance and the execution performance values required for the application, we measured the features of the application for a certain instance. The aim is to propose a method for selecting instances that can be executed in the shortest execution time with the lowest usage costs, by referencing similar applications for the measured execution performance values. In the evaluation experiment, we verified the effectiveness of selecting resources based on similarity. The match rate of the results was approximately 55%, and a large impact was present in CPU-bound applications. As this considers the impact on execution time when making resource selections

more than for memory or I/O, it is necessary to focus on CPU performance.

Our experiment showed that the proposed method based on similarity can usually select the best instance for Hadoop or Sysbench-type applications. Furthermore, scientific computing workflow applications are mainly executed using few system functions, like Hadoop and Sysbench-type applications. Thus, our method would be effective in selecting resources for many types of scientific computing workflows. On the other hand, if the application requires many system functions, like UnixBench, the similarity calculation requires weighting factors to handle complicated behavior.

References

- [1] T. Koita Performance Evaluation of Memory Usage Costs for Commercial Cloud Services, Proc. of the IEEE 3rd Int'l Conf. on Big Data Analysis (ICBDA2018), pp.307-311, 2018.
- [2] Ewa Deelman, Pegasus and DAGMan From Concept to Execution: Mapping Scientific Workflows onto Today's Cyberinfrastructure, Proc. of the Advances in Parallel Computing, vol.16, pp.56-74, 2008.
- [3] Open Science Grid, <https://www.opensciencegrid.org/>.
- [4] Amazon Elastic Compute Cloud (EC2), <http://www.amazon.com/ec2/>.
- [5] Y. Manabe, Performance comparison of scientific workflows on EC2, IPSJ technical report, 2016.
- [6] S. Ostermann, A Performance Analysis of EC2 Cloud Computing Services for Scientific Computing, Proc. of the Cloud Computing, pp.115-131, 2010.
- [7] G. Juve, Scientific workflows and clouds, Crossroads, vol.16, pp.14-18, 2010.
- [8] G. Juve, Scientific workflow applications on Amazon EC2, Proc. of the 5th IEEE International Conference on e-Science Workshops, pp.59-66, 2009.
- [9] B. Tovar, A Job Sizing Strategy for High-Throughput Scientific Workflows, IEEE Trans. On Parallel and Distributed Systems, vol.29, no.2, pp.240-253, 2018.
- [10] I. Sfiligoi, Estimating job runtime for CMS analysis jobs, Proc. of J. Physics: Conf. Series, vol. 513, no. 3, 2014.
- [11] Y. Manabe, Resource provisioning method for scientific workflows on commercial cloud services, graduation thesis, Doshisha University, 2017.
- [12] Sysbench, <http://imysql.com/wp-content/uploads/2014/10/sysbench-manual.pdf>, 2009.
- [13] UnixBench, <http://code.google.com/p/byte-unixbench/>.
- [14] Hadoop, <http://hadoop.apache.org>.

Visualizing Affordances of Everyday Objects Using Mobile Augmented Reality to Promote Safer and More Flexible Home Environments for Infants

Miho Nishizaki*

Department of System Design, Tokyo Metropolitan University, 191-0065, Japan

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 20 October, 2018

Online: 01 November, 2018

Keywords:

Augmented Reality (AR)

Affordance

Visualization

Accident

Development

Infant

ABSTRACT

This study presents a prototype augmented reality (AR) application that helps visualize the affordances of everyday objects for infants in their home environments to prevent accidents and promote development. To detect and visualize affordances, we observed 16 infants, 4 to 12 month of age, on how they perceived and handled common objects in their homes in Tokyo, Japan and in Lisbon, Portugal. Based on the longitudinal data, we developed an AR application for handheld devices (iPhone and iPad) and tested two types of vision-based markers. Ten types of basic objects were selected from the results of the observation and embedded into the AR markers. AR contents illustrated infants' actions toward objects based on actual video data recorded for security purposes. To confirm the prototype's advantages and improvements, informal user interviews and user tests were conducted. The results demonstrated that the prototype reveals the relationship between infants and their home environments, what kinds of objects they have, how they perceive objects, and how they interact with these objects. Our study demonstrates the potential of this application's AR contents to enable adults to better understand infants' behavior towards objects by considering the affordances of everyday objects. Specifically, our app assists in improving the perspective of adults who live with infants and promotes the creation of more flexible and safer environments.

1. Introduction

1.1. The Environment and Augmented Reality

Augmented reality (AR) technology enables us to visualize our living environment in novel ways. In 1968, Ivan Sutherland created the first AR experience generated using a head-mounted display [2], [3][†]. In the early 1990s, the term “augmented reality” was coined with the first technological development underlying augmented devices [4], [5]. Azuma described AR as systems with three technical components: (1) use of real and virtual elements, (2) real-time interactivity, and (3) 3D registration [6]. The third characteristic “3D registration,” refers to the system’s ability to anchor virtual content in the real world, that is, in a part of the physical environment [7].

AR enhances user perceptions of and interactions with the real world by seamlessly merging virtual content with reality [5], [7]. Whereas virtual reality technology replaces reality by creating a whole immersive virtual environment, AR enhances the reality of

an experience without changing it by projecting partially virtual/digital content in a non-immersive way.

Current AR devices require four elements: displays, input devices, tracking, and computers. Displays include head-mounted, handheld, and spatial displays. Mobile AR systems use handheld displays such as cellular phones, smartphones, and tablets. The first mobile-phone-based AR application was demonstrated in 2004 [8], and the mobile AR market has since grown consistently [6], [9], [10–12]. Such applications became well established with the advent of Pokémon Go in July 2016. This successful location-based mobile game required users to find virtual characters in the real world, and it immediately became an international sensation. It infiltrates users’ physical lives and changes their locomotor behavior with potential effects to their health. Pokémon Go was launched for the Apple Watch in December 2016 and helped further promote health effects. Even now, in 2018, this game remains popular [13].

AR application development for mobile devices has become a large and fast-growing area as AR glasses and head-mounted displays are being improved further. The AR market is valued at \$83 billion, whereas the VR market is valued at only \$25 billion

*Miho Nishizaki, Department of System Design, Tokyo Metropolitan University, 191-0065, Japan. [†]This paper is an extension of work originally presented in 2017 Intelligent Systems Conference (IntelliSys) [1].

[14]. Azuma suggested the following potential application areas for AR: medical visualization, equipment maintenance and repair, annotation, robot path planning, entertainment, and military aircraft navigation and targeting [7]. Thus far, AR has been used in fields such as architecture, clinical psychology, cognitive and motor rehabilitation, education, and entertainment. AR has become popular; nonetheless, more research is needed on how to use it to better living experiences. The use of AR could be increased by improving device and recognition technologies. However, few studies and applications of AR currently exist for solving everyday issues.

1.2. Accidents and Developments in Home Environments

Young children spend most of their time at home while they grow up. Although most people think of accidents as things that happen outside the house, accidental injuries among infants and young children at home have been a major issue worldwide. For example, 14.5% of aged 0–4 years children’s deaths are caused by home accidents in Japan [15]. Similarly, in the UK and the USA, the 0–4 year age group is at the highest risk of home accidents and the accidents at home are the leading cause of deaths [16–18]. Therefore, many studies have explored how to prevent accidents at home [19–23]. However, suitable methods for collecting actual data at home and for conducting quantitative and qualitative analyses have not yet been established.

Most injury alerts are based on hospital reports [24–27]. The Japan Pediatric Society publishes “Injury alerts” since 2008 and “Follow-up articles” since 2011 based on members’ reports on their website. Although such information is useful and important, it is often incomprehensible to ordinary people. Therefore, such information needs to be presented clearly and be accompanied by detailed explanations. Several studies have highlighted that accidental injuries suffered by children at home are preventable [21–24]. This study uses an AR system to visualize and simulate infants’ behavior in a concise and clear manner to provide caregivers of children with insights into their own living environments.

1.3. Visualizing Possibilities-Affordances

It is too difficult to protect infants from all dangerous possibilities. At the same time, just to prevent accidents, children may be banned from exploring various things at all stages of development. It is important to avoid overly protective behavior or overcontrolling children. Therefore, we employed James Gibson’s theory of affordance [28,29] to deal with infants’ unexpected behaviors at home during growth. The affordance is an essential concept of ecological approach; it refers possibilities or opportunities for behavior that the environment offers an animal. Gibson proposed that an affordance is a fact in the environment as well as a fact embodied by an act. It implies the complementarity of the animal and the environment [29]. Thus, the affordances of the home environment in which we spend most of the time and our way of life are inseparable. Furthermore, according to Gibson, “the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill [29].”

Perceiving the information of affordances is not the same even if people share the same objects or space. Although adults know the way to use objects at home as designed, infants explore every

object and find a variety of possibilities. We do not have to know the name of the object or how to use it to perceive what it affords. Empirical studies on individual affordance have been advanced in a series of pioneering studies since 1960 [30–34]. However, researchers have defined affordance in a variety of ways. The generalization methods are still being investigated as they have not yet been thoroughly established [35–39]. Experiments on infants’ natural activities have been conducted in a laboratory playroom [40–42]; however, these kinds of studies are few in number and little is known about everyday home environments. Therefore, the current study aims to present a method adaptable to the home environment that would help visualize resources to support healthy growth of children from a variety of aspects.

This study has three main objectives: First, to identify infants’ unique behaviors in home environments by the longitudinal observations during the first year; second, to examine data for two countries whether there is a difference or not; and third, to present an AR mobile application prototype that enables users to simulate infants’ behaviors toward objects that assists caregivers to understand intuitively and hence provide better environments for infants.

2. Methods

2.1. Participants

We recruited ten healthy infants (6 males, 4 females) in Japan and six healthy infants (4 males, 2 females) in Portugal from middle or upper middle-class families for the longitudinal observations. In Japan, eight Japanese families and one Hungarian family were from the Tokyo area. In Portugal, all Portuguese families were from the Lisbon area.

2.2. Procedure

Observation. Longitudinal observations were conducted to clarify how infants behave at home during their first year. We observed the activities of 16 infants at their homes from 4 to 12 months of age. An experimenter or infants’ caregivers recorded the infants’ behaviors with digital video cameras (EX-ZR500; Casio Corp) for over 60 min per month. The recordings needed to cover several natural behaviors (e.g., it is preferable to not film only one activity such as napping). When caregivers recorded the videos, they were asked to send the recorded video data to us every month via a mail or file transfer service. Although parts of monthly data from both countries were unavailable because of family reasons, we collected 122 h of digital videos for Japan and 53 h for Portugal in total.

Design and Development. This process consisted of four steps: (1) sketch and paper prototyping; (2) design of an iOS application using Adobe Illustrator CC and Photoshop CC; (3) development of the iOS application using OpenCV 2.4.9, Objective-C, and Swift 4.0; and (4) creation of AR movies based on videos from the observations using Adobe Photoshop CC and Adobe Lightroom CC. All AR movies were illustrated manually as line drawings to protect privacy and make it easy to zoom in infant–object interactions. More than three drawings were needed for every second and each playtime is less than 60 s. Line drawings were digitized using Lightroom and Photoshop. Every image was saved at 448 × 336 pixel resolution and combined and exported as

QuickTime movies. An example of line drawings of a cabinet with an infant (11 months, 17 days) is shown in Figure 1.

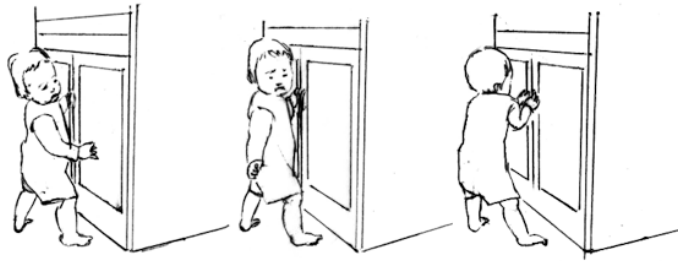


Figure 1. Example of line drawings (cabinet).

2.3. Data Coding

Two coders scored the infants' attribute, house style, behaviors, postures, milestones, ages, places, durations, objects/people (if infants directly touched), and surfaces from digital video recordings by using the Datavyu coding software [43], with reference to the caregiver's comments or notes.

2.4. Ethical Considerations

All procedures for this study were approved by the relevant institutional review boards in consideration of ethical standards for research on human subjects. Informed consent was obtained from all families.

3. Results

3.1. Reality from Observation

A prototype AR application was developed to enable the simulation of children's actions in the first year based on recorded data from observations. We observed infants' spontaneous interactions between surrounding objects at their home from the recorded data. Two infants (one Japanese and one Portuguese) were excluded from the analysis because the data for these babies were not enough 9 months.

Number of Objects at Home. Infants' interactions with surrounding objects were counted using the observational data. All objects that infants chose to interact with met the following two requirements: 1) Infants directly touched the object by him/herself, and 2) objects observed at infants' homes were items that were used daily, such as furniture.

On the whole, Japanese infants, between 4 to 12 months of age, encountered 292 types of objects at their homes. As for Portuguese infants, 177 types objects were observed at their homes. Figure 2 shows the kinds of objects that exist and how many objects exist in each country. According to the ecological psychologist James Gibson's definition of the term objects, in this study, we categorized the two types of objects: detached objects and attached objects [29]. A detached object can be displaced from the surface while an attached object cannot. We identified 262 detached objects including furniture, daily used items, toys, and 30 attached objects including enclosures/corners, aperture, steps, and concave/convex objects, in our recordings in Japan. Similarly, in Portugal, we observed 153 detached objects and 24 attached objects. A number of differences were found in the detached objects, especially daily used items and toys. Japanese infants

could reach more stationary items (e.g., crayons, erasers, scales, pencils, pens, drawing boards, papers, and notebooks), and kitchen items (e.g., cups, dishes), laundry items (e.g., clothespins, hangers, laundry basket), and room conditioner (e.g., thermometer, humidifier, hot water bottle, heater, stove) than Portuguese infants. Regarding toys, more handmade toys, digital video games, and character items were observed in Japanese homes than in Portuguese homes.

Figure 3 shows the percentage of the number of objects that infants interacted with during the observation period in each country. The detached objects, especially the daily-used objects, were observed most frequently among all categories in both countries. It suggested that the tendency of interaction with objects of these categories is similar even though the country is different.

Based on this result, we selected ten objects which were common, continuous, and frequently-observed, for this prototype AR application. Detached objects included cabinet, chair, cushion, futon/mattress, sofa, and table. Attached objects included bathtub, door, threshold, and wall.

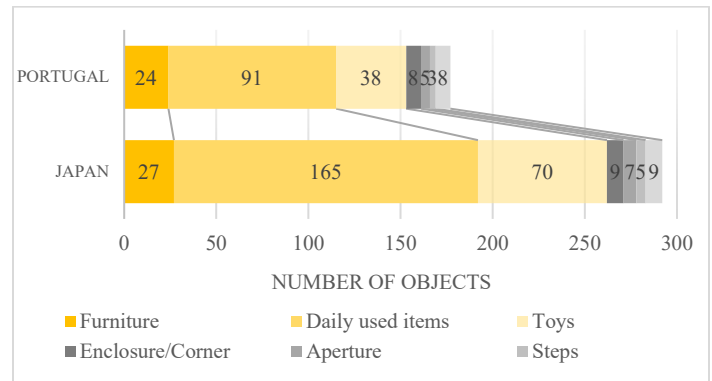


Figure 2. The number of objects that infants directly touched at home in each country. The yellow bars indicate detached objects and the monochrome bars indicate attached objects.

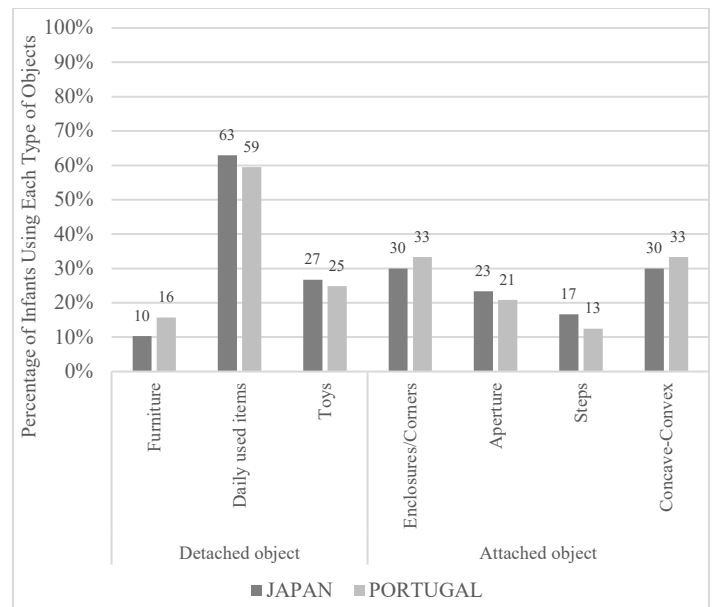


Figure 3. Percentage of each object-category in each country.

Actions and Perceptions. The results of the observations showed infants' spontaneous activities involving objects in their homes. It combines both developmental possibilities and accident risks. To clarify how often every action occurred or were supposed to occur, we counted the number of possibilities of development, accidents, and both accidents and development from all recordings. In this case, we measured the possibilities because we attempted to protect the infant from the accident before accidents occurred during observation. Therefore, the moments also applied to such cases: Parents say "watch out" to the infant or supported the infant's posture before falling.

The number of sessions were 1771 in total ($M = 196.78$, $SD = 110.17$) for Japanese infants, and 910 ($M = 182$, $SD = 117.64$) for Portuguese infants. The conditions of three types of selections are as follows. Accidents were linked to injury-inducing behavior such as losing balance (slip, trip, and fall down), drinking or eating harmful substances/objects, and touching something dangerous for infants. Demonstrations of preventative actions by people in the immediate vicinity before the occurrence of accidents are also included. This includes the same aforementioned actions. Development is associated with motor, perceptual, and cognitive progress to a greater degree than before. Both accident and development are recognized in the same session.

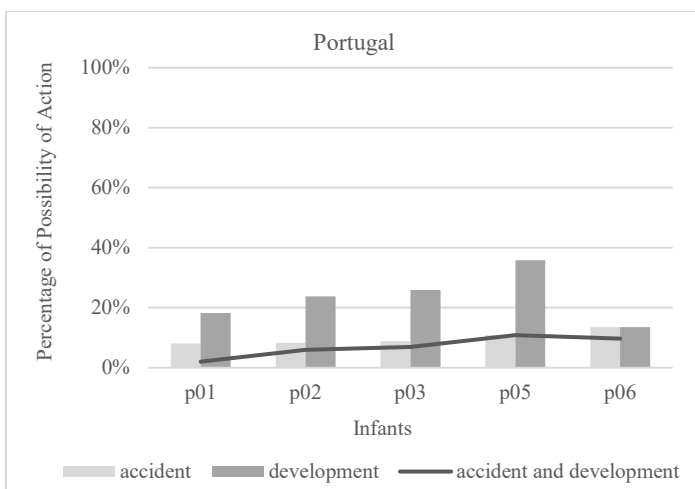
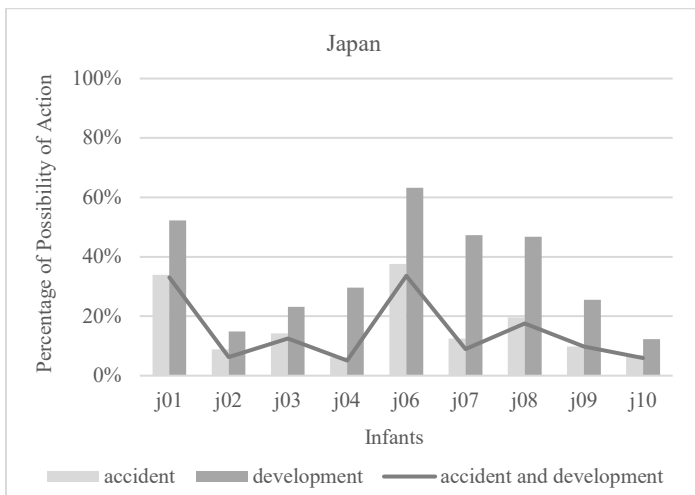


Figure 4. Percentage of action possibilities related to development, accidents, and both in each country per infant.

In both countries, for all sessions, the number of developments was higher than accidents and the combination of accident and development. Figure 4 shows the percentage of action for possibilities of development, accidents, and both in every infant's total sessions per country. The values for accidents and both accidents and development were close in most cases. It indicates that the accidents tend to involve developmental issues that might have affected the situation.

There were over 30% of possibilities of accidents for two infants (Japan: j01 and j06). Common features in these incidents included the detached objects, furniture, and daily used items in both countries. These objects were adapters, cords of light/AV equipment/laptop PC/phone, pencils on tables, cream of tube package, washing machines, bathtubs, sinks, fence, tripods, strings, steps, low tables, cabinets, and doors. However, the only observed baby products that caused accidents were the walker and the high chair. Needless to say, the accidents are related to motor developmental changes; however, the results suggest that many normal products that are designed for a particular purpose were perceived differently by infants, and this difference may cause accidents (e.g., dressers, cabinets and washing machines sometimes offer opportunities to touch, hang on, climb on, and enter inside, for infants. However, there are a number of accidents related to washing machines in particular [44]).

3.2. AR application and vision-based tracking

Vision-based tracking is defined as registration/recognition and tracking approaches [5, 6]. In this study, two types of vision-based marker codes were examined (Figure 5). In Figure 5 (a), OpenCV, MATLAB, and Illustrator, are used for compounding dots and object icons. Figure 5 (b) uses Vuforia and Illustrator for combining textures and icons. Both markers can be attached to real-world objects that can be printed using a printer.

We measured the tracking times to determine if there is a difference between two types of markers (Figure 5) along the two directions of surface (horizontal and vertical) under two lighting conditions (A: Fluorescent light, B: Flat-surface Fluorescent light.). Tables 1 and 2 show the relation between velocity and illuminance of the two directions of surface and the distance from marker to device. To confirm the accuracy of the marker recognition related to two different light condition, we used an illuminance meter (KONICA MINOLTA T-10). Differences between two markers at distances greater than 30 cm, regardless of light conditions, were observed (Table 1 and Table 2). Marker (b) indicates more unstable properties than marker (a) in both conditions. Moreover, additional differences were found between the two orientations of the surface under the lighting (B). Although the illuminance was sufficient to capture an AR marker, the tracking time was slightly longer for vertical surface than for horizontal surface. This suggests that surface orientation affects tracking in the same manner as other AR marker systems.

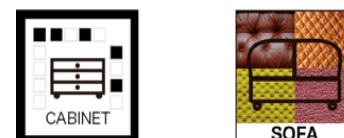


Figure 5. Examples of AR markers. Left: dot-type marker (a), Right: texture-type marker (b).

Table 1. Change of velocity related to illuminance and distance between marker and device under fluorescent light (A).

Type of Marker	Surface	lx	20cm	30cm	40cm
(a)	vertical	289	3 sec	4-5 sec	3 sec
	horizontal	330	3 sec	4-5 sec	3 sec
(b)	vertical	290	5 sec	—	—
	horizontal	300	5 sec	—	—

Note. Fluorescent light's product # is FPL36EX-N, 4 tubes × 20, 2900 lm/tube.

Table 2. Change of velocity related to illuminance and distance between marker and device under flat-surface fluorescent light (B).

Type of Marker	Surface	lx	20cm	30cm	40cm
(a)	vertical	916	3-7 sec	4 sec	3 sec
	horizontal	865	2 sec	2 sec	3 sec
	vertical	427	4-5 sec	4 sec	4-6 sec
	horizontal	330	1-2 sec	1-2 sec	1-2 sec
	vertical	57.5	3 sec	3-5 sec	3 sec
	horizontal	53	2-3 sec	2-3 sec	3 sec
(b)	vertical	916	8 sec	—	—
	horizontal	865	5 sec	—	—
	vertical	427	3 sec	—	—
	horizontal	330	3 sec	—	—
	vertical	57.5	5 sec	—	—
	horizontal	53	3 sec	—	—

Note. Flat-surface fluorescent light's product # is ELF-554P, 4 tubes × 1, F55bx/Studiobias32, 4100 lm/tube, 3,200 K.

Overview. Figure 6 depicts an overview of how to use the prototype application. First, in order to watch AR movies, the user needs to put markers on objects at his/her home. After starting the application, the user can tap the camera icon. Once tapped, the camera window appears and the user capture the AR marker in a manner similar to taking a photograph. When the camera recognizes the marker, a pop-up showing the list of movies according to the age in months is shown. As the user select the movie, AR movie begins to play.

Every marker contains approximately 10 AR movies that illustrate the interactions of infants between 4 and 12 months of age, with objects. The prototype enables the user to simulate how infants interact with the environment and know the difference according to the objects and infants' age. Furthermore, users can also learn about infants and objects. We introduced a notification window of an example for positive and negative tips in Figure 7. Users can find further information as they tap the object icon. In the case of risk information, it is composed with recorded conversations from the data, comments from caregivers, and links to websites.

Figure 8 shows another example of attached objects for further references. This small dictionary was implemented in the application to serve as an introduction for the application users. The users were able to select the content to understand objects at home from the perspective of infants from the floor plan, attached objects, and detached objects.

3.3. User Experience

To determine whether our application is functional, informal user interviews and tests were conducted with 142 people from six countries (Italy, Japan, Portugal, Spain, UK, and USA) from 2013 to 2017 as part of testing phase for the prototype. The application has been improving since 2013 with the addition of longitudinal observation data and updating functions. The interviewees were spread across 18 fields and nine occupations, and consist of architects, designers, developers, engineers, office workers, researchers, students, teachers, and infants' parents. The interviews revealed the advantages of the prototype and possible improvements. To examine what potential users need to know about infants, we conducted a SWOT analysis based on all results of the interviews. Table 3 shows a summary of this analysis.

As a result of the interviews, the application's advantages were categorized into strengths and opportunities. These indicate that visualization of infant behavior using mobile devices using AR technology was able to gain adults' understanding of infants' and the home environment's specific affordances as a whole. Regardless of nationality, occupation, and sex, the interviewees showed interest in the infants' behaviors as "unexpected" interactions with daily used items.

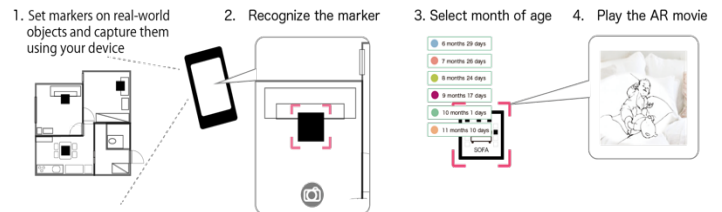


Figure 6. How to use the application.

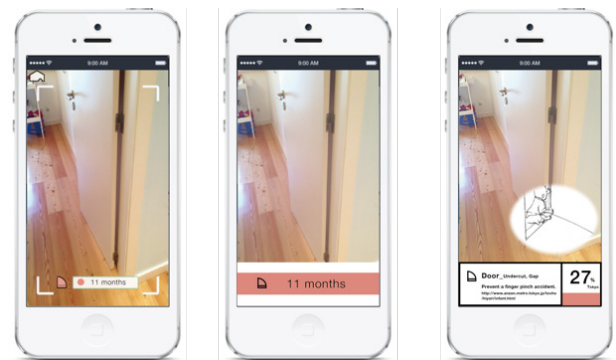


Figure 7. Example of notifications.

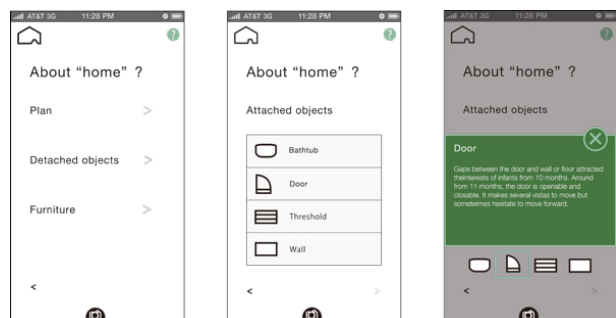


Figure 8. Example of content with pop-up windows of an attached object.

Table 3. Summary of SWOT analysis for the prototype.

CAT	Strengths	Weaknesses	CAT
【Usability】			
R, O, P, S	Instant access to information and camera.	Improve recognition.	R
R, P	Protected privacy with line drawings.		
P	Add more function (e.g., recommendation of products).		
【Method】			
R, P, D	All animations are based on longitudinal data is great.	Hard to collect longitudinal data and archive data.	R
		Rich video data; however, difficult to analyze.	R
		Difficult to draw all movies.	R, D
		Need collaborators from several fields.	R
【Effect】			
R, P, S	Easy to understand the characteristics of infants.		
R, P	Reduced risk of inaccurate perceptions or misperceptions.		
R, E, P	Increase interests in infants' behaviors for adult.		
R, P, D	Adds opportunities to reevaluate development.		
R, D	Interactive applications. Enhance communication.		
CAT	Opportunities	Threats	CAT
【Method】			
R	Aid in elderly people's accident analysis.	Radical change in technology.	R, E
R	Applicable to use for autistic children.	Frequent update of programming languages and libraries.	E
R, E	Increase adaptation to wearable devices.		
【Effect】			
R, P	Increase safety to reduce preventable accidents.		
R, D	Promote making the home childproofing.		
O, P	Good for adults, especially new parents.		
R	Potential in the psychology/healthcare/educational fields.		
【Possibility】			
R, D, E	Can bridge several fields through collaboration.	Big data of hospital or healthcare areas.	R
R, P	Collaborative products with this application.		

Note. Interviewees' categories (CAT) are as follows: D = designers/architects; E = engineers/developers; O = office workers; P = parents; R = researchers; S = students.

In terms of strengths, almost all interviewees supported the method of converting digital recorded videos to line drawings. In particular, line drawings ensured a greater feeling of security for the infants' parents compared to embedded raw videos. In terms of reality and security, this method is assumed to be appropriate at the present time. With respect to usability, we found that all interviewees were familiar with the operation of smartphone devices. This suggests that the development of applications for handheld devices are appropriate for everyday use. In the case of interface, the application's simple graphic and motion design

enabled users to access the desired information and to reach the camera icons easily. However, it was also observed that tracking times were affected by the lighting conditions, especially in the indirect lighting condition, it appeared to unstable at times.

Regarding opportunities, researchers who specialized in robotics, design, architecture, and computer science suggested possibilities for expanding the scope of the application to several other fields, such as preventing accidents among the elderly or children with autism. Furthermore, derivatives of the product possibilities based on the application and the longitudinal data were suggested.

Improvements are suggested in the sections where weakness and threats are discussed. Most of weaknesses in the collection of data and its processing were identified. Researchers concerned that data collecting in natural settings are rich resources for visualization; however, they also referred to the difficulty in analysis compared to the data from laboratory experiments. In contrast, this study is considered to be better at collaborating with other fields because of its interdisciplinary work; hence, engineers suggested technical improvements of the AR technology through continuous updates and researchers pointed out the analysis methods.

4. Discussion

The first year of a child's life is full of indefinite positive and negative opportunities that enable children to perceive, act, and learn. These affordances can be changed based on the relationship between an infant and the surrounding environmental conditions, namely, age, sex, motor skills, body scale, order of birth, object, surface, layout, space, and other people. The affordances of the environment that are available to infants are different from those available to adults and infants' actions are also unpredictable to adults. This may result in carelessness or overprotection. Our AR application attempts to consider both positive and negative affordances. In this respect, our application differs from previous injury prevention systems. This feature of the prototype could make a contribution by changing "preventable" incidents into prevention and even more promotion, whichever is appropriate.

4.1. Everyday Life and Objects from Observational Data

In this study, we examined the data from two countries based on longitudinal observations conducted in Japan and Portugal during the first year and measured infants' activities and objects' types and numbers that they directly encountered. Even though the number of infants were smaller than those in our cross-sectional method, the 175 hours of recorded data captured infants' changing processes in the real world. Therefore, we found the common tendencies in the daily use of products that tend to be associated with the accidents. This means that the infants' interactions with these objects are sometimes unpredictable to adults and vary depending on the conditions of the body-environment relationship, layout of objects and people. Adopting only an adult viewpoint in this aspect leads to carelessness or overprotection.

Concretely, this prototype attempted to visualize how infants cope with ten types of normal objects based on the recorded data as evidence. This visualization was enabled using AR technology and facilitated adults' intuitive understanding by focusing on the decisive moments of the infants' action toward the object.

Essentially, this application provides objects' potential issues and values from the view of infants. This means that the object was re-defined from an infant's perspective because the number of possible accidents related to normal detached objects compared to those of special objects for babies remained the same. Moreover, the number of normal, daily-used objects were observed to be greater than that of baby products. This is necessary to provide different affordances to infants. The application shows concrete and specific examples with AR. Thus, it is possible to help to prevent "preventable" accidents.

4.2. AR Mobile Application

We measured the recognition rate and the accuracy of the vision-based system for AR applications to determine whether the AR systems functioned appropriately. The results showed differences between two AR systems depending on the conditions. We found no differences for distances in the 20–40 cm range; however, the tracking time was affected by the surface orientation. It took slightly longer to recognize the AR marker on the vertical surface than on the horizontal surface. Although this kind of problem is also found in Apple's ARKit, the subsequent update of ARKit will make it possible to recognize not only horizontal surfaces but also vertical surfaces as well. This platform is expected to become available in the summer of 2018 [45]. Therefore, we must re-examine and determine which platform is ideal to develop the application in the long run. Moreover, in real-world scenarios, marker tracking is sometimes unstable under indirect lighting. This suggests that the proposed application needs further improvements for everyday use. In summary, the marker-based AR system would be the most practical and adaptable application for the unique situation of a given user, considering the function and size of application at present. However, the number of objects in our everyday environments is increasing and its arrangements are also changing with time. A marker-less vision-based tracking system will be developed in the future using our ongoing longitudinal recorded video resources.

We referred to Apple's human interface guidelines [46] to develop the application's interface. We focused on helping people understand and interact with the content when designing the interface. Considering that the evaluation of the AR system is important even though its conception is fairly recent, compared to developing AR studies. In 2008, Gabbard and Swan suggested the necessity to learn from user studies for user interface design, usability, and discovery, early in the development of emerging technologies such as AR [47,48].

4.3. Visualizing Home Environment with AR

We examined the validity of the aims and methods of this AR application through informal users interviews and user tests. Since 1995, when the first user-based experiments in AR literature were presented, usability studies have been published to address issues in three related areas, namely, perception, performance, and collaboration [2]. In this study, we focused on perception and performance.

The results revealed how users interact with virtual movies in an AR application in the real-world. Experts from 18 fields of research showed interests in the prototype and suggested improvements. This means that the visualization of affordances

using AR technology is acceptable and the applications abilities are not limited to only parents. Furthermore, it indicated that our approach is actually suitable for interdisciplinary research with aging, architecture, computer science, design, education, healthcare, psychology, rehabilitation, and robotics. This may lead to the development of safety measures from positive and negative aspects in the everyday environment in the form of attractive and safe products, measurements and standard, AI-ready data, and smart homes.

Regarding data, almost all interviewees agreed that collecting longitudinal data in natural settings is worth exploring and considered as a good resource for finding solutions. This is one of our application's features. It is valued by professionals such as researchers, developers, and designers. However, this might also be a weakness because every AR animation requires significant time and effort to draw, because even a single movie requires at least 100 hand-drawings at the present. There is room for improvement and a more efficient method.

We fill our homes with objects; therefore, the application is aimed at enabling a re-examination of the way everyday objects are used, designed, and improved flexibly from the perspective of infants. Visualizing affordances related to an infant's home environment with an AR application would provide another structure to explore the creative processes of individuals who are responsible for coordinating the placement of objects at home, and the thinking behind possible solutions to the challenges we are facing currently to reduce preventable accidents and promote development for infants.

5. Conclusions and Future Directions

The work reported in this study represents an essential first step in adapting the concept of affordance theory to visualizing infants' perceptions and actions within the home environment and an attempt to illustrate the reality in a different perspective to caregivers of infants to prevent the accidents at home. Numerous incidents in the home environment result in injuries and even death of young children. Almost all the ordinary incidents of everyday life can pose danger to infants depending on the conditions. In this study, we demonstrate that AR technology is a promising and useful tool for visualizing invisible affordances of everyday objects. Our AR movies are based on the developmental level, or rather, the incidents anticipated owing to the developmental level of young children. To the scenery, we add the facts about infants and objects and enable caregivers to simulate and visualize other possible incidents. Thus, our app is different from injury reporting or warning apps.

Several limitations exist in this study. First, since we needed to add information on the developmental process of children, we conducted longitudinal observation; we could not determine how one specific fall becomes dangerous, because infants experienced thousands of steps and dozens of falls per day during natural locomotion [40]. It may be difficult to finish collecting data for a whole developmental period as that would delay analysis; however, we continue to collect data and increase the number of participants from year to year. Second, technical problems exist; adding AR markers and creating line drawing illustrations from videos are time-consuming. Therefore, an automatic process for generating line drawings from videos needs to be designed.

More broadly, from the results of interviews, we found that the prototype AR application has been easy and efficient for everyday use because smart devices and AR are common now. Our goal is to build an integrated system that will allow researchers, engineers, designers, and users worldwide to collaborate and contribute to the use of the proposed technology to promote the development of young children and prevent accidents in the home environment. We have started collecting data in one more country, as more evidence must be obtained by processing large amounts of relevant data in future works to generate further meaningful insights.

Conflict of Interest

The authors declare that there are no conflicts of interest associated with this manuscript.

Acknowledgment

We thank all the children and parents who participated in this research. We also thank Sachio Uchida, Tesuaki Baba and Shohan Hasan for their help, and João Barreiros and Rita Cordovil for their support in conducting the observations. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 16K21263.

References

[1] M. Nishizaki, "Visualizing positive and negative affordances in infancy using mobile augmented reality" in *Intelligent Systems Conference (IntelliSys)*, 1136-1140, 2017. DOI: 10.1109/IntelliSys.2017.8324272

[2] I. E. Sutherland, "A head-mounted three-dimensional display," in *Proceedings of FJCC 1968*, Thompson Books, Washington DC, 757-764, 1968.

[3] M. Billinghurst, A. Clark, G. Lee, "A Survey of Augmented Reality," *Foundations and Trends in Human-Computer Interaction*, 8(2-3), 73-272, 2015. <http://dx.doi.org/10.1561/1100000049>

[4] T. P. Caudell, D. W. Mizell, "Augmented reality: An application of heads-up display technology to manual manufacturing processes," in *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference*, 2, 659-669. IEEE, 1992.

[5] S. Aukstakalnis, *Practical augmented reality: a guide to the technologies, applications, and human factors for AR and VR*, Addison-Wesley 2016.

[6] R. Azuma, M. Billinghurst, and G. Klinker, "Editorial: Special section on mobile augmented reality" *Comput., Graph.*, 35(4), 7-8, 2011. <https://doi.org/10.1016/j.cag.2011.04.010>

[7] R. T. Azuma, "A survey of augmented reality" *Presence-Teleop. Virt.*, 6(4), 1892, 355-385, 1997.

[8] M. Mohring, C. Lessig, O. Bimber, "Video see-through AR on consumer cell-phones" in *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, 252-253. IEEE Computer Society, 2004. DOI: 10.1109/ISMAR.2004.63

[9] O. Smordal, G. Liestøl, and O. Erstad, "Exploring situated knowledge building using mobile augmented reality" in *QWERTY 11*, 1, 26-43, 2016.

[10] S. Rattanarungrot, M. White and B. Jackson, "The application of service orientation on a mobile AR platform — a museum scenario" *2015 Digital Heritage, Granada*, 329-332, 2015. DOI: 10.1109/DigitalHeritage.2015.7413894

[11] D. Wagner, and D. Schmalstieg, "Making augmented reality practical on mobile phones, part 1" *IEEE. Comput. Graph.*, 29(3), 12-15, 2009. <http://doi.ieeecomputersociety.org/10.1109/MCG.2009.46>

[12] D. Beier, R. Billert, B. Bruderlin, D. Stichling and B. Kleinjohann, "Marker-less vision based tracking for mobile augmented reality" *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2003. *Proceedings*, 258-259, 2003. DOI: 10.1109/ISMAR.2003.1240709

[13] SUPERDATA Games & Interactive Media Intelligence, <https://www.superdataresearch.com/us-digital-games-market/>, last accessed 2018/08/13.

[14] Digi-Capital Homepage, <https://www.digi-capital.com/news/20170/01/after-mixed-year-mobile-ar-to-drive-108-billion-vr-ar-market-by-2021/>, last accessed 2018/05/03.

[15] The Ministry of Health, Labour and Welfare of Japan. List of Statistical Surveys. Vital Statistics 2016, https://www.e-stat.go.jp/en/stat-search/files?page=1&layout=datalist&tstat=000001028897&year=20160&month=0&tclass1=000001053058&tclass2=000001053061&tclass3=000001053066&result_back=1, last accessed 2018/05/03.

[16] The National Safety Council of US Homepage, <http://www.rospa.com/homesafety/adviceandinformation/childsafety/accidents-to-children.aspx>, last accessed 2018/05/03.

[17] Consumer Safety Unit. 24th Annual Report, Home Accident Surveillance System. London: Department of Trade and Industry, 2002.

[18] The U.S. Consumer Product Safety Commission, <https://www.cpsc.gov/s3fs-public/5013.pdf>, last accessed 2018/05/03.

[19] A. Carlsson, A. K. Dykes, A. Jansson, A. C. Bramhagen, "Mothers' awareness towards child injuries and injury prevention at home: an intervention study" *BMC Research Notes*, 9, 223, 2016. <https://doi.org/10.1186/s13104-016-2031-5>

[20] D. Kendrick, C. A. Mulvaney, L. Ye, T. Stevens, J. A. Mytton, S. Stewart-Brown, "Parenting interventions and the prevention of unintentional injuries in childhood: systematic review and meta-analysis" *Cochrane Database System Review*. 2013 Mar 28; (3):CD006020. Epub. <https://doi.org/10.1111/j.1365-2214.2008.00849.x>

[21] B. A. Morrongiello, S. Kiriakou, "Mothers' home-safety practices for preventing six types of childhood injuries: what do they do, and why?" *J. Pediatr. Psychol.*, 29(4), 285-97, 2004.

[22] A. Carlsson, A. K. Dykes, "Precautions taken by mothers to prevent burn and scald injuries to young children at home: an intervention study" *Scand. J. Public Health*, 39(5), 471-478, 2011. <https://doi.org/10.1177/1403494811405094>

[23] SAFE KIDS WORLDWIDE, https://www.safekids.org/safetytips/field_venues/home, last accessed 2018/08/13.

[24] Japan Pediatric Society, Injury Alert and Follow-up report <http://www.jpeds.or.jp/modules/injuryalert/>, last accessed 2018/08/13.

[25] K. Kitamura, Y. Nishida, Y. Motomura, H. Mizoguchi, "Children Unintentional Injury Visualization System Based on Behavior Model and Injury Data" *The 2008 International Conference on Modeling, Simulation and Visualization Methods (MSV'08)*, 2008.

[26] K. Kitamura, Y. Nishida, Y. Motomura, T. Yamanaka, H. Mizoguchi, "Web Content Service for Childhood Injury Prevention and Safety Promotion," in *Proceedings of the 9th World Conference on Injury prevention and Safety Promotion*, 270, 2008.

[27] Y. Nishida, Y. Motomura, K. Kitamura, T. Yamanaka, "Representation and Statistical Analysis of Childhood Injury by Bodygraphic Information System" *Proc. of the 10th International Conference on GeoComputation*. 2009.

[28] J. J. Gibson, "The Senses Considered as Perceptual Systems, Houghton. Mifflin Company, Boston, 1966.

[29] J. J. Gibson, "The ecological approach to visual perception," Boston: Houghton Mifflin, 1979.

[30] K. E. Adolph, "A psychophysical assessment of toddlers' ability to cope with slopes" *J. Exp. Psychol. Human*, 21, 734-750, 1995.

[31] E. J. Gibson and R. D. Walk, "The 'visual cliff'" *Scientific American*, 202, 67-71, 1960.

[32] T. Stoffregen, "Affordances and events" *Ecol. Psychol.*, 12, 1-28, 2010. https://doi.org/10.1207/S15326969ECO1201_1

[33] M. Turvey, "Affordances and prospective control: An outline of the ontology" *Ecol. Psychol.*, 4, 173-187, 1992. https://doi.org/10.1207/s15326969eco0403_3

[34] W. H. Warren, "Perceiving affordances: Visual guidance of stair climbing" *J. Exp. Psychol. Human*, 10, 683-703, 1984.

[35] W. Gaver, "What in the world do we hear? An ecological approach to auditory source perception" *Ecol. Psychol.*, 5, 1-31, 1993a. https://doi.org/10.1207/s15326969eco0501_1

[36] W. Gaver, "How do we hear in the world? Explorations in ecological acoustics" *Ecol. Psychol.*, 5, 285-313, 1993b. https://doi.org/10.1207/s15326969eco0504_2

[37] L. S. Mark, "Perceiving the preferred critical boundary for an affordance," in *Studies in perception and action III*, B. G. Bardy, R. J. Bootsma, and Y. Guiard (Eds.), Mahwah, NJ: Lawrence Erlbaum Associates, 1995.

[38] L. S. Mark, K. Nemeth, D. Gardner, M. J. Dainoff, J. Paasche, M. Duffy, and K. Grandt, "Postural dynamics and the preferred critical boundary for visually guided reaching" *J. Exp. Psychol. Human*, 23(5), 1365-1379, 1997.

[39] W. H. Warren, and S. Whang, "Visual guidance of walking through apertures: Body-scaled information for affordances" *J. Exp. Psychol. Human*, 13(3), 371-384, 1987.

- [40] K. E. Adolph, W. G. Cole, M. Komati, J. S. Garciaguirre, D. Badaly, J. M. Lingeman, G. L. Y. Chan, R. B. Sotsky, "How do you learn to walk? Thousands of steps and dozens of falls per day" *Psychol. Sci.*, 23, 1387-1394, 2012. <https://doi.org/10.1177/0956797612446346>
- [41] W. G. Cole, S. R. Robinson, & K. E. Adolph, "Bouts of steps: The organization of infant exploration" *Dev. Psychobiol.*, 58, 341-354, 2016. <https://doi.org/10.1002/dev.21374>
- [42] J. M. Franchak, and K. E. Adolph, "Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults" *Vision Res.*, 50, 2766-2774, 2010. <https://doi.org/10.1016/j.visres.2010.09.024>
- [43] Datavyu, <http://www.datavyu.org/> last accessed 2018/10/10.
- [44] Consumer Affairs Agency, Government of Japan, "Policy of children's accidents prevention in 2017" http://www.caa.go.jp/policies/policy/consumer_safety/child/children_accident_prevention/pdf/children_accident_prevention_180328_0004.pdf, last accessed 2018/08/13.
- [45] Apple, Newsroom <https://www.apple.com/jp/newsroom/2018/06/apple-unveils-arkit-2/> last accessed 2018/08/13.
- [46] Apple, human-interface-guidelines, <https://developer.apple.com/design/human-interface-guidelines/ios/overview/themes/>, last accessed 2018/08/05.
- [47] J. E. Swan J. L. Gabbard. "Survey of user-based experimentation in augmented reality" in *Proceedings of 1st International Conference on Virtual Reality*, 1-9, 2005.
- [48] J. L. Gabbard and J. E. Swan, "Usability engineering for augmented reality: Employing user-based studies to inform design" *Vis. Comput. Graph., IEEE*, 14(3), 513-525, 2008. <http://doi.ieeecomputersociety.org/10.1109/TVCG.2008.24>

Enhanced Ship Energy Efficiency by Using Marine Box Coolers

Abdallah Aijjou*, Lhoussain Bahatti, Abdelhadi Raihani

Laboratory: signals, distributed systems and Artificial Intelligence, ENSET Mohammedia, University Hassan II Morocco.

ARTICLE INFO

Article history:

Received: 13 August, 2018

Accepted: 24 October, 2018

Online: 01 November, 2018

Keywords:

Ship energy efficiency

Box / keel coolers

Central cooling

Sea water pumping

ABSTRACT

Climate change, increasing fuel oil prices and new international regulation on ship emissions lead to more focus on shipping fuel consumption and energy efficiency. There are various solutions for improving the ship energy efficiency. In this manuscript, we aim to present a real case of energy saving by adopting the central cooling system with box cooler on the ship instead of conventional system. The electric energy power necessary for operating the machinery cooling system of the ship is calculated for conventional cooling system and compared to the cooling system using box coolers in term of fuel oil consumption and CO₂ emissions. This study quantified the fuel saving potential that could be achieved with use of keel coolers. Adopting central cooling with box coolers may contribute in reduction of fuel oil consumption and improving the ship energy efficiency. systems. should not contain citations.

1. Introduction

The ship machinery is essentially composed by propulsion diesel engine, auxiliary engines for electric power production and boiler for heating purposes. Other auxiliary equipment necessary for the operation are also fitted such air compressors, air conditioning plant, steam condensers, hydraulic power pack for ship mooring and cargo operation etc.

All this equipment produces undesirable heat and need cooling which remove excessive heat out of surfaces of material, safeguard the metal mechanical properties and keep the temperatures within the limits specified by makers for maximum performances during operations.

The propulsion plant (mostly diesel engines) is the largest source of energy to dissipates by cooling system [1]. The cooling system is designed to covers machinery cooling needs when most of the equipment are in operation at its maximum powers. Safe margin is also added.

When we look at the electric energy balance of various types of ships we notice that cooling water pumps are among the largest and the most equipment running over the time, since they are solicited either when the vessel is at sea or alongside. Therefore, cooling pumps are the largest electric power consumers on board and are accounted for 10% of the kilowatt- hour consumed.

This fact explains why several researches dedicated to ship energy efficiency focus on cooling system components. These

studies have focused mainly on improving the performance and efficiency of the conventional cooling system (piping and sea water pumps) i.e. [2,3,4,5] studies on energy saving by improving the pumping system, [6] diagnosis and corrosion protection, [7] cooling system reliability.

Very few studies discussed the possibilities of elimination of sea water pipes and pumps from cooling system [8]. The use of box or keel coolers is one of the solutions available for the ship building industry.

Adopting box coolers for cooling systems allows to eliminate the pumps, filters, valves and reduces the piping length therefore the installation is more simple and cheaper. In addition, elimination of these components will reduce also the cost of the operational maintenance.

This paper aims to demonstrate the effect of box cooler use on ship energy efficiency hence after this introduction we give an overview of different ship cooling systems in section II, in section III box and keel coolers principles are described, in section IV we study an application of box cooler for real ship in operation. Conclusion is given in section V.

2. Ship machinery cooling system:

There are three basic types of cooling system commonly used in the marine machinery on board the vessel.

1-Direct cooling system with sea water (Fig.1): sea water is used as cooling media for heat exchangers in open circuit. Sea

*Abdallah Aijjou, Email: thalassa1310@yahoo.fr, Tel: +212 661 42 34 17,

water is drawn from sea and pumped directly through the machinery system before being discharged overboard.

The inconveniences of this system are:

- The sea water temperature must be below 50°C to avoid scale formation.
- contamination of the water supply with consequent deposition inside the piping and cooled equipment.
- the system is subject to high rate of corrosion and erosion due to the nature of sea water.

2-Indirect cooling system with freshwater (Fig.2): machinery components are cooled by treated fresh water in a closed circuit. This fresh water is pumped through sea water coolers where it is cooled by sea water.

This system reduces the length of sea water piping inside engine room and thus eliminates the corrosion problems linked to sea water. Furthermore, sea water pipes may be made from brass with limited costs.

Also, fresh water as cooling media permits the use of plate cooler instead of tubular cooler for their efficiency and easy maintenance. Figures 1&2 illustrate the simplified drawings of the two conventional cooling systems respectively direct and indirect. Only main components are shown.

3-Keel / box cooler based cooling system: In this system, the cooling fresh water is cooled in coolers fitted outside the engine room. Sea water cooling pump and pipes are eliminated. This system is further explained in next chapters.

Box cooler is a vessel cooling system, in the form of U-tube-bundle (Fig.3 and Fig.4) that is fitted in a sea-chest (Fig.5) on the side of the vessel. The sea chest is equipped with inlet and outlet-grids for circulating cooling sea water. The heat exchange takes place in the sea chest by natural convection of the water when the vessel is stationary or by a circulation due to the speed of the vessel.

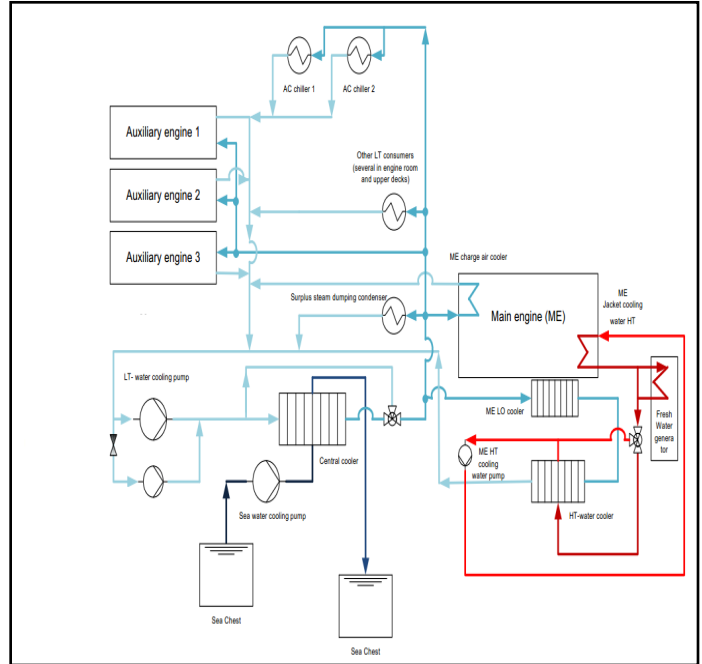
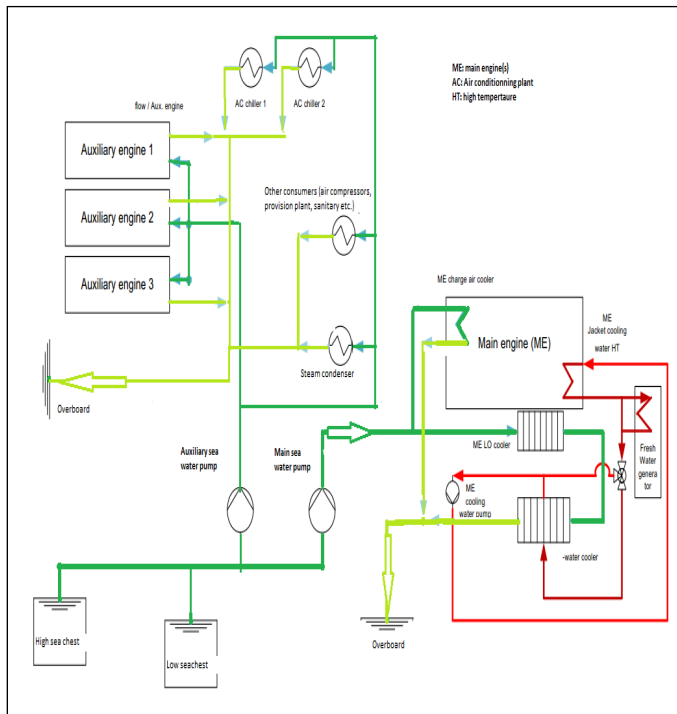


Figure 2. Indirect cooling system with fresh water.



3. Description of keel cooler and box cooler.

The terms box cooler and keel cooler refer to the heat exchangers mounted on the underside of ship outer shell or in sea chests under the water line, saving space in the engine room.



Figure 3. Box cooler: Main engine (left), Auxiliary engine (right). Photos taken during the vessel “ASD” stay in dry dock (2017).

Keel cooler (Fig.6): The closed circuit of fresh water is pumped in the spiral tubes installed outside the ship’s hull below the water line, its contact with the sea water ensures the heat transfer and achieving the cooling effect. Compared to

conventional cooling systems, a keel / box cooling system provides several advantages.

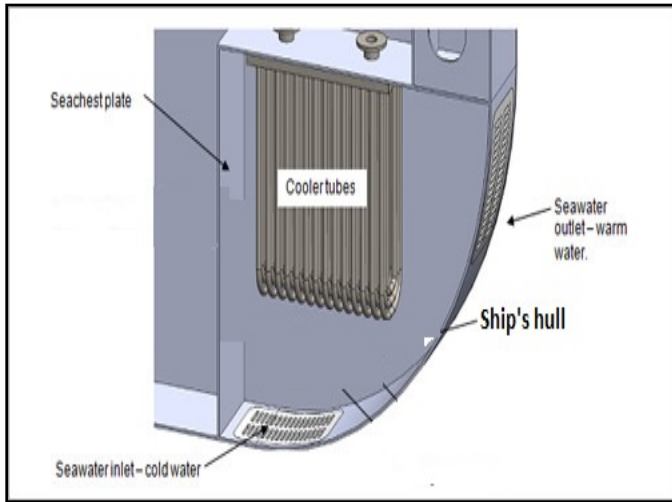


Figure 4. Box cooler and sea chest. (Google image).

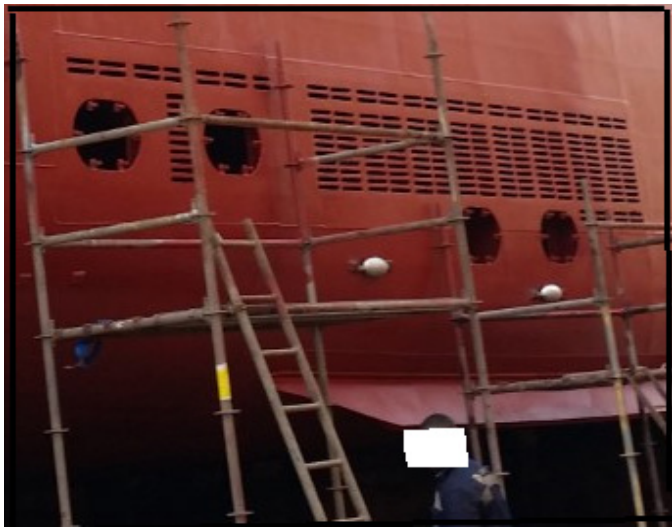


Figure 5. Sea chest where box coolers are fitted ("ASD" at dry dock 2017).

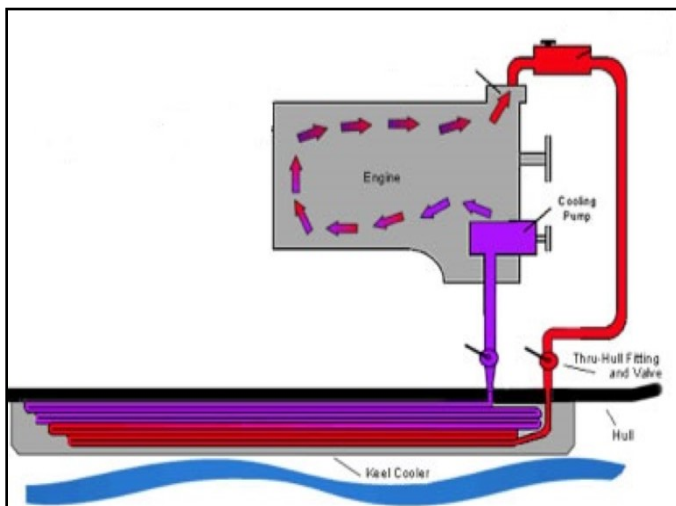


Figure 6. Basic keel cooling system. (Google image).

The sea water pumps, filters, sea water piping are eliminated, operational cost for cleaning, repairing and renewing the system components are reduced. It can also operate in all sea water conditions: icy, silt or polluted. However, the use of box and keel coolers raises some challenges, mainly the material corrosion and the biological fouling.

Nowadays the corrosion caused by box coolers may not a big problem due to technological advances in manufacturing resistant materials. The coolers are fabricated from noble materials such as copper alloys deemed corrosion resistant. The galvanic effect caused by material difference between the ship's hull made from steel and coolers is overcome by coating the sea chest surface and in some cases by ICCP.

The biological fouling by barnacles, algae and other shellfish depends on the area where the ship is trading and operational conditions. Nowadays several solutions exist to prevent this problem: antifouling coating, electrolytic, chemical injections etc.

The use of box coolers is also limited by its capacity, according to makers, existing box coolers in the market may be used for cooling engines with output up to 30,000KW. Most of the commercial ships are fitted with engines of less power than this limit.

4. Case study

For this paper, we take for study the machinery arrangements of the vessel "Aline Sitote Diatta (ASD)", this vessel is a "Car Ferry" type used for the transport of passengers and vehicles between ports in Senegal. The particulars of this vessel are summarized in table I. The vessel is a car ferry type fitted with two four strokes main engines (ME) with a rated power of 1800 Kw each.

The electric power is produced by three sets of diesel generators 433 KW each. One generator can supply all necessary power during normal seagoing. The ship is equipped with two air conditioning units for all accommodations: passengers and crew cabin etc. As fuel, the ship's main engines and auxiliary engines are burning marine diesel oil.

Table 1: The ship's particulars.

Length	76 meters
Speed	14.5 Knots
Gross tonnage	3481
Main engine output	Two diesel engine 1800 KW each
Propeller type	Controllable pitch propeller
Capacity	504 passengers, 28 vehicles

Figure 7 is a simplified representation of the machinery cooling installation on board the vessel "ASD". Sets of box coolers are fitted in sea chests on both sides of the ship. Each engine's water coolant (fresh water) is circulated by the engine cooling pump through one box cooler, where the coolant is further cooled by sea water before returning to the engine. Heat transfer is achieved by natural convection and circulation due to ship movement.

Compared to figures 1 and 2, we can notice that main engine sea water pump and auxiliary engine pumps are eliminated, including associated piping, valves, and strainers.

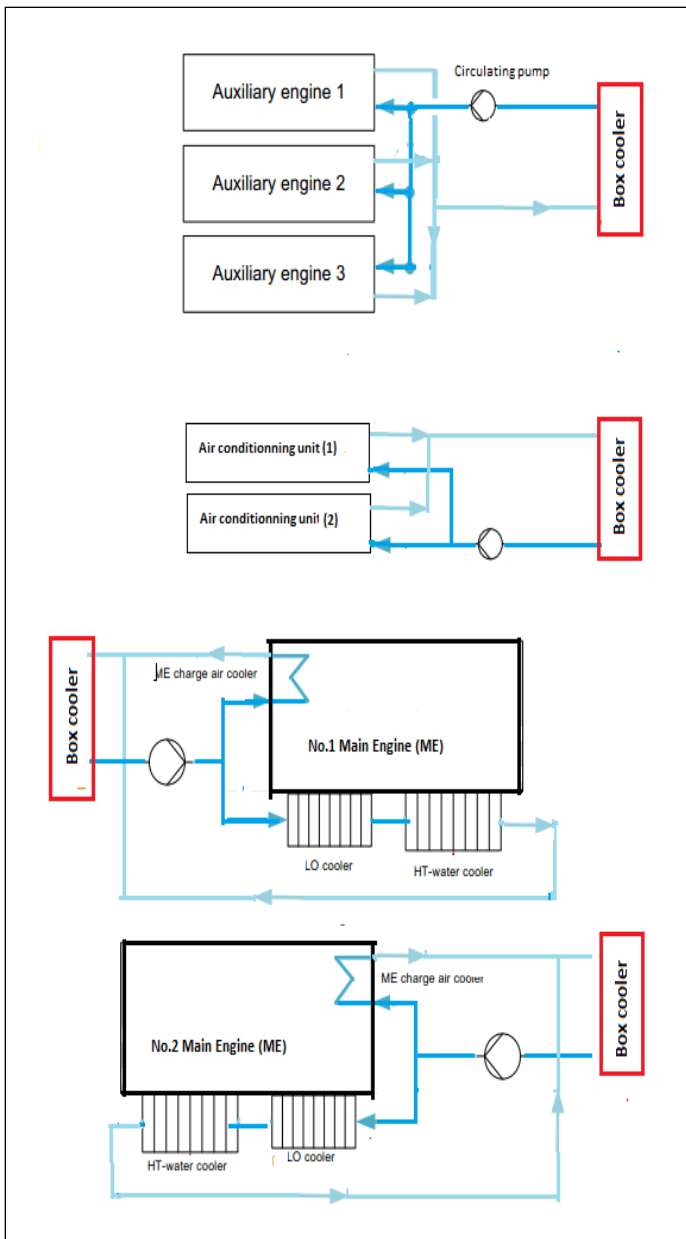


Figure. 7. Simplified representation of the cooling diagram of the vessel "ASD".

4.1. Cooling system parameters calculation:

Determination of the sea water flow:

The quantity of sea water (V_0) expressed in *cubic meter per hour* (m^3/h) necessary for cooling a single diesel engine is defined by formula (1).

$$V_0 = \frac{Q}{C_{sw} \cdot \rho \cdot \Delta T} \quad (1)$$

Where

Q : Heat rejection of Diesel Engine in KJ/h

C_{sw} : Specific heat of sea water = 3.925 KJ/Kg°C.

ρ : Sea water density = 1027 Kg/ m^3

ΔT : Difference between the inlet and the outlet temperature of engine sea water cooling.

As per the engine specification published by the Maker, the total heat rejection (Q) the case study engine is [9]: Jacket cooling water, High Temperature (HT) circuit (375 KW); Charge air cooler, Low temperature (LT)-circuit (607 KW); Lubricating oil, LT-circuit (270 KW).

Note: Due to soft scale deposit inside the sea water piping and coolers which affect the water flow and cooler efficiency, the sea water cooler outlet temperature is generally maintained below to 50°C. Taking in consideration that sea water temperature in certain navigation zones is up to 36°C, we take $\Delta T \approx 14^\circ\text{C}$.

By substitution into formula (1) we can calculate the volumetric rate of sea water flow necessary for one engine cooling needs which is approximately $\approx 80 \text{ m}^3/\text{h}$. The ship is fitted with two main engines. The ship is also equipped with three auxiliary engines for electric power generation but in worst case only two engines are sufficient to cover the ship power needs. Based on the auxiliary engine technical data, the sea water cooling flow necessary for cooling two auxiliary engines running in parallels is $\approx 30 \text{ m}^3/\text{h}$.

As the ship is intended for passenger transport the air conditioning plant is important and its capacity is considerable hence the sea water flow required for the AC condenser as per its specification is $20 \text{ m}^3/\text{h}$.

The total cooling sea water flow is the sum of:

$$(2 \times 80 + 30 + 20) \text{ m}^3/\text{h}$$

Taking safety margin as 10%, The total sea water rate necessary for cooling the machinery at full load is approximately $230 \text{ m}^3/\text{h}$.

Sea water pump power:

As per Bernoulli equation, the hydraulic power (p) (expressed in watt) transmitted to the sea water as it passes through the pump is calculated by formula (2)

$$p = V_s \cdot H \cdot \rho \cdot g \quad (2)$$

Where

V_s is the sea water rate in m^3/s , ρ is the density of sea water ($\rho = 1027 \text{ kg}/\text{m}^3$ at the sea surface), H is the total manometric height of the pump and g is acceleration due to gravity, average $g = 9.81 \text{ m}\cdot\text{s}^{-2}$.

The total manometric height H is defined by formula (3):

$$H = \left(\frac{p_2 - p_1}{\rho \cdot g} \right) + J_r + J_n \quad (3)$$

p_1 : pump suction pressure, in Pascal (Pa)

p_2 : pump discharge pressure, in Pascal (Pa)

J_r : geometrical height of the pumped medium lift, in meter (m)

J_n : head overall loss, in meter (m)

The difference in dynamic heights is negligible.

J_n includes all losses due to friction inside piping, flanges and valves. Its value depends on the fluid velocity, piping material, piping geometry etc....

Sea water pump discharge pressure is globally around 2.5 bars (total $H \approx 25m$), based on data published by marine diesel engine makers. By solving the equation (2) we define the hydraulic power absorbed by the pumped sea water flow of 230 m^3/h .

$$p \approx 16 \text{ Kw.}$$

The sea water pump is driven by an electric motor, the electric motor power (P) is obtained from formula (4).

$$P = \frac{p}{\eta_e \cdot \eta_{pm}} \quad (4)$$

- η_{pm} : pump efficiency, sea water pump is a centrifugal type. The average efficiency for such pump is 60%. [10]. η_e : is the efficiency of the pump electric motor approximatively 90%.

From formula (4) we defined the electric power necessary for driving the cooling sea water pump $P = 16/06 \times 0.9 \approx 30Kw$, equivalent to 1.7% of the main engine output. The specific fuel oil consumption for the modern four stroke engine at optimal service conditions is [7,11]:

$$SFOC = 200 - 210 \text{ g/Kwh.}$$

Therefore, eliminating sea water pump means saving 150 kg of fuel oil. This is equivalent to 450kg CO_2 emission prevention per day. Considering that main engines are used for an average of 260 days per year, the total fuel oil save per year is 39 tons and 117 tons of CO_2 emission prevention. The average price of marine fuel oil is 500 \$/ton [12]. The application of box cooler for the cooling system of the vessel under study permits to:

- enhance the ship energy efficiency, by reducing the fuel oil consumption and CO_2 emissions.
- reduce the ship operational cost.

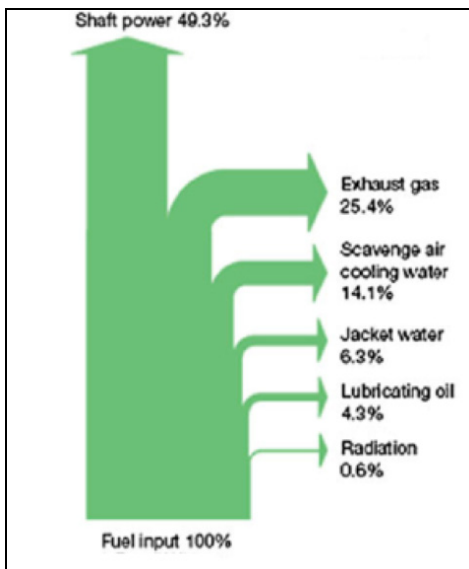


Figure. 8. Energy balance for typical marine diesel engine

4.2. Typical diesel engine energy flux and cooling system energy

The transformation of chemical energy to mechanical work by burning fuel oil is subject to various heat losses. The diagram in Figure 8 also called Sankey diagram, illustrates the basic energy

balance of a modern Diesel. The heat losses by cooling engine lubricating oil, engine cooling water and the exhaust gases are mainly produced by the irreversibility of the engine cycle and only part of the energy is transformed into useful mechanical work.

The diagram in Figure 6, also called Sankey diagram, illustrates the overall energy balance of a modern Diesel engine, the values shown on the diagram are an average, slight deviations are possible depending on engine category, type, and power. Globally, the heat losses in marine diesel engines are estimated to 26-30%, this amount needs to be absorbed by the cooling system mainly with sea water.

Taking in consideration that the most used type of fuel for ship bunkers is the heavy fuel oil having a lower calorific value (LCV) of 42,700 KJ/Kg at ISO conditions. The specific fuel oil consumption for the modern two stroke engine is better than for the four-stroke engine.

SFOC = 160 g/Kwh. The actual measured values for various engines is within range of 185- 205.

The input quantity of chemical energy necessary per KWh output is given by formula (3).

$$Q_i = SFOC \cdot LCV, \text{ (KJ/KWh).} \quad (3)$$

The total fuel energy consumed by the engine per hour (input) is given by:

$$Q_t = Q_i \cdot P, \text{ (KJ/h).} \quad (4)$$

(P) in KW: the engine power transmitted to propulsion shaft (output). The heat losses Q_l to be evacuated by the cooling system is within 26% - 30% of Q_t (KJ/h). Figure 8 as example.

$$Q_l \approx 26-30\% Q_t$$

$$Q_l \text{ (KJ/h)} \approx (26 - 30) \% (0.160 \times 42700 \times P)$$

The heat to be dissipated in Kw is equal to $Q_l/3600$

From formulas (3) and (4) we can assume that the quantity of heat losses (in KW) to be dissipated by cooling system may reach up to 60% of diesel engine output. The cooling sea water rate and the electric power necessary for its pumping are defined by equations (1) and (2).

Table II is extracted from engine specification [11] it shows the cooling water capacity and power for modern diesel engines with different outputs. The column 5 is calculated based on SFOC of 200 g/Kwh, the optimal value for diesel generators. The table II covers only the main engine cooling system. The ship cooling system must be designed for also the auxiliary machinery components: diesel generators, steam condensers, cargo plant, air conditioning etc. The sea water flow and the pump capacity are higher than the values indicated, the fuel oil save and emissions preventions as well. TABLE II shows that the cooling sea water pump for single engine may absorb up to 0.4% of the engine output.

Depending on the size of this machineries, cooling system power needs may reach up to 1.7% as per the case studied.

Table 2: Propulsion power X Cooling power

Engine output (KW)	Heat dissipation (KW) (Maker data)	Cooling pump capacity (m3/h)	Cooling pump power (KW)	*Equivalent Fuel consumption per day for cooling (kg)
29000	17230	850	110	528
40670	24080	1177	153	735
58100	34480	1680	217	1042
81000	48000	2400	310	1488

World fleet outlook; Fuel oil consumption:

Table 3 is an extract from “DNVGL” report on 2030 world fleet outlook [14]. The total world marine fuel oil consumption is estimated to 325 Million tons per year equivalent to 1.028 Million tons of CO2 emissions.

Table 3: World fleet X Fuel oil consumption

Ship's type	Number of ship	Max. engine power	Average engine power	Average consumption per year (in tons)
Crude	2037	36,941	15000	10760199
Products	5272	20,080	2390	7242117
Chemical	3895	14,758	4185	14469811
LPG LNG	1725	39,902	8252	11293652
Bulk	7392	30,099	7830	49570494
General cargo	18473	16,550	2270	31986087
Container	4138	80,911	21880	63722049
Ro-Ro - Ferry	8859	52,799	4660	34889212
Cruise - Yacht	1550	75,627	6710	5945856
Offshore	5086	45,199	3821	4564645
service	17303	82,510	2162	14041725
Miscellaneous	24516	49,910	1060	15647195
	100246			280965660.

Total fuel oil consumed by shipping sector is evaluated to 325 million tons, 281 for propulsion and 44 consumed by auxiliary engines.

Reduction of 0.5% of fuel consumption by adopting box cooler means saving 1.6 million tons of fuel oil yearly. From the same table, we notice that less than 3% of ships have propulsion power of more than 30,000KW therefore 97% of world fleet may be fitted with central cooling based on box coolers.

5. Conclusion

The machinery cooling system of the vessel “Aline Siteo Diatta” was studied. Two cases were considered, use of conventional cooling system and application of box cooler. Both cases are compared for energy efficiency and fuel oil consumption. The main conclusions are:

The ship cooling sea water pumps are the largest auxiliary pumps on board and are running most of the time therefore there are the largest electric power consumers, up to 10% of ship’s needs in kilowatt – hour. For single engine, cooling water pumps absorbs as a minimum 0.4% of the engine output.

In the case studied cooling system of all main and auxiliary machineries requires up to 1.7% of the propulsion engine power. In the opinion of authors, the use of the keel or box coolers for ship machinery cooling system may contribute to reduce fuel oil consumption of the ships, to improve the energy efficiency accordingly and meet the legal requirements [13].

The application of box cooler is limited to small size ships but further to recent technology developments it’s possible to adopt on ships up to 30,000 KW (only 3% of the world cargo vessel is equipped with engine of more than 30,000 KW, mainly container vessels).

With large power, the box coolers may be part of the solution which will contribute to reduce the size and the cost of the cooling system and energy consumption.

References

- [1] C. T. Wilbur “Pounder’s Marine Diesel Engines” Sixth 6th edition.
- [2] Chun-Lien Su, Wei -Lin Chung, Kuen-Tyng Yu. « An enegy saving evaluation method for adjustable frequency drives on sea water cooling pumps on ships. Industrial and commercial Power system techniques. 2013/ IEEC/IAS 49th
- [3] Gazi Cocak, Yalcin Durmusoglu “Energy efficiency analysis of a ship’s central cooling system using variable speed pump”. Journal of Marine Engineering and Technology: Published on 2017-01-31.
- [4] Mia Elg, Maunu Kuosa, Markku Lampinen, Risto Lahdelma, Panu Mäkipeska, Juuso Raita, Guangrong Zou, Kari Tammi “Advanced auxiliary cooling system for energy efficient ships”. Technical Research Centre of Finland Ltd. -C. T. Wilbur
- [5] Gerasimos Theotokatos, Konstantinos Sfakianakis and Dracos Vassalos. “Investigation of ship cooling system operation for improving energy efficiency” Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, G4 0LZ, Glasgow, UK.
- [6] Application of titanium in shipboard sea water cooling systems. Wayne L. Adamson Vol.99, Issue 3.
- [7] Ait Lallal Abdelmoula, Khalifa Mansouri “Toward a reliable sea water central cooling system for a safe operation of autonomous ship”/ ENSET Mohammedia, University Hassan II Casablanca.
- [8] Andrzej Młynarczyk. Box coolers as an alternative to existing cooling systems Scientific Journal. Maritime University og Szczecin . 2013, 36(108) z. 2 pp. 131–136.
- [9] Wartzilla 20 product guide. Issue 1/2017
- [10] Sulzer_Centrifugal_Pump_Handbook_3rd_Ed.M.
- [11] MAN B&W S90ME project guide ed. 05/2014
- [12] <https://bunkerindex.com/>
- [13] IMO. (2011). Resolution MEPC.203(62), Amendments to the annex of the protocol of 1997 to amend the international convention. [http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Marine-Environment-Protection-Committee-\(MEPC\)/Documents/MEPC.203\(62\).pdf](http://www.imo.org/en/KnowledgeCentre/IndexofIMOResolutions/Marine-Environment-Protection-Committee-(MEPC)/Documents/MEPC.203(62).pdf)
- [14] World fleet MACC 2030 Ver.24 DNV. 2012-02-12 GL Library..
- [15] DNV GL SE, (2015). Rules for Classification and Construction, I-1.2, July 2015, Hamburg, Germany DNV.GL website publication 2017.

A Resolution-Reconfigurable and Power Scalable SAR ADC with Partially Thermometer Coded DAC

Hao-Min Lin*, Chih-Hsuan Lin, Kuei-Ann Wen

Department of Electronic Engineering, University of National Chiao Tung (NCTU), Hsinchu 300, Taiwan

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 19 October, 2018

Online: 01 November, 2018

Keywords:

Reconfigurable

SAR ADC

Thermometer Coded

ABSTRACT

Power consumption is becoming more and more important in the Internet of Things (IOT). The ADC is the main power hungry in multi-sensor electronic systems and effectively reducing ADC power consumption without affecting ADC characteristics is an important. This paper is extended from the conference paper. The segmented SAR ADC presents reconfigurable 9 to 12-bit DACs with rail-to rail input range, and 3 MSB segmented capacitor arrays are used to improve linearity and lower switching energy than conventional architectures. The dual supply voltage skill separating digital and analog voltage is implemented for achieving low power consumption. In the provided 9 to 12 bits mode, this structure consumes 2.5, 2.8, 3.9 and 9.7 μ W and SNDR achieve 52.3, 57.7, 63.2 and 68.6 db respectively, resulting in figure of merit (FoM) 148, 88.8, 66.3 and 88.4 fJ/conversion-step

1. Introduction

“This paper is an extension of work originally presented in A Low Power Reconfigurable SAR ADC for CMOS MEMS Sensor” [1]. Modern consumer electronics use multi-sensor more and more frequently for CMOS MEM process, the main reason is that has low cost and high integration characteristics of electronic systems. The reported sensors for different capacitive sensitivity have different ranges, such as CMOS MEMS accelerometers are often less than 1 fF/g. To provide greater sensitivity, the readout circuit needs to provide a large capacitance-to-voltage conversion gain. However, large conversion gain amplifies noise and reduces signal-to-noise ratio (SNR). The Correlated double sampling (CDS) and chopper techniques are often used to reduce flicker noise. Moreover, in order to provide a large dynamic range for low to high sensitivity sensors, the programmable gain amplifier (PGA) is implemented. Followed by segmented successive approximation register (SAR) ADC, we choose low power, high resolution, using thermometer code to improve linearity for 3 MSB, and easy-controlling logic circuit to design a reconfigurable ADC that can adjust resolution and power consumption for different requirement. In addition, the power consumption reduction when scaling down the resolution can still maintain the FoM.

On the other hand, the conventional capacitor array architecture performs approximation action and the energy

consumption is not efficient. The monotonic capacitor array architecture consumes less than the energy of a conventional capacitor array architecture [2]. The MCS consumes less than monotonic capacitor array architecture [3]. The segment capacity array architecture consumes the same energy as MCS.

In this paper, the proposed SAR ADC can be fabricated under UMC 0.18 mm standard CMOS-MEMS process, which is highly area efficient with MEMS sensor being integrated in single chip. This paper is organized as follows: Section 1 describes ADC ARCHITECTURE including Analysis average energy of Capacitance DAC array structure, analysis of linearity. Section 2 presents ADC ARCHITECTURE DESIGN including system architecture, bootstrap switch and sample-hold, scalable resolution design, control logic and multiplexer Scalable voltage design, comparator, Level shifter. Section 4 describes RESULT and CONCLUSION.

2. ADC Architecture

2.1. Analysis average energy of Capacitance DAC array

In order to achieve greater than 10 bits accuracy, using differential architecture to suppress substrate noise and power noise and have good common mode noise suppression. The conventional SAR ADC architecture is shown in Figure. 1. and often uses of binary weighted capacitor arrays for better linearity. The function block has sample-and-hold, comparator, capacitance DAC array, successive approximation registers. The conventional

*Hao-Min Lin, Email: ken970054@gmail.com

SAR ADCs are complementary in terms of the fully differential architecture and the following describes the operating procedures on the positive side. In the sampling phase, the bottom plate capacitor on the positive side is charged to the V_{ip} and the top plate capacitor on the positive side is connected to the common mode voltage V_{cm} . Next phase, the maximum capacitance bottom plate on the positive side is switched to V_{ref} and the other capacitors on the positive side are switched to GND. At this time, the comparator compares the node voltage V_{xp} and V_{xn} . When the node voltage V_{xp} is greater than V_{xn} , the most significant bit (MSB) “S11p” is high. Otherwise, “S11p” is low. Then the second maximum capacitor “C2” is switched to V_{REF} and the comparator compares the nodes voltage V_{xp} and V_{xn} . The SAR ADC will continue to repeat this process until the least significant (LSB) is determined. Although this trial-and error action is simple, it is not a save power switch procedure.

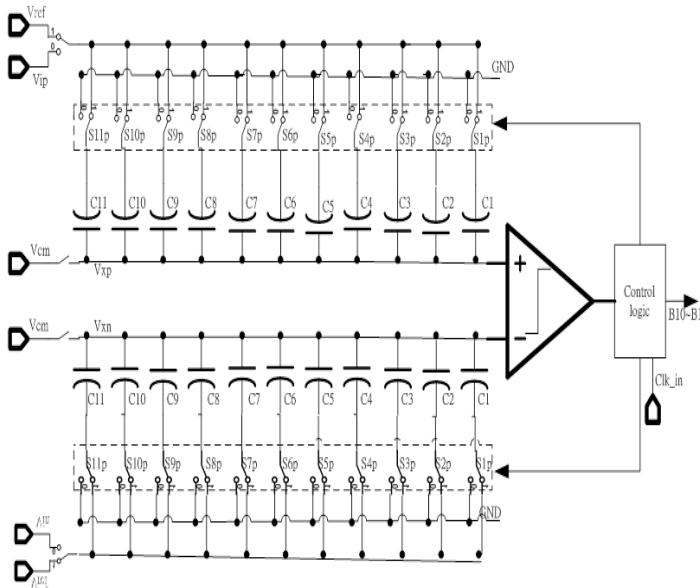


Figure 1. The conventional SAR ADC architecture

Figure 2. is a monotonic switching architecture. The monolithic switching method operating procedure is as follows: The input signal via the bootstrapped switch and input signal v_{ip} is switched to the capacitance DAC array the top plate on the positive side, which reduces the settling time and increases the input bandwidth. At the same time, the bottom plate capacitor is switched to VREF. Next phase, v_{ip} switch to floating and the comparator can directly compare the node voltage of both V_{xp} and V_{xn} without switching any capacitor. When the comparator input V_{ip} is greater than V_{in} , the comparator output “S10p” (MSB) is high. The maximum capacitance bottom plate is switched to GND on the positive side, and the maximum capacitance bottom plate remains unchanged on the negative side. The SAR ADC will continue to repeat this process until the LSB is determined. In this procedure, only one capacitor switch is switched to reduce charge conversion for each phase. In addition, the input signal is switched to the maximum capacitance top plate on the capacitance DAC array through the bootstrapped switch, so that the comparator can directly compare the node voltage both V_{xp} and V_{xn} . The number of unit capacitors is half of the conventional unit capacitor.

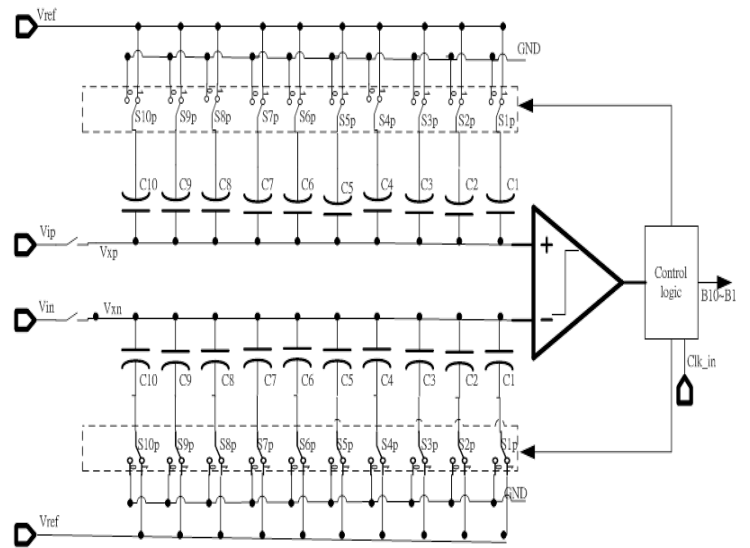


Figure 2. The monotonic switching architecture [2]

Figure 3. is a v_{cm} -based switching architecture. The operation procedure of the v_{cm} -based capacitor switching method is as follows: The input signal v_{ip} via the bootstrapped switch and is switched to the top capacitance DAC array on the positive side. The bottom plate capacitor is switched to common-mode voltage V_{cm} . Next phase, v_{ip} switch to floating and the comparator can directly compare the node voltage V_{xp} and V_{xn} without switching any capacitor. When the comparator input voltage V_{xp} is greater than V_{xn} , the “S10p” (MSB) is high. The maximum capacitor bottom plate is switched from V_{cm} to V_{ref} on the positive side and the maximum capacitance bottom plate is switched from V_{cm} to GND on the negative side. The SAR ADC will continue to repeat this process until the LSB is determined. Figure 4 is v_{cm} -based switching method flow chart.

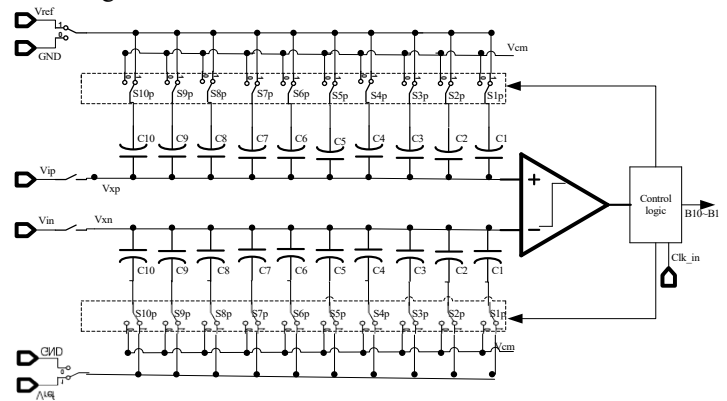


Figure 3. v_{cm} -based switching architecture

Moreover, the following is a list of different methods for average switching energy. The average switching energy of conventional switching method, monotonic switching method and v_{cm} -based capacitor switching method are $5459.3 CV_{ref}^2$, $1023.8 CV_{ref}^2$ and $341 CV_{ref}^2$ respectively.[4, 5]

$$E_{conventional,avg} = \sum_{i=1}^n 2^{n+1-2i} \cdot (2^i - 1) CV_{ref}^2 \quad (1)$$

$$E_{monotonic,avg} = \sum_{i=1}^n 2^{n-2-i} \cdot CV_{ref}^2 \quad (2)$$

$$E_{vcm\text{-based,avg}} = \sum_{i=1}^{n-1} 2^{n-3-2i} \cdot (2^i - 1) \cdot CV_{ref}^2 \quad (3)$$

Table I compares the number of switches, the number of unit capacitors, and switching energy for different methods.

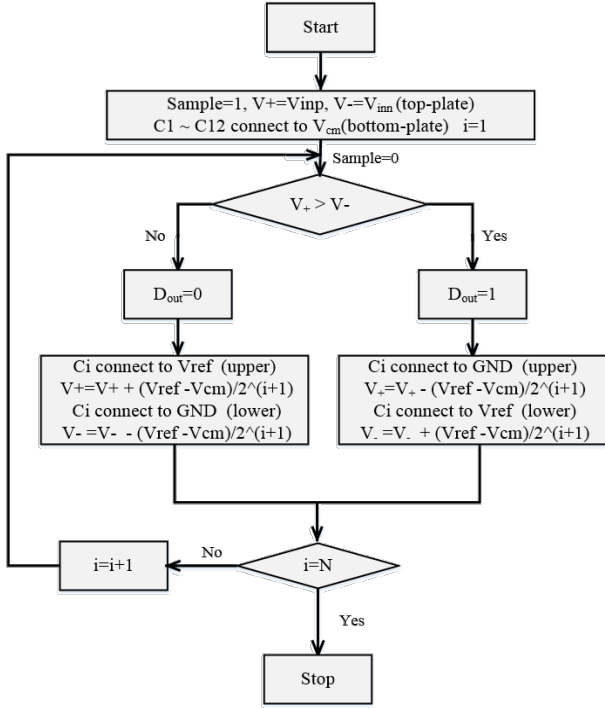


Table 1 Comparison of differentially switch capacitor method

Parameter	Conventional	Monotonic	V _{cm} -based
No. of Switches	6n	4n	4n
No. of Unit Capacitors	2 ⁿ	2 ⁿ⁻¹	2 ⁿ⁻¹
Switching Energy	“(1)”	“(2)”	“(3)”

2.2. Analysis of linearity

The conventional binary weighted capacitor array realizes with "radix of 2" and the total number of capacitors is 2ⁿ unit capacitors. The proposed partial thermometer coded (or segmented) capacitor array is divided into lower bits and higher bits to implement. The lower bit is implemented with "radix of 2" and the higher bits are implemented with the same capacitor size which is 2^{N-T-1} unit capacitor. The T is the number of higher bits. The total number of capacitors is same with the capacitor number of capacitors the conventional binary weighted capacitor array, as shown in Figure. 3. The advantage of the proposed partial thermometer coded based DAC is that it can make higher bits (MSB) have better linearity, can reduce DNL, ensure monotonic characteristics, and reduce the glitch phenomenon caused by voltage peak.

According to the binary weighted capacitor array, each weighted capacitance can be expressed as [6]

$$C_i = 2^{i-1}C_u + 2^{i-1}\sigma_u = 2^{i-1}C_u + \delta_i$$

where "i" is an integer representing bit position, "δ" is the error term. "C_u" is unit capacitor.

after the DAC passes n conversion phases, V_{xp} and V_{xn} can be expressed as the following expression, "D_n" is the digital output, n is the number of bits, and "C_p" is the parasitic capacitance.

$$V_{xp} = V_{inp} - \frac{(2D_n - 1) \cdot C_{n-1} + (2D_{n-1} - 1) \cdot C_{n-2} + \dots + (2D_2 - 1) \cdot C_1}{\sum_{i=0}^{n-1} C_i + C_p} \cdot (V_{ref} - V_{cm})$$

$$V_{xn} = V_{inn} - \frac{(1 - 2D_n) \cdot C_{n-1} + (1 - 2D_{n-1}) \cdot C_{n-2} + \dots + (1 - 2D_2) \cdot C_1}{\sum_{i=0}^{n-1} C_i + C_p} \cdot (V_{ref} - V_{cm})$$

ignore the parasitic capacitance and subtract V_{xp} and V_{xn} to get the error term, which can be expressed as follow

$$\Delta V_{x,binary} = V_{xp} - V_{xn} = V_{inp} - V_{inn} + V_{ref} - \frac{4D_n C_{n-1} + 4D_{n-1} C_{n-2} + \dots + 4D_2 C_1}{\sum_{i=0}^{n-1} C_i} \cdot \frac{1}{2} V_{ref}$$

then, subtracting the nominal value ΔV_{x,nominal} which means no error term of ΔV_{x,binary} from (B-5) expresses as follows

$$V_{error} = \Delta V_{x,binary} - \Delta V_{x,nominal} = \frac{2D_n \delta_{n-1} + 2D_{n-1} \delta_{n-2} + 2D_2 \delta_1}{2^{n-1} C_u} V_{ref}$$

the expected value of V_{error} is

$$E[V_{error}^2(y)] = E\left[\frac{2 \sum_{i=1}^{n-1} D_{i+1} \delta_i^2}{2^{2n-2} C_u^2} V_{ref}^2(y)\right]$$

where "y" is the digital output. Differential nonlinearity (DNL) is the difference of two adjacent code as shown in below:

$$DNL(y) = \frac{V_{err}(y) - V_{err}(y-1)}{LSB}$$

the maximum error is generated from 10...0 to 011...1, variance of the maximum DNL error can be expressed as

$$\begin{aligned} & E[V_{error}^2(100 \dots 0) - V_{error}^2(011 \dots 1)] \\ &= E\left[\left(\frac{2\delta_{n-1}^2}{2^{2n-2}C_u^2}\right) - \left(\frac{2\delta_{n-2}^2 + 2\delta_{n-3}^2 + \dots + 2\delta_1^2}{2^{2n-2}C_u^2} V_{ref}^2\right)\right] \\ &= \frac{2(2^{n-2}\sigma_u^2) - 2(2^{n-3}\sigma_u^2 + 2^{n-4}\sigma_u^2 + \dots + \sigma_u^2)}{2^{2n-2}C_u^2} V_{ref}^2 \\ &\approx \frac{\sigma_u^2}{2^{n-1}C_u^2} V_{ref}^2 \end{aligned}$$

$$DNL_{max(binary)} = \frac{\sqrt{E[V_{error}^2(100 \dots 0) - V_{error}^2(011 \dots 1)]}}{LSB}$$

$$= \sqrt{\frac{\sigma_u^2}{2^{n-1}C_u^2} V_{ref}^2} = \sqrt{2^{n-1}} \frac{\sigma_u}{C_u} \quad (4)$$

integral nonlinearity (INL) is the difference between the ideal code and the actual code as shown below:

$$INL(y) = \frac{V_{error}(y)}{LSB}$$

the maximum error occurs during the code in '100...0', so the maximum INL is shown a

$$INL_{max(binary)} = \frac{\sqrt{E[V_{error}^2(100 \dots 0)]}}{LSB}$$

$$= \sqrt{\frac{\sigma_u^2}{2^{n-2}C_u^2} V_{ref}^2} = \sqrt{2^{n-2}} \frac{\sigma_u}{C_u}$$

segmented DAC's higher bits MSB, MSB-1, MSB-2 are divided into 7 equal $2^{n-1-3} \cdot C_u$ capacitors. The maximum error is generated from 10...0 to 011...1, variance of the maximum DNL error can be expressed as

$$DNL_{max(segmented)} = \frac{\sqrt{E[V_{error}^2(100\dots0) - V_{error}^2(011\dots1)]}}{LSB} =$$

$$\frac{\sqrt{2(4 \cdot 2^{n-1-3} \cdot \sigma_u^2) - [2(3 \cdot 2^{n-1-3} \cdot \sigma_u^2) + 2(2^{n-5}\sigma_u^2 + \dots + \sigma_u^2)]}}{\frac{2V_{ref}}{2^n}}$$

$$\approx \sqrt{\frac{\sigma_u^2}{2^{n-1-3}C_u^2} V_{ref}^2} = \sqrt{2^{n-4}} \frac{\sigma_u}{C_u} \quad (5)$$

the quotient of “(4)” and “(5)”, can obtain variation of the variance of the maximum DNL error of the binary weighted DAC and the variance of the maximum DNL error of the segmented DAC as follows:

$$\frac{DNL_{max(binary)}}{DNL_{max(segmented)}} \approx \frac{\sqrt{2^{n-1}} \frac{\sigma_u}{C_u}}{\sqrt{2^{n-4}} \frac{\sigma_u}{C_u}} = \sqrt{2^3} \quad (6)$$

for typical metal-insulator-metal (MIM) capacitor, the standard deviation of capacitor mismatch can be derived as

$$\frac{\sigma_u}{C_u} = \frac{K_\sigma}{\sqrt{A} \cdot \sqrt{2}} \text{ and } C_u = K_C \cdot A \quad (7)$$

where "K_σ" is the mismatch coefficient, "A" is capacitor area and "K_C" is the capacitor density parameter. Inserting “(7)” into “(6)” give the capacitor area ratio is 2³. From the results, the DNL in segmented capacitor array is smaller than DNL in binary-weighted array and the size of capacitor array in segmented is smaller than binary-weighted array. In addition, the INL doesn't change. So, the INL becomes to the main influencing factor and then we can re-calculate the value of unit capacitor through 4.5 INL_{max} < 1/2 LSB again. At last, a minimum unit capacitor is about 17.83 fF in 12-bit situation.

3. ADC Architecture Design

3.1. System architecture

The segmented SAR ADC system architecture is shown in Figure. 5., divided into sample and hold ,bootstrap switch, dynamic two-stage comparator, synchronous digital control logic including multiplexer and shift register + vcm-based control logic, 3 MSB segmented capacitive DAC including capacitor array and switch, level shifter, resolution Scale control(RS).

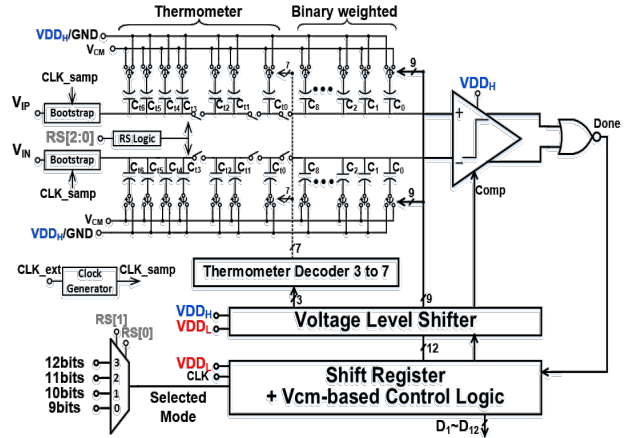


Figure. 5. The segmented SAR ADC system architecture

3.2. Sample and hold

The bootstrap switch with body effect reduction is shown in Figure. 6., which perform sample-and hold function. The input signal is rail-to rail, and must suppress the distortion to at least 12 bits. The operation of the bootstrap switch with body effect reduction is as shown in Figure. 7. [7]. When clk=high, input signal is Vin and the node VG voltage is fixed at the voltage AVDD+Vin. So that the on-resistance of MOS (“SW”) on-resistance keep a small constant value, which can improve the linearity. In addition, “M15” is turned on to make the bulk node of MOS (“SW”) is connected to Vin, which can cancel the body effect. Therefore, this can reduce the significant distortion. When clk=low, the bulk of MOS (“SW”) will be connected to GND to avoid back-gate driven. According to simulation and the sample rate 50 Ks/s and and Nyquist input frequency in 8192 sample point, taking FFT(fast Fourier transform) for bootstrap switch and output capacitor and get SNDR is greater than 87.2 db, ENOB(effective number of bits) is greater than 14.29 bits. The relation of SNDR to ENOB equation can be derived, SNDR=6.02*ENOB+1.76.

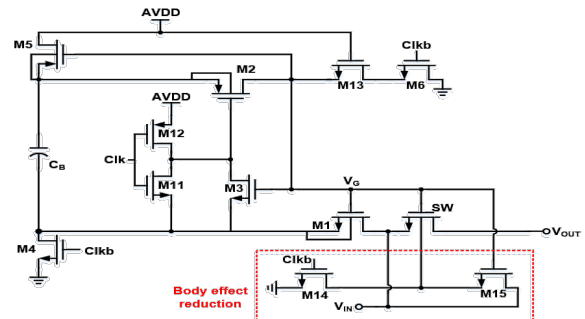


Figure. 6. The bootstrap switch with body effect reduction

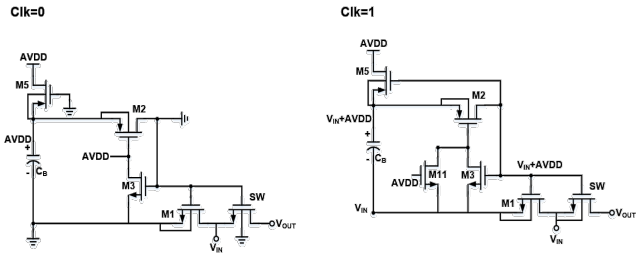


Figure 7. The operation of the bootstrap switch with body effect reduction

3.3. Scalable resolution design

The proposed segmented capacitor array is shown Figure 8. The 3 MSB capacitor bottom plate is divided into seven equal 2^{n-1-T} capacitances and the T is 3. Moreover, the seven equal capacitances can be controlled by three-to-seven bit binary-to-thermometer decoder logic. The remaining switches can be controlled by binary weighted mode. Insert the switch on the MSB capacitor top plate to decouple from the other capacitor array, using these insertion switches to divide into different resolutions and the corresponding FOM is obtained. Moreover, the resolution signal (RS1, RS2) is used to control insertion switch on the 3 MSB capacitor top plate and the control method is two-to-three bit binary-to-thermometer decoder. The three-to-seven binary-to-thermometer logic expression is as follows.

$$T_6 = D_1 \cdot D_2 \cdot D_3 \quad T_5 = D_1 \cdot D_2 \quad T_4 = D_1 \cdot (D_2 + D_3)$$

$$T_3 = D_1 \quad T_2 = D_1 + (D_2 \cdot D_3) \quad T_1 = D_1 + D_2 \quad T_0 = D_1 + D_2 + D_3$$

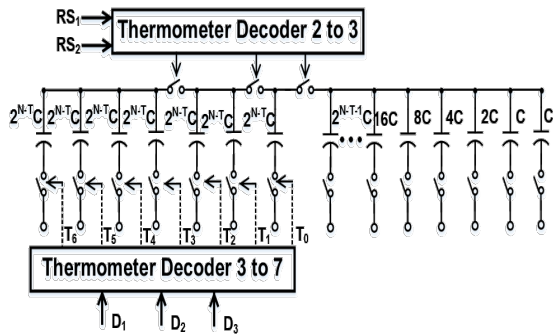


Figure 8. The proposed segmented capacitor array

3.4. Control logic and multiplexer

The synchronous clock control is adopted in the proposed SAR ADC and is shown Figure 9. The external clock (CLK_COMP) triggers the comparator and input in the clock divider to generate sampling rate (CLK_SAMP) of system. In addition, extra NOR gate is used to detect the result of comparator. When CLK_COMP is low in the reset mode, the output V_{op} and V_{on} are both low and the Done signal is high. When CLK_COMP is high in the comparison mode, the output V_{op} and V_{on} are either high or low and the Done signal goes low that means the comparison is finished.

The shift signal $S[1:13]$ shows the current conversion cycle and will trigger the corresponding digital control logic and capacitor switching signal as shown in Figure 10. The multiplexers control the necessary shift signal to pass through shift register and stop each conversion with resolution scale ($RS[1], RS[0]$) signal. For

9 to 12 bit mode, after the corresponding shift signal $S[10]$ to $S[13]$ goes to high, the STOP signal will be triggered and reset all the block of SAR ADC for avoiding the waste of power consumption.

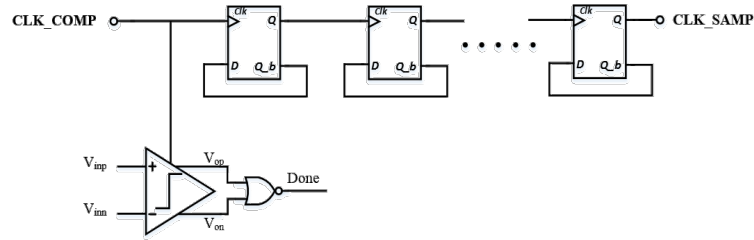


Figure 9. Synchronous clock generator

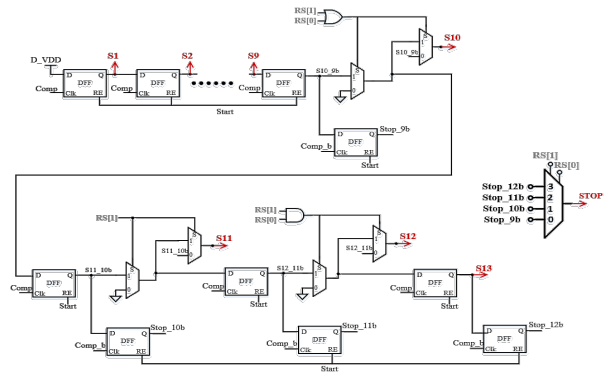


Figure 10. Shift register with multiplexer

The synchronous clock waveform of digital circuit is shown Figure 11. The clock waveform shows that START goes to high when sampling end. Then the first COMP_CLK goes to high and so does $S[1]$, which means SAR ADC is in first conversion cycle. The second period of COMP_CLK is also the same way to enter to the next conversion. In the end, the last COMP_CLK goes to low and STOP will rise to high to finish the number bit-mode of conversion.

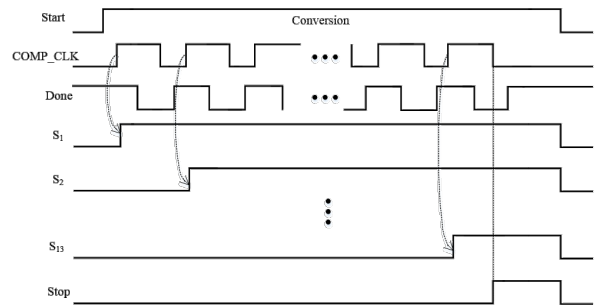


Figure 11. Synchronous clock waveform of digital circuit

The control logic circuit of V_{cm} -based switching is shown Figure 12. For the first comparison cycle, we use XOR gate with $S[1]$ and $S[2]$ to generate $X1$ that is the range of first comparison cycle. Before the first result of comparator is decided, the MSB capacitor is still connected to V_{cm} . Therefore, reuse the $X1$, COMP_CLK and Done input to the AND gate to generate the CLK_V_{cm1} . After the first result of comparator is decided, MSB capacitor is connected to V_{ref} or GND depending on the output of comparator in the rest of the comparison cycle. Here we utilize

the XOR gate with X_{i2} and $CLK_V_{cm_i}$ to generate $CLK_V_{ref_i}$. X_{i2} is the signal created by $S[1]$ and $S[13]$ with XOR gate, and it represents the range of first to last comparison cycle. From the second comparison cycle to the last comparison cycle, the same methods are used to generate $CLK_V_{CM_2}$ to $CLK_V_{CM_{12}}$ expect for addition of the OR gate with the previous X_i comparison cycle to keep the capacitor in V_{cm} voltage. $CLK_V_{ref_2} \sim CLK_V_{ref_{12}}$ also utilize the XOR gate with X_{i2} and the current $CLK_V_{cm_i}$ to generate.

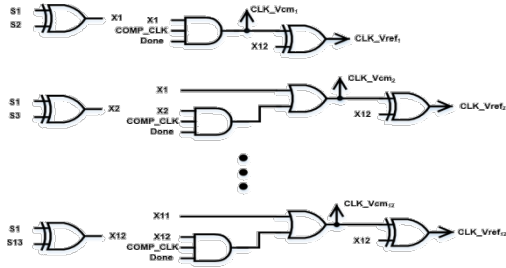


Figure. 11. Control logic circuit of V_{cm} -based switching

Schematic and timing diagram of the DAC control logic based on [2] is shown in Figure. 12. At the rising edge of $COMP_CLK$, a D-flip-flop samples the output of the comparator at the current conversion. If the output is high, the relevant capacitor is switched from V_{cm} to GND. If the output is low, the relevant capacitor is switched from V_{cm} to V_{ref} . After the decision signal of control voltage is confirmed, the level shifter should be used to switch the digital voltage domain to analog voltage domain that makes correct charge distribution. The output buffer stores the digital output code decided by comparator and reveals when the $STOP$ signal triggers in the end of the bit-conversion.

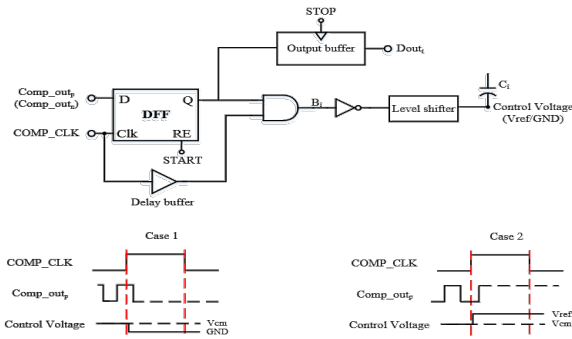


Figure. 12. DAC control logic

3.5. Scalable voltage design

The system architecture energy analysis can be simplified into analog block, comparator, digital block, level shift. The system architecture of the SAR ADC with the Dual supply voltage is shown Figure. 13. Since the level shift is the bridge between analog and a digital and can be divided into analog block and digital block. The energy-per-conversion is given by

$$E_{DIG,CLK} = C_L(n) \cdot V_{DD,digital}^2 \quad (8)$$

where " $C_L(n)$ " is the effective capacitance being charged and discharged, and is relative to the resolution .

In [8], the energy of a dynamic regenerative comparator is derived. When applied to a n-bit SAR ADC that requires n comparisons per conversion, the comparator energy-per-conversion can be derived as [9]

$$E_{COMP,reset} + E_{COMP,reg} = nC_{load}V_{DD,analog}^2 + 2 \ln 2 \cdot n^2 \cdot C_{load}V_{eff}V_{DD,analog} \quad (9)$$

where " C_{load} " is the capacitive load of comparator, " V_{eff} " is the transistor overdrive voltage, and " V_{DD} " is the power supply of comparator. The energy is also proportional to the analog supply voltage.

The V_{cm} -based average switching energy of DAC we use in this ADC design is shown in formula "(3)".

$$E_{vcm-based,avg} = \sum_{i=1}^{n-1} 2^{n-3-2i} \cdot (2^i - 1) \cdot CV_{ref}^2$$

The energy is proportional to V_{ref}^2 and capacitor size. Capacitor size depends on the limitation of capacitor mismatch and V_{ref} is corresponding to the power supply voltage of comparator. According to formula "(3)", "(8)", "(9)", scaling down V_{DD} is an effective method to reduce energy.

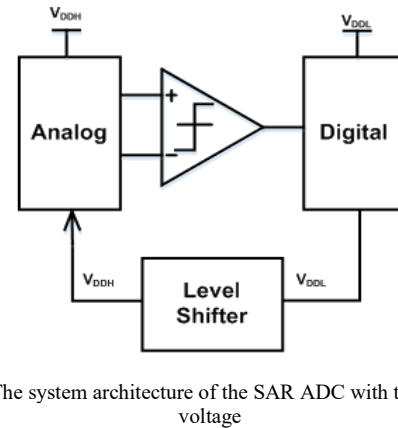


Figure. 13. The system architecture of the SAR ADC with the Dual supply voltage

3.6. Comparator

The comparator consists of two stage is shown in Figure. 14. The first stage Preamp has a fixed current source (" $M5$ ") and the current source operates in saturation. Consequently, the drain current is only slightly changed and the dynamic offset is only slightly changed. The pre-amp's voltage gain is about 5~10 and helpful to reduce the second-level input referred noise. In addition, the input pair operates in weak inversion to achieve lower input referred noise. The second stage consists of a simple voltage amplifier and a positive feedback amplifier that makes the output reach rail-to-rail. The operation of dynamic comparator has two phases. In reset phase, the CLK is low voltage and the node TI_N , TI_P has been pre-charged to V_{DD} by device " $M3$ ", " $M4$ ". The output V_{OP} , V_{ON} of comparator has been discharged to ground by device " $M13$ ", " $M14$ ". When it comes to comparison phase, the CLK goes to high voltage. CLK enables " $M5$ " which producing a current path and starts to discharge the capacitors on node TI_N , TI_P through " $M1$ ", " $M2$ ". " $M11$ " and " $M12$ " are used as a switch to sense the voltage difference between input signal V_{IP} and V_{IN} .

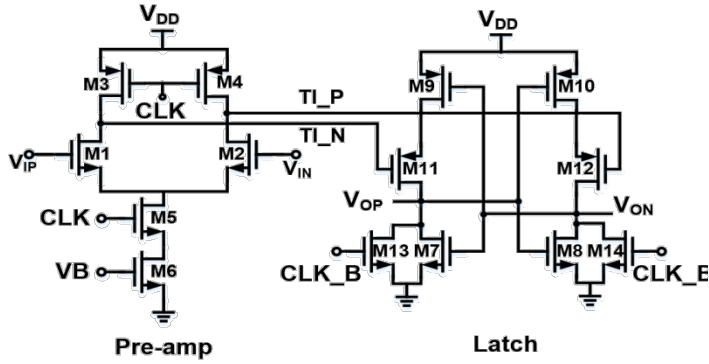


Figure. 14. Dynamic two-stage comparator with a current source

In comparison phase, the input-referred noise can be estimated by [10]:

$$\sigma_v = \sqrt{4 \cdot k \cdot T \cdot \frac{2}{g_{m1,2}} \cdot \frac{1}{2 \cdot T_{int}}} \quad (10)$$

where “k” is Boltzmann constant, “T” is the Kelvin temperature which is set to 300 K, transconductance “ $g_{m1,2}$ ” and integration time “ T_{int} ” is shown below:

$$g_{m1,2} = \frac{I_{MOS}}{2 \cdot V_{thermal}} \quad (11)$$

$$T_{int} \approx \frac{V_{threshold} \cdot C_{TI_P}}{I_{MOS}} = \frac{V_{threshold} \cdot C_{TI_N}}{I_{MOS}} \quad (12)$$

where “ $V_{thermal}$ ”, the thermal voltage, is equal to 25 mV, “ $V_{threshold}$ ” is threshold voltage of input pair equals to 460 mV, C_{TI_P} and C_{TI_N} is the parasitic capacitance of node. TI_P and TI_N . Inserting “(11)” and “(12)” into “(9)” gives

$$\sigma_v = \sqrt{\frac{k \cdot T}{C_{TI_P}}} \cdot \sqrt{8 \cdot \frac{V_{thermal}}{V_{threshold}}} \quad (13)$$

At 12-bit mode with 1.8 analog supply voltage, 1/2 LSB is equal 0.44 mV. It means that “ σ_v ” should be designed smaller than 0.44 mV. According to (13), re-derived C_{TI_P} , C_{TI_N} as follow:

$$C_{TI_P} = C_{TI_N} \geq \frac{k \cdot T}{\sigma_v^2} \cdot 8 \cdot \frac{V_{thermal}}{V_{threshold}}$$

substituting all the value that mentioned before, C_{TI_P} , C_{TI_N} can be designed over 11 fF to let the thermal noise of comparator in the same order as quantization noise of the ADC.

For V_{cm} -based switching method, the input common voltage of comparator maintains at V_{cm} voltage that the offset belongs to static offset and it doesn’t affect the accuracy but it will decrease input range, thus degrading the signal-to-noise ratio [11]. From Monte-carol simulation, the comparator offset for 1.8 V, 1.2 V, 1.0 V is 7.83 mV, 8.28 mV and 7.86 mV respectively. The formula of SNR can be derived as follow:

$$SNR = 20 \times \log \frac{V_{in(rms)}}{V_{Q(rms)}} = 20 \times \log \frac{V'_{ref}}{2\sqrt{2} \cdot \frac{V_{LSB}}{\sqrt{12}}}$$

where “ V'_{ref} ” is the input voltage range influenced by offset and “ V_{LSB} ” is the least significant bit (LSB) voltage. Compare without offset simulation and the SNR of with offset simulation decrease by 0.04, 0.06 and 0.07 dB respectively. Therefore, the effect of ENOB is little influence.

3.7. Level shifter

The conventional level shifter has large delay and power consumption. The CMLS (contention mitigated level shift) has less delay and power consumption, the reason is that “MN1”, “MP3”, “MN2”, “MP4” form quasi-inverter, the node OUT voltages are pulled faster than conventional level shifter. The schematic of level shifter is shown in Figure. 15. The two PMOS act as a swing-restoring load. Assuming the input signal, IN , is low, “MN1” is turned on and provides a conducting path to ground while “MN2” is cut off. Therefore, OUT is pulled down to ground. The operation reverses when the input signal, IN , is switched to high

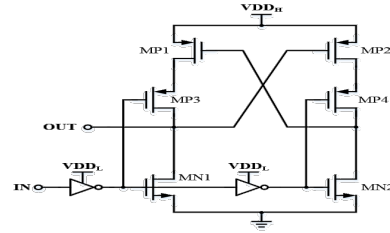


Figure. 15. Schematic of level shifter [12]

Table 2 shows comparison of the SAR ADC simulation result

	This work				[4]	[13]
	12 bit	11 bit	10 bit	9 bit	10 bit	12 bit
Resolution	12 bit	11 bit	10 bit	9 bit	10 bit	12 bit
Technology (um)	0.18				0.13 □	0.18 □
Supply voltage(V) (Analog/Digital)	1.8/0.9	1.2/0.9			1.0/0.4	1.8/1.8
Area(mm²)	0.35				0.19	2.38
Sampling Rate(KS/s)	50				1	200
SNDR(dB)	68.6	63.2	57.7	52.3	56.54	69.6
Power(uW)	9.7	3.9	2.8	2.5	0.05	41.5
FoM^a(fJ/conversion)	88.4	66.3	88.8	148	94.5	84.6

4. Result and Discussion

This paper presents a reconfigurable SAR ADC for multi-sensor application. The transistor level simulation is operated by Cadence Spectre for 1P6M 0.18 um CMOS technology. The maximum sampling rate is 50 KS/s, and input frequency is at Nyquist rate and gets 256 number of points for the FFT analysis. Analog supply voltage is at 1.8 V (12 bit), 1.2 V (11, 10, 9 bit). Digital supply voltage is at 0.9 V and the clock duty cycle is 50%. The reconfigurable SAR ADC has achieved 4 mode including 12/11/10/9 bits and corresponding performance SNDR are 68.6/63.2/57.7/52.3 dB at input frequency (f_{in}) 6.25/25/25/25 KS/s respectively. Table II shows comparison of the SAR ADC result.

For the accuracy of capacitor array, the placement of the DAC are arranged by common-centroid layout to enhance matching and the dummy capacitors are added around capacitor array to keep from etching effect around edge. The layout of capacitor array separate into two parts, one is for 10 to 12 bit that T0 to T6 means the corresponding capacitor controlled by thermometer decoder and the other is for the 1 to 9 bit. The overall core area is 810 x 430 μm^2 . Layout of capacitor array is shown Figure. 16. And the segmented SAR ADC layout plan including capacitor array, switch array, comparator, sample-and-hold(S/H), SAR logic is shown Figure. 17.

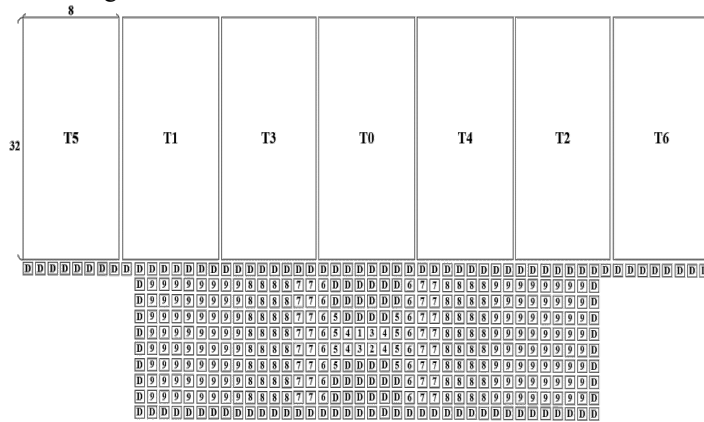


Figure. 16. Layout of capacitor array

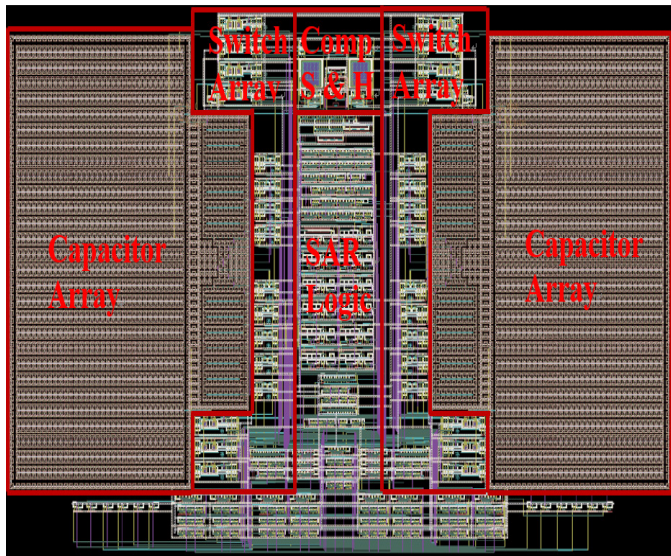


Figure. 17. The segmented SAR ADC layout plan

Acknowledgment

This research was sponsored in part by the National Science Council of Taiwan under grant of MOST 106-2911-I-009-301. The authors appreciate the UMC and the National Chip Implementation Center (CIC), Taiwan, for supporting the CMOS chip manufacturing.

References

[1] H.-M. Lin and K.-A. Wen, "A low power reconfigurable SAR ADC for CMOS MEMS sensor," in *SoC Design Conference (ISOC), 2017 International*, 2017, pp. 7-8.

[2] C. C. Liu, S. J. Chang, G. Y. Huang, and Y. Z. Lin, "A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 4, pp. 731-740, 2010.

[3] V. Hariprasath, J. Guerber, S. H. Lee, and U. K. Moon, "Merged capacitor switching based SAR ADC with highest switching energy-efficiency," *Electronics Letters*, vol. 46, no. 9, pp. 620-621, 2010.

[4] S. Haenzsche and R. Schüffny, "Analysis of a charge redistribution SAR ADC with partially thermometer coded DAC," in *2013 European Conference on Circuit Theory and Design (ECCTD)*, 2013, pp. 1-4.

[5] Yuan-Fu Lyu, "A Low Power 10-Bit 500-KS/s Delta-Modulated Successive Approximation Register Analog-to-Digital Converter for Implantable Medical Devices," Master, Institute of Electronics, National Chiao Tung University, Hsin chu, 2012.

[6] Liao, Bo-Shi, "Power-Efficient Successive-Approximation Register Analog-to-Digital Converter," Master, Institute of Electronics Engineering, National Taiwan University, 2016.

[7] S. Wang and C. Dehollain, "Design of a rail-to-rail 460 kS/s 10-bit SAR ADC for capacitive sensor interface," in *2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS)*, 2013, pp. 453-456.

[8] Z. Dai, C. Svensson, and A. Alvandpour, "Power consumption bounds for SAR ADCs," in *2011 20th European Conference on Circuit Theory and Design (ECCTD)*, 2011, pp. 556-559.

[9] M. Yip and A. P. Chandrakasan, "A Resolution-Reconfigurable 5-to-10-Bit 0.4-to-1 V Power Scalable SAR ADC for Sensor Applications," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 6, pp. 1453-1464, 2013.

[10] M. v. Elzakker, E. v. Tuijl, P. Geraedts, D. Schinkel, E. A. M. Klumperink, and B. Nauta, "A 10-bit Charge-Redistribution ADC Consuming 1.9 mW at 1 MS/s," *IEEE Journal of Solid-State Circuits*, vol. 45, no. 5, pp. 1007-1015, 2010.

[11] D. Zhang, A. Bhide, and A. Alvandpour, "A 53-nW 9.1-ENOB 1-kS/s SAR ADC in 0.13- μm CMOS for Medical Implant Devices," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 7, pp. 1585-1593, 2012.

[12] C. Q. Tran, H. Kawaguchi, and T. Sakurai, "Low-power high-speed level shifter design for block-level dynamic voltage scaling environment," in *2005 International Conference on Integrated Circuit Design and Technology, 2005. ICICDT 2005.*, 2005, pp. 229-232.

[13] Z. Yan, U. F. Chio, W. He-Gong, S. Sai-Weng, U. Seng-Pan, and R. P. Martins, "Linearity analysis on a series-split capacitor array for high-speed SAR ADCs," in *2008 51st Midwest Symposium on Circuits and Systems*, 2008, pp. 922-925.

A Novel Technique for Enhancing Color of Undersea Deblurred Imagery

Chrispin Jiji*, Nagaraj Ramrao

Department of Electronics and Communication, The Oxford College of Engineering, Oxford Institutions, Bangalore, India

ARTICLE INFO

Article history:

Received: 18 September, 2018

Accepted: 23 October, 2018

Online: 01 November, 2018

Keywords:

Underwater

Image Deblurring

Image Enhancement

ABSTRACT

Exploring the ocean underneath has always been an area of great scientific and environmental concern. However, the study of underwater environment was very difficult due to the extreme conditions. Undersea descriptions undergo severe distortion attributed to absorptive as well as scattering properties. Absorption substantially removes illumination, whereas a ray of light redirected in several path when it interacts by substance. Because of these, undersea descriptions encompass blur as well as color loss. In this paper we suggested an effective technique namely, a turbidity removal method for deblurring the image. If the deblurred image has a lighting problem, we make use of a color-correction method to find the clear image. Our substantial qualitative and quantitative assessment expose that the proposed algorithm progress the excellence as well as lessen color distortion loyally, also improves the state-of-the-art undersea technique.

1. Introduction

Images captured in undersea has plays a vital basis of interest within various branches of technical and systematic explores [1], such as examining underneath infrastructures [2] as well as cables [3], detecting manmade objects [4], managing undersea vehicle [5], marine biology investigate [6], and archaeology [7]. Apart of normal descriptions, undersea descriptions undergo reduced visibility ensuing attenuation of the propagated illumination, mainly owing to absorption along with scattering effect. In this paper, we use image processing acting extensive interest over earlier years due to its challenging nature and its importance for the surroundings. Improving undersea scene excellence separates the problem into image restoration and image Enhancement

Visibility in undersea imagery is usually blurry, but having large number of particles underneath cause's cloudiness or more haziness, called turbidity, which causes blur in undersea imagery. To remove any blur, we usually use restoration problem with estimated or known PSF matrix.

$$b = h * d + n \quad (1)$$

where d represents deblurred representation, h denotes PSF kernel, and b denotes turbid representation. The main challenge is estimating blur kernel [8-12]. The blur kernel cannot be estimated directly as it varies depending on blur itself. Existing system uses some prior to estimate point spread function (PSF) for restoration of undersea blurred imagery.

If the deblurred image has any difficulty in lighting, we use image enhancement to enhance the picture excellence. The difficulty during lighting owing towards absorption substantially reduces illumination, practically, in undersea images shown in Figure 1. The objects by remoteness more than 10m are about imperceptible, and colors go down by the deepness of water.

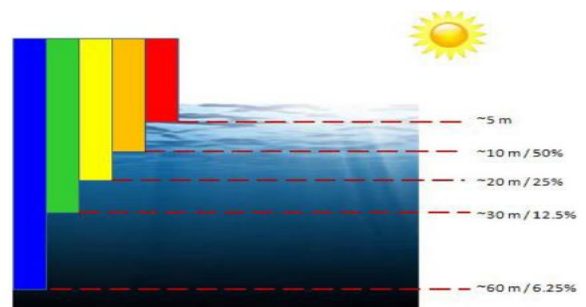


Figure 1: Dissimilar illumination weakened by dissimilar charge

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Do not use abbreviations in the title or heads unless they are unavoidable. In general, red illuminations vanish by deepness of 5m, after that brown followed by yellow radiance, lastly green as well as blue illumination vanish by deepness with 30m as well as 60m. Thus undersea imagery is subject to blue-green color which changes the picture excellence. As a result, we used color adjustment scheme to compensate light condition. There have been several attempts for restoration as well as enhancement, since degraded picture

*A. Chrispin Jiji, 8951627124 & chrispinjiji@gmail.com

outcomes the understanding of multiplicative by means of additive process [13]. Conventional enhancement methods namely gamma correction, histogram equalization strongly restricted for such a task. This difficulty deal with modified attainment approach via various imagery [14], specific hardware [15] or polarization filters [16]. Even though their important attainment, these approach undergo numeral concern that decrease their convenient applicability.

As shown in Figure 2, our technique uses two stages namely image restoration stage for deblurring and enhancement stage for enhancing its excellence. In this paper we suggest an adaptive sparse domain selection (ASDS) system to restore the image. By training compact sub-dictionaries from high quality example image cluster the window. Since each cluster employ several windows by related prototype, compact sub-dictionary learns for every cluster. Particularly, for simplicity we use principal component analysis (PCA) technique towards learning sub-dictionaries. The most excellent sub-dictionary that is mainly applicable to given window is chosen, because the given window is better represent sub-dictionary is accurately reconstruct entire image. Besides sparsity regularization, other terms as well initiate for improving its performance. Later use autoregressive (AR) models, pre-learned from training dataset characterizing confined structures. For every confined window, we choose the AR model. On the other hand, considering fact that there are often several repetitive image structures in an image, we introduce a non-local (NL) self-similarity constraint served as another term, helpful for preserving boundary sharpness and restrain noise. After introducing ASDS using two constraints (AR & NL) into IR structure, we present a weighted Gray Edge method for solving lighting problem. Substantial experimentation on image deblurring and enhancement show that the projected approach effectively reconstructs picture details along with excellence, outperforming different state-of-the-art IR with IE methods in terms of both excellence metrics in addition to ocular insight.

The paper is planned as follows. Section 2 provides concise outline about earlier art. Section 3 signifies our restoration approach, about turbidity removal method especially used for undersea descriptions. Section 4 describes novel enhancement approach, mainly to improve its excellence. Section 5 presents comparative qualitative and quantitative estimation of undersea system and Section 6 provides closing comments.

2. Previous Art

This part reviews main advances to deblur or else improve imagery confined undersea. In computer vision, methods to handle ambiguity is roughly on some knowledge or assumptions known beforehand, *i.e.*, *priors*. The priors impose extra constraints/dependency among the unknown variables. In the following, we review the previous turbidity removal methods. We do not discuss the technical details among those methods. Instead, we concern about the extra constraints. All methods reformulate in a same framework expressed indifferent forms in the original works. Single image turbidity removal methods have to rely on some priors.

The prior is statistical/physical properties, heuristic assumptions, simplifications, along with application-based rules. The blur imaging model in (1), discrepancy between some equations and unknown, the prior expected to introduce at least

one constraint for each pixel. The challenge of recovering d from b is under-constrained. To make it solvable, extra knowledge has been built-into the restoration method. The former understanding is often built-in with a regularization term, principal to the later energy minimization problem:

$$\hat{d} = \arg \min \left\{ \|b - Hd\|_2^2 + \lambda J(d) \right\} \quad (2)$$

where λ denotes Lagrangian multiplier matching the exchange among former term $J(d)$ along with likelihood $\|b - Hd\|_2^2$. The former term $J(d)$ act as a main part in the restoration method.

Sparsity based IR process guess that the natural representation is sparse in few fields. Sparse representation lately has paying attention to investigators for resolving complications including deblurring, denoising, super resolution. In [17], the ARM prototypes are in the neighborhood calculated commencing an originally expected image besides bring about much better-quality for TV prior trendy reforming boundary associations. Now, we resolve to suggest knowledge established using adaptive prior, where the AR prototypes remain knowledgeable commencing great feature of training imageries, towards raising AR forming accuracy. The approach presents exact modest: the blotches that make sure alike forms can exist spatially distant, besides gather the entire appearance. In [18-20], the NLM prior combined through the sparse domain, as of alike appearance blotches be concurrently implied towards the strength of converse reestablishment.

Image enhancement based methods are not required to solve the physical form, but rather directly enhance its contrast as well as improving its excellence from human visual perception. Within undersea, color is extremely related by deepness, and a significant difficulty is green-bluish form desires to resolve. Since as the illumination go through undersea, reduction process affects wavelength spectrum, thus affecting the gray level along with appearance of colored surface. Existing white balancing process is a key to our domain. Next, we briefly change those techniques and give details for our novel approach projected by undersea imagery. Generally a scheme formulates exact guess to estimate color, in addition to meet color constancy via color channel will standardize the radiance. Projected scheme use a statement that majority of patch show off extremely small gray level in at-least single color channel that openly about blur mass and progress vivid colors. It cannot sufficiently deal with color deformation and complex arrangement. The deblurred imagery will attribute color swing with artifact effect. The Grey World system [21] believe that the averages of the three color constituent, illuminate during impartial radiance basis further acquire poorer color patch. The Max-RGB [21] resting on the RGB color. Shades-of-Grey [22] use Minkowski norm-p is achromatic. The Grey Edge method [23] employs items boundary data towards adjusting color fidelity; it assumes that normal derivative of color mechanism is achromatic. Every color provides pre procedure by Gaussian filter is typically 1-2 toward determining the statement. It might worsen the performance. Weighted Grey Edge method [24] use boundary data of variety of substance, intended for instance, shadow and mechanism. This paper is an extended work initially presented during ICPCSI [25]. Also the journal paper intended novel technique for enhancing undersea imagery. Revised result much

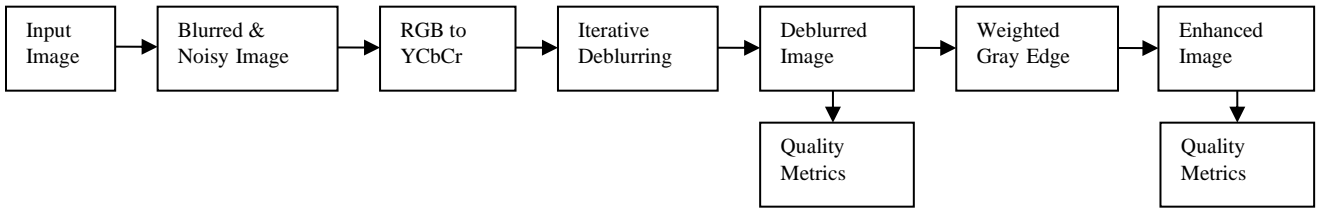


Figure 2: Block diagram of Proposed Method

improves the lighting problem in undersea imagery. Enhancement based methods not required to solve physical form of degraded image, but rather directly enhance contrast and improve image quality from human visual perception.

3. Underwater image restoration for turbidity removal method

Underwater turbidity removal is very challenging with essential part of picture suffers from turbidity. Several algorithm deals with turbidity elimination process recognized as Image deconvolution (ID). To get back the undersea deblurred picture, we used [26-27]. Here we adaptively study compacted sub-dictionaries to every confined patch. Let $d_m \in R^n$ denote the image block dimensions $\sqrt{n} \times \sqrt{n}$ attained from deblurred picture, we can describe $d_m = R_m D, m=1,2,\dots,N$ wherever R_m denotes matrix from d_m . Rendering towards sparse prior d_m can be signified by means of outmoded dictionary, namely $d_m = \phi_k \alpha_m$, thus D can be conveyed by way of

$$D = \left(\sum_{m=1}^N R_m^T R_m \right)^{-1} \sum_{m=1}^N R_m^T \phi \alpha_m \quad (3)$$

For convenience, we define:

$$D = \phi \alpha = \left\{ \sum_{m=1}^N R_m^T R_m \right\}^{-1} \sum_{m=1}^N R_m^T \phi \alpha_m \quad (4)$$

Now, the primary restored image denoted as $D^{(k)}$ constantly drop some information in the input image and comprises some artifacts. In order to further appropriate for human visual scheme, we update the restored image in every iteration.

For each patch we adaptively choose the sub-dictionary based on least distance specified through

$$k_m = \left\| d_m^h - \mu_k \right\|_2^2 \quad (5)$$

3.1 Adaptively Reweighted Sparsity Regularization

Existing work uses $\lambda=1$ for whole patches. In our work λ differs from patch to patch to expand the excellence of restored image. The reweighted sparsity regularization is specified in place of

$$\hat{\alpha} = \left\{ \left\| b - H \phi \alpha \right\|_2^2 + \sum_{m=1}^N \sum_{n=1}^o \lambda_{m,n} \left| \alpha_{m,n} \right| \right\} \quad (6)$$

Then weight λ_{mm} stands as

$$\lambda_{m,n} = \frac{2\sqrt{2}\sigma_{n1}^2}{\sigma_{m,n} + \varepsilon} \quad (7)$$

Reweighted procedure is as follows:

- Step 1 Fix $\ell = 0$ as well as $\lambda_n^l = 1$
- Step 2 Resolve the subjective l_1 minimization problematic $d^{(l)} = \arg \min \left\| \lambda_n^l \phi^T d \right\|_1$ Focused towards $d = \phi \alpha^{(l)}$
- Step 3 Set $\alpha^{(l)} = \phi^T d^{(l)}$ and describe $\alpha_n^{(l+1)} = \frac{2\sqrt{2}\sigma_{n1}^2}{\sigma_{m,n} + \varepsilon}$
- Step 4 Proceed until l reaches l_{\max} Else, increase l and drive towards step 2

3.2 Auto Regressive model

The ASDS progress extensively by the usefulness of sparse modeling and later outcome image deblurring. To further get better excellence of deblurred imagery, we established autoregressive (AR) models to formalize the image confined smoothness. For each high pass filtered patch adaptively choice AR dictionary by means of calculating the Euclidean distance. For each present pel we make sure four neighbouring pels and its weights are specified as;

$$w_{m,n} = \frac{1 / \left(\left\| d_m^h - \mu \right\|_2^2 + \varepsilon \right)}{\sum 1 / \left(\left\| d_m^h - \mu \right\|_2^2 + \varepsilon \right)} \quad (8)$$

Therefore, the weighted sum of entire adjacent pels expressed as

$$\chi_m = \sum w_{mn} a_{k_m} \quad (9)$$

Assemble the exceeding matrix into 3x3 window we resolve the value of local geometry model. Let d_m remains the center pel, and χ_m stays the vector comprising of the adjacent pels which is adjacent to middle pel d_m , now the best middle pel value would

lessen $\left\| d_m - a_{k_m}^T \chi_m \right\|_2^2$. For the ease of expression we write $\sum_{d_i \in d} \left\| d_m - a_{k_m}^T \chi_m \right\|_2^2$ as $\left\| (I - A) \phi \alpha \right\|_2^2$. By integrating this restraint, the restoration equation as follows

$$\hat{\alpha} = \left\{ \|b - H\phi\alpha\|_2^2 + \sum_{m=1}^N \sum_{n=1}^o \lambda_{m,n} |\alpha_{m,n}| + \gamma \|(I - A)\phi\alpha\|_2^2 \right\} \quad (10)$$

where γ represents a constant balance the role of AR formalization term.

3.1 Nonlocal Self Similarity model

The AR model utilizes the confined information in every patch. On the other hand, there are often numerous cyclical prototypes all through the representation. Thus non-local redundancy mainly improves the excellence of imagery. Along with AR models an additional term called non-local similarity added into the IR structure. The source of global mean filtering is very simple, aimed at current image block; we catch numerous alike blocks to restrain it. Aimed at each image block d_m , we discover all image blocks alike to it in the complete image. Take

on d_m^s is one of blocks similar to d_m , then $e_m^s = \|d_m^s - d_m\|_2^2$ would be small enough. If we estimate the summation of entire midpoint pels voguish these alike blocks, besides effects would satisfy

$$d_m \approx \sum_{s=1}^L b_m^s d_m^s \text{ where } b_m^s \text{ denotes the weight owed to } d_m^s.$$

Apparently, the more two blocks are alike; the bigger weight would be assigned. Thus, we can define the expression for weight calculation as $b_m^s = \exp(-e_m^s / h) / \sum \exp(-e_m^s / h)$ where h stays a constant directing the window shape. The global model in an image can be estimated such as

$$\sum_{m=1}^N \left\| d_m - \sum_{s=1}^L b_m^s d_m^s \right\|_2^2 \quad (11)$$

Mathematically in form of $\|(I - B)\phi\alpha\|_2^2$. Besides incorporate equally the AR regularization and the NLSS based sparse depiction in Eq. (12), thus solves the problem using AR to regularize image confined smoothness with NLSS towards utilizing the image Non-Local redundancies as given as

$$\hat{\alpha} = \left\{ \|b - H\phi\alpha\|_2^2 + \sum_{m=1}^N \sum_{n=1}^o \lambda_{m,n} |\alpha_{m,n}| + \gamma \|(I - A1)\phi\alpha\|_2^2 + \eta \|(I - A2)\phi\alpha\|_2^2 \right\} \quad (12)$$

Meanwhile η panels the stability among local adaptation plus nonlocal strength.

The projected technique advances the existing schemes into excellence metrics despite undersea picture is crucial to recover color. In next section, we therefore propose color adjustment approach, relying on weighted grey model. As depicted within Figure 2, our approach adopts a novel strategy, for compensating color cast so as to enhance detail scene.

4. Underwater image enhancement of deblurred image

The deblurred image is subject to global enhancement with no prior required. This is necessary to lessen the architects due to deblurring, and for enhancement of color information otherwise degraded in the underwater image formation process. Different depth levels allow different colors to distort. Thus the faded as

well as non-uniform color distribution will describe the undersea representation. Mauricio [28, 29] proposed an algorithm to readily pre-process the underwater imaging using homomorphic filtering, wavelet de-noising anisotropic filtering, and RGB color channel equalization to enhance color. The technique is usual and need no prior or constraint change. Another such scheme uses distribution of the histogram to make the bins change according to color density. In the proposed work, we use a weighted grey edge scheme to improve the color constancy.

4.1 Color adjustment section

Color adjustment is crucial in undersea, we apply our system towards deblurred imagery thus improving the picture look by means of neglecting redundant color casts caused by diverse radiance. To obtain our output we achieve a gamma correction method which endeavours to correct global dissimilarity; hence progress color constancy via weighted gray edge. It dispenses dissimilar mass to diverse boundaries according to the information of boundaries called weighted grey edge [24]:

$$\left(\int |w(f_i)^K f_{i,D}(D)|^p dD \right)^{\frac{1}{p}} = Ki_f \quad (13)$$

Where $w(f_i)$ denotes mass purpose, K remains constant to impose the mass of the boundaries, $f(D)$ stays on input, i_f denotes final-illuminant to make proper color. This modifies colors as close as original colors, considerably improving ending outcome. Numerous scheme incorporates towards accurately estimating color, an appropriate weighting system must impose appropriate data about color and ignore unrelated data. Given that cue intended for estimating color, an obvious choice towards computing weight using specular edge detection methods.

4.2 Different boundary methods

A variety of boundary type as material, shadow or shading, specular and inter-reflection boundaries. i) Material boundaries are conversion among two dissimilar surface or else items. ii) Shading boundaries are conversion that reasons the geometry of an item, intended for an alteration in surface direction by means of light. It is also cause obstruct in the beam source. Obstruct intensity gradient, however occasionally introduce faint color gradient. iii) Inter-reflection is the cause of radiance show as of one surface against a second surface. Thus, further transform of entire lighting towards inmost near next surface, therefore color, perceived via subtracting the derivative of invariant as well as true image. The projected scheme includes various photometric descriptions. Further exclusively, quasi invariants use many flexible weighting systems, resultant within a well-designed weighting system. Quasi invariants calculate the derivative of a representation, with three photometric variants

$$f_D = (f_{R,D}, f_{G,D}, f_{B,D})^T \quad (14)$$

Besides eliminating inconsistency, a set of derivatives making called quasi invariants. Using quasi invariants portray, three dissimilar weighting schemes.

4.2.1 Specular boundary system

The quasi invariants decay picture into three paths. Ridge of derived term resting on luminosity basis called specular variant which definite as

$$O_D = \left(f_D \cdot \hat{C} \right) \hat{C} \quad (15)$$

where $\hat{C} = \frac{1}{\sqrt{3}}(1,1,1)^T$ denotes color source and dot signify vector product. The derivative reasons with places of interest and subtracted with variation as of its derivative termed as

$$O'_D = f_D - o_D \quad (16)$$

It merely enclosed shadow shade with material boundaries and is not sensitive towards emphasizing boundaries. While the entire derivative power enclosed in three various paths, relation among specular variation vs whole sum of energy be a suggestion that a boundary is specular or else not. The specular weighting system as:

$$W_s(f_D) = \frac{|o_D|}{\|f_D\|} \quad (17)$$

Where $|o_D|$ describes complete rate with o_D in addition to $\|f_D\| = \sqrt{f_{R,D}^2 + f_{G,D}^2 + f_{B,D}^2}$.

4.2.2 Shadow boundary system

By means of similar analysis happening, the shadow-shading route an invariant in addition to quasi invariant attained as:

$$S_D = \left(f_D \cdot \hat{e} \right) \hat{e} \quad (18)$$

$$S'_D = f_D - S_D$$

Where $\hat{e} = \frac{1}{\sqrt{R^2 + G^2 + B^2}}(R, G, B)^T$. This quasi-invariant is not sensitive towards shadow boundaries. Translating this yields the next effect:

$$W_{SD}(e) = \frac{|S_D|}{\|f_D\|} \quad (19)$$

4.2.3 Material boundary system

At last, shadow system along with quasi invariant creates extrapolate derivative taking place hue path:

$$H_D = \left(f_D \cdot \hat{b} \right) \hat{b} \quad (20)$$

$$H'_D = f_D - H_D$$

where denotes hue path, perpendicular on the way to preceding paths:

$$\hat{b} = \frac{\hat{e} \times \hat{C}}{\|e \times C\|} \quad (21)$$

where H'_D do not enfold specular or else shadow boundaries, which then employ better weights to objects boundaries resembling

$$W_M(f_D) = \frac{|H'_D|}{\|f_D\|} \quad (22)$$

To assess the control of boundary -kind classifier on the color constancy outcome. Out of these three methods specular edge type used to estimate final illuminant.

4.3 Iterative Weighted Gray Edge

The projected scheme initially correct deblurred imagery by an approximated illuminant. Then, last color corrected output image is given as

$$f_f = f_D \left(\frac{1}{i_f} \right)^D \quad (23)$$

Subsequently, in suggested method, a novel weighting scheme is to compute color corrected imagery for every iteration. For clearness, we will not change the weighting scheme throughout the iterations. Additionally, the early light approximate white basis $(1,1,1)^T$ otherwise end the color adjusted result. At last, convergence defined for predestined several iterations.

5. Experiment Results

Experiments results focussed towards removing blur and enhancing color. The improved underneath picture illustrates rigorous color alteration; however not succeed to improve colors completely. Deblurring method in Figure.3 gives deblurred picture; furthermore Figure 4 yield enhanced result to improve the color. Therefore, the projected scheme yields improved performance than the conventional system via four major branches: (A) Underwater Deblurring Evaluation (B) Underwater Enhancement Evaluation, and (C) Quantitative Evaluation. These parts depict individual performances of each module.

5.1 Underwater Deblurring Evaluation

It is an effective scheme for reconstructing original representation. Estimated PSF is added with original to make blurry which then added with Gaussian noise to get a blurry representation, which subjected through iterative deblurring schemes to recover novel imagery. Our algorithm in Figure 3 applied to only luminance element because human vision is more responsive to luminance variation.

5.2 Underwater Enhancement Evaluation

The undersea turbidity removal imagery improves the eminence although it produces severe color deformation owing towards absorbing band with underneath element. Thus inaccurate statement is based on every hue band is evenly absorbed in underneath. Thus projected technique in Figure 4 employ weighted gray edge process towards make even every RGB means and thus resolve distortion problem captured in underneath. Assessment among preceding technique exposed in Figure 5. While our method can proficiently get away blur with color deformation to get proper color exclusive by artifacts and gamma adjustment to correct global contrast moreover undersea imagery tend to emerge too light. The connection among true, deblurred as well as color output illustrated in Figure 6.

5.3 Quantitative Evaluation

Color adjustment using final illuminant gives final output, used by computer vision systems in many applications. Then projected algorithms give improved results as evaluated to existing scheme [30]. Table summarize metrics using Image quality Assessment and unique Image quality Assessment (SIQA). Thus projected scheme be superior than conventional scheme gives improved result.

a) Ordinary IQA

It mainly assesses contrast and structural adjustment. The most commonly used IQAs introduced as follows: *Mean*: Mean is the sum of all values in matrix.

$$\mu = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{m,n} \quad (24)$$

b) *Standard deviation (SD)*: It reflects the degree of dispersion within picture on its standard significance, and contrast in certain range. The larger SD, better the visual effect will be:

$$SD = \sum_{m=1}^M \sum_{n=1}^N \sqrt{\frac{(f(m,n) - \mu)^2}{MN}} \quad (25)$$

where M with N denotes row, column of imagery; $f(i, j)$ denoted intensity of pel moreover μ belongs to standard rate of whole image.

c) *Entropy (E)*: An image taken as a source of random output sets $\{a_i\}$ and the probability with a_i is $P(a_i)$ then standard amount by data in image as shown as

$$H = -\sum_{i=1}^L P(a_i) \log_2 P(a_i) \quad (26)$$

Higher value of E, more the information in picture.

d) *Mean squared error (MSE)*: The full-reference excellence metric computed via average squared intensity differences of distortion along with reference representation pels as

$$MSE = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N [f(m,n) - f'(m,n)] \quad (27)$$

where M and N denotes row and column of the imagery, $f(m, n)$ remains original with $f'(m, n)$ denotes deblurred picture.

e) *Peak SNR (PSNR)*: It is used as index for signal distortion. Larger the PSNR smaller the distortion and expressed as:

$$PSNR = 10 \log \frac{f_{\max}^2}{MSE} \quad (28)$$

where f_{\max} stays on largest gray value, in general $f_{\max} = 255$.

f) *Structural similarity (SSIM)*: Generally, the ocular view is extremely modified to extract data on picture, thus measures restored picture excellence using three mechanism specifically; luminance $L(D, E)$, contrast $c(D, E)$ along with structure comparison $S(D, E)$. All these combined to give up whole similarity computation as:

$$S(m, n) = F(L(D, E), c(D, E), S(D, E)) \quad (29)$$

The similarity of the two images is dependent on SSIM, and has a value between [0,1]. When the value is close to 1, the two images are more similar. SD reflects dissimilarity of the image; E reflects the data; later MSE, PSNR and SSIM reflects the degree of distortion. Higher MSE, lower PSNR and SSIM scores imply greater dissimilarity among enhanced results and referenced deblurred image. This measures often used for simple calculation enclose clear and suitable for optimization. Conversely, these approaches cannot be simply adopted, because preceding IQA metrics are usually unsuitable for relevance intend to measure distortion intensity rather than visibility in imagery.

5.3.2 Special IQA

Some IQAs designed particularly for image from different views as follows.

a) Image Visibility Measurement (IVM)

Inspired with blind assessment indicator, an extra picture visibility measurement employ perceptible boundary segmentation as

$$IVM = \frac{n_v}{n_{total}} \log \sum_{D \in \mathfrak{S}} C(D) \quad (30)$$

where n_v represents the amount of perceptible boundaries, n_{total} denotes amount of boundaries, $C(D)$ remains average dissimilarity, along with \mathfrak{S} denotes picture region with perceptible boundary.

b) Histogram correlation coefficient (HCC)

A good deblurring technique should allow improved picture towards Histogram allocation. It employs dual color imagery as standard to measure the act of color enhancement.

c) Contrast gain

The dissimilarity of clear picture is much higher than that of deblurred imagery, that check dissimilar enhancing system. Higher contrast, better the scheme to be. Global dissimilarity is for evaluating different techniques. It signifies mean contrast comparison among enhanced as well as deblurred image by

$$C_{gain} = \bar{C}_J - \bar{C}_I \quad (31)$$

where \bar{C}_J in addition to \bar{C}_I represents mean contrast of enhanced and deblurred representation, respectively.

d) Ocular dissimilarity assess (ODS)

It compute the amount of perception using

$$ODS = 100 * R_v / R_l \quad (32)$$

where R_v represents amount with limited region, SD is higher than known boundary with R_l remains entire amount of limited region. We chose OTSU image segmentation algorithm that adaptively compute the limit. It uses local standard deviation that denotes dissimilarity picture towards determining visibility.



(i) (ii) (iii) (iv)
 Figure 3: Underwater Turbidity Removal. (i)Blurred picture (ii) ASDS (iii) ASDS using AR (iv) ASDS using AR and NL



a) Deblurred b) grad_im c) weight_map d) Ours
 Figure 4: Proposed Method



a) GW b) SOG c) WP d) GE e) Ours
 Figure 5: Assessment among preceding technique



a) Blurred Image b) Deblurred c) Ours
 Figure 6: Input and output imagery

Table 1: Estimation of eminence metrics of Image quality Assessment (IQA) and Special Image quality Assessment (SIQA) of Proposed Method

	Methods	M	SD	E	PSNR	RMSE	SSIM	ODS	HCC	CG	UQI
ID	ASDS	127.3	76.98	7.911	25.31	19.916	0.694	67.867	0.444	0.465	0.948
	ASDSAR	127.31	76.834	7.919	25.73	19.117	0.716	64.82	0.475	0.422	0.962
	ASDSARNL	127.34	76.66	7.933	26.16	17.816	0.742	58.449	0.531	0.334	0.940
CA	GW	129.6	72.443	7.798	65.038	0.142	0.9983	62.604	0.070	0.054	0.976
	SOG	128.73	72.057	7.794	65.745	0.132	0.9986	62.05	0.096	0.051	0.976
	WP	127.15	76.134	7.923	69.992	0.081	0.9999	58.449	0.477	0.045	0.981
	GE	127.55	76.954	7.931	69.993	0.081	0.9999	58.449	0.978	0.044	0.981
	Ours	127.42	76.812	7.910	70.002	0.081	0.9999	58.449	0.962	0.044	0.981

Generally, higher ODS, clearer the improved picture.

e) Universal quality index (UQI)

It mainly assess performance of UQI among original as well as the improved imagery given as

$$Q = \frac{4\sigma_{de} \bar{d} \bar{e}}{(\sigma_d^2 + \sigma_e^2) \left[\left(\frac{\bar{d}}{\sigma_d} \right)^2 + \left(\frac{\bar{e}}{\sigma_e} \right)^2 \right]} \quad (33)$$

Where

$$\bar{d} = \frac{1}{N} \sum_{m=1}^N d_m, \quad \bar{e} = \frac{1}{N} \sum_{m=1}^N e_m$$

$$\sigma_d^2 = \frac{1}{N-1} \sum_{m=1}^N (d_i - \bar{d})^2, \quad \sigma_e^2 = \frac{1}{N-1} \sum_{m=1}^N (e_m - \bar{e})^2$$

$$\sigma_{de} = \frac{1}{N-1} \sum_{m=1}^N (d_m - \bar{d})(e_m - \bar{e})$$

Traditional UQI criterion both uses a picture with high excellence as the reference image. Thus, higher UQI, better the compared image. However, the deblurred imagery is always chosen as of reference imagery, so large UQI do not means that the imagery is of high excellence. The improved representation with best visibility may have smallest UQI.

6. Conclusion

We projected an effective technique namely an underwater turbidity removal scheme with adaptive sparse domain selection (ASDS) methodology significantly improves undersea imagery, and consequently, results of image deblurring. If the deblurred imagery has any lighting problem, then; we used color corrected method using Weighted Gray Edge to find the clear image.

The experimental results shown that the projected means illustrate towards removing turbidity outperforms several state-of-the-art restoration scheme in excellence metrics along with ocular excellence. Finally, we use a color correction scheme to clear the lighting (color) in state-of-the-art images.

Acknowledgment

The authors thank the reviewers for their thorough and helpful remarks.

References

[1]. M. Kocak, Fracer Dal, M, F Cai and Sche, "A focus on recent development and trends in Underwater Imagng," *Mare Tech. Socie. Jorn*, Mar-2008.
 [2]. Gian. Lu. F, "Visual inspection of Sea Bottom Structures by an Autonomous Underwater Vehicle," *IEEE Transac System., Man, Cybern.* Oct-2001.

[3]. Alberto O, Miquel S, and Gab.O, "A Vision System for an Underwater Cable Tracker," *Machn. Vision. Appln.*, 2002.
 [4]. Adria. O and Eman. T, "Detecting Man-Made Objects in Unconstrained Subsea Videos," *Prod. BMV-2002*,
 [5]. BA. Level and L. Berg, "Control of Underwater Vehicles in full unsteady flow," *IEEE Jrn. Oceac. Engg-2009*.
 [6]. CH. Maze, "In Situ Measurement of Reflectance and Fluorescence spectra to support Hyperspectral Remote Sensing and Marine Biology Research," *IEEE oceanc-2006*.
 [7]. Yaac. K and Jeffrey R, "Analysis of Hull remains of the Dor D vessel," *Intn. Jrn. of Nautial Archeo- 2001*.
 [8]. BL. McGla, "A Computer model for Underwater Camera System," *Proc of spie1979*.
 [9]. J. Jaf, "Computer Modeling and Design of Optimal Underwater Imaging Systems," *IEEE Journ of Oceanic Engi -1990*.
 [10]. J. Fun, B. Brya, and J. Heck, "Handbook of Underwater Imaging System Design," TP303 - 1972.
 [11]. Timo H, Thom P, F. Chap, and Rob. CT, "A Range-Gated Laser System for Ocean Floor Imaging," *Marin Tech So Jo-1983*.
 [12]. John W and Kenn. V, "Point spread functions in Ocean Water: Comparison between Theory and Experiment," *Appd Opti-1991*.
 [13]. Tor M and Gre Du, "A statistical learning based method for color correction of Underwater Image", *Advn artifi intelli theory-2005*
 [14]. Chike and Fu"Automatic white balance for digital still camera," *Inf.orm Scien Jrn/- 2006*
 [15]. Weng, Ho C, and Fu, "A Novel automatic white balance method for digital still Cameras," *IEEE_ Inter .Sym Ckts & Sys- 2015*.
 [16]. EY La, "Combining grey world & Retinex theory for automatic White Balance in digital photography," *Inter Sym Consum Elec, 2005*.
 [17]. Xial. W, Xiang. Z and Jia.W, "Model-Guided Adaptive Recovery of Compressive Sensing," *Proc Data Compre Confe, - 2009*.
 [18]. Juli. M, Franc. B, Jean. P, Guil. S and Andre. Zn, "Non-Local sparse models for Image restoration," *Intern Conf r on Compt Visn-2009*.
 [19]. Stef. K, Stanley. O, and Peter. W. J, "Deblurring and Denoising of images by Non-Local Functional," *Multisel Modl and Simun- 2005*.
 [20]. Mattan. P, Michael. E, Hiro. T, and Paymn. M, "Generalizing the Nonlocal-Means to Superresolution Reconstruction," *IEEE Transac. Img Pros- 2009*.
 [21]. Riz, Gat and Mar "Color Correction between Gray World and White Patch," *Humn vis & Elecr,2002*
 [22]. Finly and Trez "Shades of grey and colour constancy," *Clr & Imag Confr, 2004*.
 [23]. J Van, T Geve "Edge based colour constancy," *proc.IEEE Trans on Img Proces, 2010*.
 [24]. Gijs, T Gevrs, and Van, "Improving Color constancy by photometric edge weighting," *proc. IEEE_ Trans. on Patrn Analy & Machine Intelli-2012*.
 [25]. A.Chrispin Jiji, Vivek M "Underwater Turbidity Removal through Ill-posed optimization of sparse modeling," *ICPCSI (2017)*.
 [26]. Weish D, Lei Zhg, Gua Sh, Xiaon Wu, "Image Deblurring and Super resolution by Adaptive Sparse Domain Selection and Adaptive Regularization", *IEEE Transc on Imge Proc- 2011*.
 [27]. We Dong, Le Zhang, Guan Shi, Xiao W, "Image Reconstruction with Locally Adaptive Sparsity and Nonlocal Robust Regularization", *Sigl Proces: Img Commn -2011*.
 [28]. Mauo Delb, Pu Mu , Andr Alman "Non -Parametric Sub-Pixel local point spread function estimation," *Img Proces-2012*.
 [29]. Yoav. S and Nir. K, "Recovery of Underwater Visibility and Structure by Polarization Analysis," *IEEE Jol of Ocea Engg, 2005*.
 [30]. Y Xu, , J Wen, L Fei1, and Zheng Z, "Review of Video and Image Defogging algorithms and related studies on Image Restoration and Enhancement", *IEEE open acces Jom-2016*.

Analysis and Methods on The Framework and Security Issues for Connected Vehicle Cloud

Lin Dong^{1,2}, Akira Rinoshika^{2*}

¹Mechanical and Automotive Engineering School, Shanghai University of Engineering Science, 201620, China

²Department of Mechanical Systems Engineering, Yamagata University, 992-8510, Japan

ARTICLE INFO

Article history:

Received: 07 June, 2018

Accepted: 09 July, 2018

Online: 14 November, 2018

Keywords:

Internet of Things (IoT)

Connected Vehicle Cloud (CVC)

Security of Vehicle Cloud

ABSTRACT

In the world today, the rapid development of the Internet of Things (IoT) and the application of the Connected Vehicle Cloud (CVC) as the Internet of Things in the intelligent transportation are becoming widespread. They can improve people's safety, vehicle security as well as reduce the cost of ownership of an automobile. At the same time the security of the Internet is a non-negligible factor in the development of the Internet of Vehicles. Therefore, the security of vehicle networking is of great concern. This article starts with the network architecture of vehicle networking and combines the examples of vehicle networking security issues, which analyzes and researches the security problems of vehicle networking, and proposes solutions to the security problems faced.

1. Introduction

Connected Vehicle Cloud (CVC) increases the core business value of automotive OEMs by providing a platform for creating, managing, and deploying connected vehicle services as shown in Figure 1. It creates a direct channel to the driver and gives the possibility to introduce new partners to participate in the value network of the automotive industry. CVC enables OEMs to engage with different players in the automotive eco-system to deliver services while remaining in control and keeping the costs of deploying and managing the services to a minimum, as shown in fig 1.

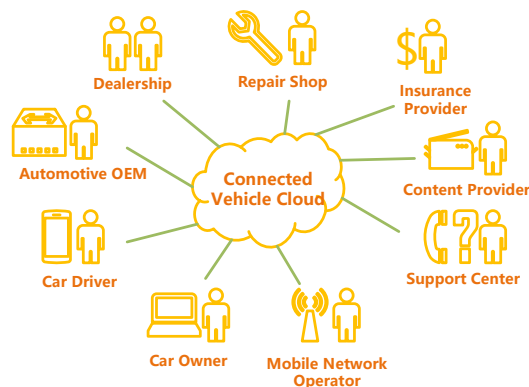


Figure 1. Connected Vehicle Cloud powering the Automotive Ecosystem.

*Akira Rinoshika, Email: rinosika@yz.yamagata-u.ac.jp

CVC is a common service delivery platform for infotainment, telematics, and other services related to connected vehicles. It is completely independent of connectivity solutions and can be deployed without any integration with a mobile network. CVC is based on the Service Enablement Platform (SEP). SEP combines functional components from Multiservice Delivery Platform (MSDP), General Composition Engine (GCE), M2M Data Management (M2M DM) and Dispatcher.

Cloud Computing is the basis of CVC [1]. Cloud Computing is the internet based new computing system which is distributing services for the interest of clients such as shared network resources, software, and platform computing infrastructure [2]. Therefore, the CVC is an open platform that supports flexible deployment and realization of services. The flexibility of the Service Enablement Platform (SEP) allows the CVC to be continuously adapted to changing business and technical requirements. SEP has functionalities needed to support creation and deployment of connected vehicle services. Each connected vehicle service is developed according to the customer needs with the support of the SEP functionality and sometimes with the support of additional third-party components (3PPs). Through these components, vehicles can access the cloud and obtain, at the right time and the right place, all the needed resources and applications that they need or want [3].

Connected Vehicle Cloud provides functionality to connect vehicles and other devices to the cloud. Vehicles are securely

connected through a bidirectional communication link supporting multiple protocols and notification mechanisms such as SMS Shoulder Tap, HTTP, and MQTT. Status information and data are collected and sent to CVC, where it is normalized, stored, aggregated and combined with data from other systems and sensors.

This data (in whole or in part) is distributed to CVC applications and participants who gained the relevant access rights. With granular access control, you can publish only the precise information necessary to subscribe to services such as analysis systems. Events are defined for notification when a particular set of conditions is applied.

Firmware update function allows the OEM to wirelessly update software and firmware of onboard units in the vehicle. CVC acts as a cache and provides software updates for many vehicles. In addition, because business rules and scheduling functions are also provided, OEM can control which software file is provided to which vehicle when and when. Many industry standard protocols for software and firmware updates are provided by the standard, and additional protocols are being added using the open SDK.

This paper firstly presents the main actors of the Connected Vehicle Cloud solution. Then the analysis of security problems are caused by CVC. Finally a solution of CVC security issues is provided.

2. Framework of connected vehicle cloud

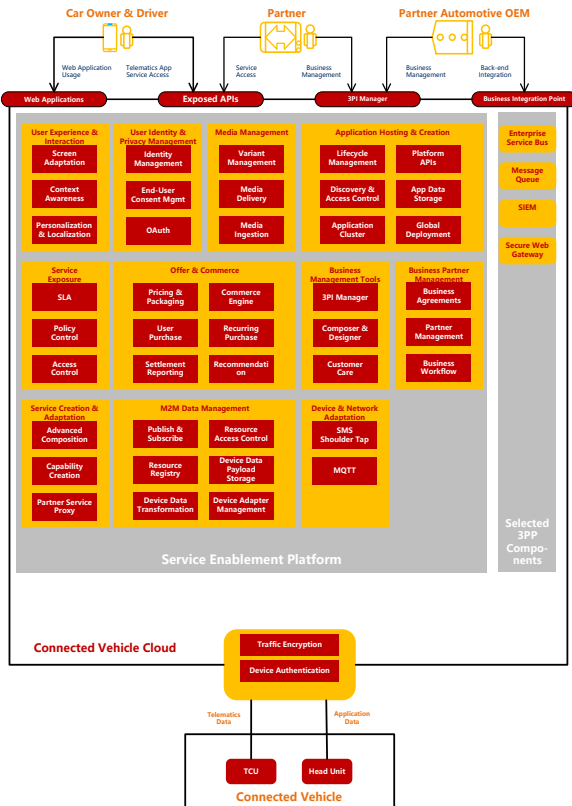


Figure 2. Connected Vehicle Cloud Functional Overview.

CVC is a comprehensive platform for creating, deploying, and managing all types of services related to connected vehicles.

CVC is an open platform that supports flexible deployment and realization of services. The flexibility of the Service Enablement Platform (SEP) allows the CVC to be continuously adapted to changing business and technical requirements. SEP has functionalities needed to support creation and deployment of connected vehicle services. Each connected vehicle service is developed according to the customer needs with the support of the SEP functionality and sometimes with the support of additional third-party components (3PPs).

An overview of the main functional areas of SEP that are used in CVC as well as some of 3PP components that typically make the foundation of a customer solution is shown in the above fig 2.. The fig 2. also shows some examples of how the main actors in the automotive ecosystem interact with the CVC.

2.1. The main actors of the Connected Vehicle Cloud framework

- **Car Driver and Owner:**
The primary end-user of CVC is a person driving a car connected to the cloud. The driver or owner accesses the services of CVC through one or more devices types connected to the cloud. For example, a built-in head unit in a car, a smartphone application, or a web portal offering services related to the vehicle. CVC separates the Driver and Owners from the Connected Vehicle; they are the end users of most of the services deployed in CVC, and they may be using services related to one or more vehicles. In some cases, an end user may be someone who is not the driver or owner of the vehicle that he is accessing services from.

- **Connected Vehicle**
A vehicle connects to CVC through one or more mobile network connections. CVC Services are delivered to one or more of the vehicle’s built-in or aftermarket devices such as an infotainment head unit or a telematics control unit (TCU). The vehicles connect to CVC using one or more components in the vehicle that is capable of sending and receiving data to and from the cloud, execute service logic, display information to the drivers and passengers, collect data from the vehicle, and interact with other micro-controllers and software components in the vehicle. The capabilities of the components in the Connected Vehicle can vary greatly depending on the vehicle platform, and the functionality can be separated into a variety of different components. In this document, we refer to these type of components as being either a component that is primarily focused on telematics services, a Telematics Control Unit (TCU), or a component responsible for infotainment services, a Head Unit.

- **Automotive OEM**
An enterprise that manufactures the vehicles connected to the CVC is the Automotive OEM (OEM). The OEM uses the services delivered through CVC to build and improve the customer relationship with car owners, capture aftermarket sales, collect vehicle information to improve quality control, and earn additional revenues from partners accessing or providing services through the platform. The OEM will typically integrate CVC with existing business support systems.

1) Partners

CVC enables any of the players in the automotive eco-system to become a partner of the Automotive OEM. The partner will provide services through CVC or access services provided by the platform.

- a) An Insurance Company that gains access to the platform to enable a Pay-as-You-Drive program.
- b) A Telematics Service Provider the uses the platform communication channels to deliver Tele Guard services to assist drivers in the case of a breakdown.
- c) A Live Traffic Information Provider that provides real-time traffic information content to multiple Intelligent Navigation services deployed on the platform.
- d) A Fleet Management Company that buys access to the platform APIs to develop its own Fleet Management Applications.

2) Service Provider

CVC service provider is responsible for operating CVC and delivering services to the connected vehicles. The service provider can be an Automotive OEM, a mobile operator, or any other party willing to take this role.

2.2. Component architecture of the Connected Vehicle Cloud framework

The CVC solution is comprised of MSDP, GCE, M2M DM, and the Dispatcher components which provide the different functionalities of the solution. This chapter gives a high-level overview of the functionalities of the different components and how they interact with each other.

This chapter describes how the different components in CVC are used together to realize and expose the functionality of collecting data and sending message to in-vehicle devices. This functionality provides a foundation for implementing any type of telematics service. Depending on the business needs and customer requirements the solution can be implemented with or without using the Service Exposure and User Identity Privacy Management functionality.

The following is a short description of how the components interact with each other.

MSDP: The MSDP realizes the business management related functionality of CVC. The 3PI Manager tool is used to configure business agreements which define the SLAs and terms of the services that are made available to the business partners of the OEM through the Service Exposure Functionality. Once the agreements have been signed and activated in MSDP the information is provisioned to GCE. This information contains information about throttling and quota rules for exposed services registered OAuth applications, and M2M data resource information. MSDP is also used to host services (for example telematics services) that directly access M2M DM to request M2M data reported from vehicles or to send M2M messages to vehicles, Services that make use of the M2M Data Management functionality may be hosted web applications making use of the Application Hosting functionality or may be services directly made available to external end-user services through the User Experience & Interaction functionality.

GCE: The Service Exposure functionality of GCE ensures that the SLAs and terms of services provisioned from MSDP are enforced on all exposed services. It also ensures that end-user consents are obtained before any M2M data can be accessed by a partner application by implementing the OAuth request flow. GCE exposes all the functionalities of the M2M Data Management component through the M2M Data Exposure and M2M Messaging Exposure services. Additionally GCE provisions information about created end-user consents as well as new OAuth applications to M2M DM in order for M2M DM to enforce the m2m data access control based on this information

M2M DM: The M2M DM component realizes the M2M Data Management functionality to store, transform, and provide access control for all data reported by vehicles as well as messages being sent from applications to vehicles. M2M DM provides interfaces for applications and services to access the data and messaging services either directly or through the services exposed through the GCE and the Service Exposure functionality M2M Messages that should be forwarded to vehicles are transformed into the correct format by M2M DM and forwarded to the Dispatcher component to be delivered to the vehicle.

Dispatcher: The Dispatcher provides a common interface for M2M devices (for example Telematics Control Units) to communicate with the different services enabled by the CVC using different communication protocols and message delivery mechanisms. The Dispatcher forwards data and messages to and from the M2M DM component from and to devices using either the MQTT protocol or an SMS Shoulder Tap mechanism.

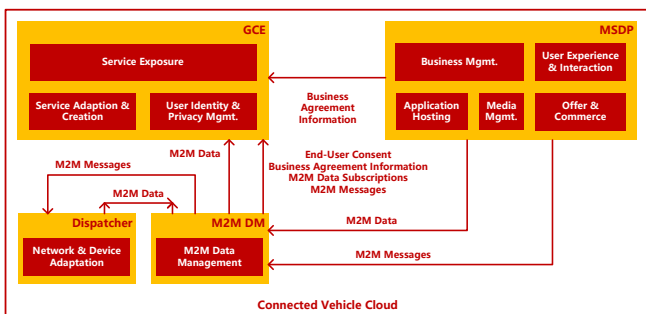


Figure 3. Overview of Components of CVC and How They Interact with Each Other.

Fig 3. provides an overview of the components, the functionalities they provide, and the primary information flows between the different components. Chapter III gives a detailed description of the different functionalities of CVC. For detailed instructions on how to install and configure MSDP, GCE, M2M DM, and the Dispatcher components, refer to Installation and Initial Configuration Guide for CVC 1.3 [4].

2.3. Case study

One of the more common applications for connected vehicles are In-Car Navigation applications that help the driver navigate to the right destination and provides information about points of interests (POIs) along the planned route.

In this example, as shown in fig 4.,the In-Car Navigation application has been developed as a native application for the head unit operating system. The application accesses one of the API

services that has been developed and deployed on the platform to send and retrieve data.

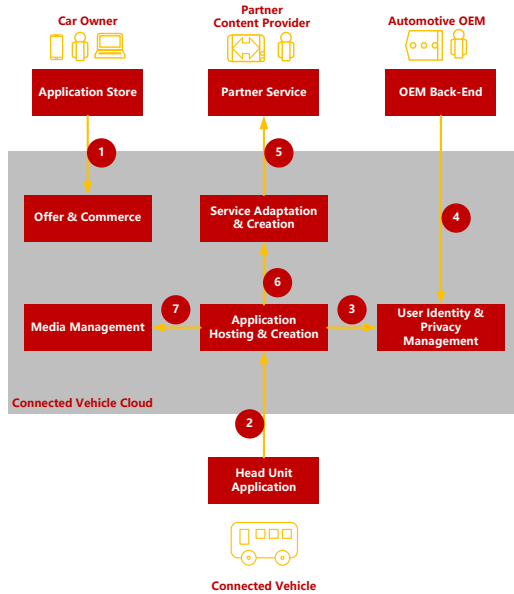


Figure 4. Overview of the CVC Functional Areas involved in Enabling an in-car Navigation Application Running in the Vehicle Head Unit.

This example describes how the CVC functionality is used to enable the In-Car Navigation application.

The following steps describe the main interactions between systems and functional components:

- The car owner uses the CVC Application Store to discover and purchase the In-Car Navigation Application. After the application is installed in the head unit of the car, it is available for the driver to use.
- When car driver uses the In-Car Navigation application, the application connects to the application back-end to download additional content. The application back-end is hosted in the CVC application hosting environment and exposes application features through REST APIs that are accessed by the application client in the head unit.
- The back-end application uses the CVC User Profile APIs to ensure that the user is authorized to use the application and retrieve the user profile.
- The automotive OEM integrates with CVC to provide user profile information from a central CRM database as well as retrieve information about user application usage from CVC.
- The automotive OEM integrates with Partner Service that provides additional content to navigation applications. To integrate external services with the hosted back-end applications the Advanced Composition can be used to create new APIs that access and retrieve content provided by partners.
- The back-end application uses the integration with a navigation information Content Provider Partner service to retrieve information about relevant POIs to display to the driver.
- The hosted back-end applications accesses the CVC Media Management functionality to retrieve the correct variants of image and video content to display in the application based on

the information received from the partner service and head unit device capabilities.

3. Analysis of Security Problems Caused by CVC

While car networking brings convenience to people, safety issues also follow. Like other information systems, the security threats of the Vehicular Network also include denial of service, information leakage, tampering, replay attacks, counterfeit identities, and denial of operations. These attacks cannot be mitigated by means of traditional security techniques, as in the case of network attacks [5]. To achieve a higher level of security for sensitive messages, one can apply active security mechanisms [6] at the cost of losing a certain amount of efficiency.

3.1. Case of vulnerability in remote Control key system

Dutch electronics industry designer Tom Wimmenhove found a serious safety design flaw in the key systems of various Subaru cars, as shown in fig 5.. The manufacturer has not yet fixed the loophole and the loophole will lead to the hijacking of Subaru cars.



Figure 5. Case of vulnerability in remote Control key system.

By receiving a data packet sent by the key system (for example, the attacker only needs to capture the packet within the signal range after pressing any key of the key system), the attacker will be able to use the data packet. To guess the rolling code generated by the vehicle key system for the next time, he can then use this prediction code or direct replay to lock and unlock the vehicle. The use of this vulnerability is not difficult, attacking devices can be made using off-the-shelf electronic components, and do not require attackers with high-end programming skills. There are many hardware hackers in underground cybercriminals and things that can be done by designers in the electronics industry. These people can easily do the same. The car thief only needs to make a simple device that collects radio signals from the car key system, calculates the next scrolling code, and then sends a similar radio signal back to the target car after the target Subaru car owner leaves, and they can hijack the car.

3.2. An attack case of Tesla vehicle networking system

Tesla has built a WIFI Tesla Service into every Tesla car. Its password is a clear text stored in QtCarNetManager and will not be automatically connected in normal mode, as is shown in fig 6..

Tesla-Guest is a WIFI hot spot provided by Tesla 4S store and charging station. This information is stored in Tesla for automatic connection in the future. Researchers can create a fishing hotspot, Tesla, where users can redirect QtCarBrowser traffic to their domain names when they use CID to search for charging piles, which can be used for remote attacks.

In addition to WIFI technology, in cellular networks, if an attacker builds enough websites, phishing techniques or user errors can also be used to achieve the purpose of the intrusion. Because it is a browser-based attack, it can be done remotely without physical contact.



Figure 6. An attack case of Tesla vehicle networking system.

3.3. Solutions to security problems

- SIEM

A Security Information and Event Management (SIEM) system is a system that enables real-time analysis of security-related information and event logs. It also provides automation of security-related tasks, and production of alarms and reports. SIEM system can be used to collect log and status information from many different subcomponents from the system, and it is based on a set of pre-defined rules take action when a security risk has been detected.

It can be used to detect vehicles that quickly connect from different geographical locations that are far away from each other. This situation can indicate that a SIM card or vehicle identity has potentially been compromised, and the SIEM system can then instruct the Cloud Entry Point or other authentication systems to temporarily block access for this vehicle and raise an alarm for system administrators to investigate.

- Boundary Defenses & Secure Gateway

On top of boundary defenses like encryption, CVC comes with an additional level of security based on a Security Incident and Event Monitoring (SIEM) system that is used to detect suspicious communication patterns and anomalies, sometimes also referred to as an Anomaly Detection system, as shown in fig 7..

Anomaly detection systems help to protect against malicious attempts to hack the vehicles. Events outside of the normal pattern will be detected and the reputation level of the device will change. CVC controls the policy for how a device is using services. For example, if the Anomaly detection system lowers reputation the CVC may enforce read access only.

Additionally the CVC includes a Secure Gateway. An automotive OEM may want to let the Car Drivers or Vehicle Passengers browse the web using the connectivity of the CVC and the in-vehicle infotainment unit or let a user connect using a mobile device that connects through the vehicle access point. To protect both the end-users and the CVC system from harmful or malicious content or software the CVC can be integrated with Secure Web Gateway 3PP that filters the web traffic of the user and ensures that no unwanted content or malicious web sites are accessed.



Figure 7. Certificates handling.

- Certificate Validation

To ensure that all communication with the Connected Vehicle Cloud is secure, the automotive OEM will often have an existing Public Key Infrastructure (PKI). Public Key Infrastructure (PKI) and digital signature-based methods have been well explored in VANETs [7]. It requires that all vehicles authenticate with a pre-provisioned certificate for all communication with the cloud. This is the process of the mutual authentication. Mutual authentication ensures that both device and cloud (i.e. server side) can verify authenticity of each other. Access can be revoked or suspended through OCSP standards, this puts the OEM in control over which devices have access. All of this can ensure that all vehicles are securely authenticated when accessing the services of the cloud and that all communication is secure and encrypted [8].

When the OEM uses a PKI, the CVC needs to implement the necessary components to validate the certificates of all vehicles and secure the communication between vehicle and cloud. This is typically achieved using Transport Layer Security (TLS) protocol to secure all communication and a CVC communication end-point (for example, a load balancer) that is capable of validating the vehicles' certificates and terminate the TLS traffic.

4. Conclusion

This paper exhibits that the Internet, Internet of Things and car networking become another major symbol in the future smart city. For improving the people's lives and increasing the convenience of travel, the car networking brings many security threats and seriously affects personal and information security. It is necessary to grasp the current development trend of the Internet of Everything, and while studying the development trend and core

technologies of the car networking. We must pay attention to the construction of safety protection under the environment of car networking and ensure the healthy and orderly development of car networking in the future.

Acknowledgment

This work is partially supported by the visiting foreign scholarship of 8th "Teacher Professional Development Project" fund by Shanghai Municipal Education Commission (No.201732), and Teaching construction project of Shanghai University of Engineering and Technology (No.P201701001).References

References

- [1] Madhusudan Singh, Dhananjay Singh and Antonio Jara. "Secure cloud networks for connected & automated vehicles," 2015 International Conference on Connected Vehicles and Expo (ICCVE), pp.330 – 335, 2015.
- [2] B. Hayes, Cloud computing, *Commun. ACM* 51(7) (July 2008) 9–11.
- [3] Gongjun Yan, Ding Wen, Stephan Olariu, and Michele C. Weigle, Security Challenges in Vehicular Cloud Computing, *IEEE transactions on intelligent transportation systems*, vol. 14, no. 1, march 2013
- [4] Tao Zhang, Fellow, IEEE, Helder Antunes, and Siddhartha Aggarwal, Defending Connected Vehicles Against Malware: Challenges and a Solution Framework, *IEEE internet of things journal*, vol. 1, no. 1, February 2014.
- [5] AlJahdali H, Albatli A, Garraghan P, Townend P, Lau L, Xu J. Multi-tenancy in cloud computing. In: *Service Oriented System Engineering (SOSE)*. 2014 IEEE 8th international symposium on, Oxford; 2014. p. 344–51. doi: 10.1109/SOSE.2014. 50.
- [6] G. Yan, S. Olariu, and M. C. Weigle, "Providing VANET security through active position detection," *Comput. Commun.*, vol. 31, no. 12, pp. 2883–2897, Jul. 2008, Special Issue on Mobility Protocols for ITS/VANET.
- [7] J. Sun, C. Zhang, Y. Zhang, and Y. M. Fang, "An identity-based security system for user privacy in vehicular ad hoc networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 9, pp. 1227–1239, Sep. 2010.
- [8] S. Almulla, Y-Y Chon, "Cloud Computing Security management", 2nd International Conference On Engineering Systems Management and Its Applications, pp.1-7, March 2010.

Non-bearing Masonry Walls Behavior and Influence to High Reinforced Concrete Buildings

Sorina Constantinescu*

Technical University of Construction Bucharest, Department of Civil Engineering, ZIP Code 011711, Romania

ARTICLE INFO

Article history:

Received: 28 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

Masonry walls stresses

Concrete walls structure

Dividing walls influence

Earthquake engineering

Finite element analysis

ABSTRACT

This is a study on non-bearing masonry walls, in a high, reinforced concrete walls building. It will be built in Bucharest, Romania. This is a high seismic area. The building will be used as a dwelling. The structure is composed of a ground floor and 10 stories above. It is interesting to see the interaction between the structure and the partitioning masonry walls. The paper presents the non-bearing walls design, the structure's behavior in the elastic and plastic stage, in particular the failing mechanism and the non-bearing walls stresses development. The paper will also compare the non-bearing walls seismic force from the design code and from the model. It will show the non-bearing walls important effect on the structure's behavior in the elastic and plastic stage.

1. Introduction

The paper presents the behavior of non-bearing masonry walls in high buildings built in high seismic areas. It is common to use masonry walls as partitions for high structures. They increase the lateral stiffness [1]. The literature contains studies showing the masonry walls capacity decreases with height, so they cannot be used as load bearing elements for a high structure like the one in study. Important cracks develop and reduce the bearing capacity by 40 to 60% [2, 3]. In many seismic countries, dwellings are built using reinforced concrete walls and nonbearing masonry walls [4]. Non-bearing masonry walls may get cracked from tensile stress from the concrete elements around them as they are subjected to deformation. [5]. On the other hand, very stiff masonry walls may crush the concrete structure [6]. Masonry does not perform well to lateral loads, as is gives in at shear stresses under 0.7 N/mm^2 [7]. The collapse process for non-bearing masonry walls in reinforced concrete walls buildings is important to be studied. This way the weakest elements can be strengthen and the collapse mechanism can be modified as needed [8]. Masonry stress-strain diagram shows stiffness degradation. The axial loading is a material degradation phenomenon caused by internal defects. This causes the process of crack extension until failure [9]. Pushover analysis can be used to predict a structure's failing mechanism, the maximum base force reached and rigidity loss. The plastic hinges are presumed to develop first at the beams ends due to exceeding the bearing bending moment [10, 11]. The

analyzed building is a dwelling. The structure contains reinforced concrete walls, beams connecting these walls and slabs. There are also non-bearing masonry walls. They can change a structure's behavior, as it will be shown here. The codes in force used to design the building are: [12–18]. It is important to establish the masonry walls behavior and the influence they have on the structure for the elastic and plastic stage. Nonlinear analysis for masonry is not something very common, as masonry is not a ductile material in itself.

2. Building Components

2.1. Building Description

The building in study is composed of a ground floor and 10 stories above it. Story height is 3m. It will be built in Bucharest, Romania. This is considered a high seismic area according to the seismic code in force, as the seismic acceleration is $0.30g$ (g is the gravity acceleration) [18]. For high buildings it is accustomed to use reinforced concrete walls, as both vertical and seismic loads reach important values. The structure is composed of reinforced concrete walls, beams and slabs. There are also non-bearing masonry walls at each story. These walls may or may not be used as elements in the structure's model. They do, however change the building's behavior, by increasing its stiffness. Both concrete and masonry walls are 25cm thick. The floor plan is presented in Figure 1. The structure in 3D is shown in Figure 2. Figures 3 and 4 explain the bays dimensions and the placement of concrete and masonry walls. In Figure 1 slabs are green and in Figure 2 they

*Sorina Constantinescu, 0742265890, sorina.constantinescu@yahoo.com

are grey. In both these figures beams are blue and walls are red. Figure 3 shows the concrete walls red with grey filling.

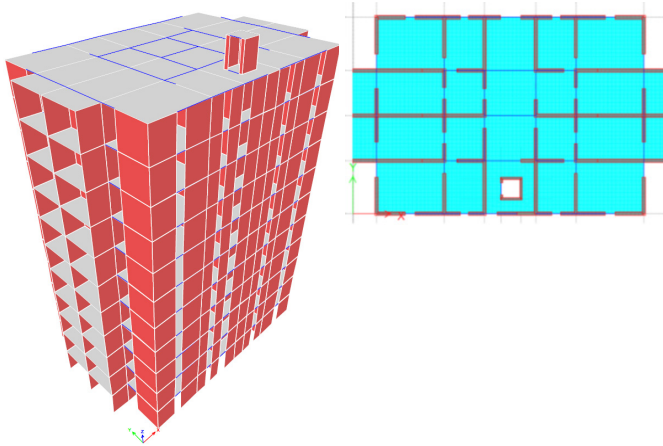


Figure 1: Structure in 3D

Figure 2: Floor plan

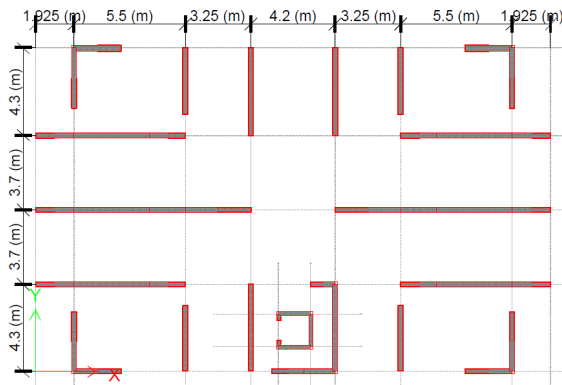


Figure 3: Reinforced concrete walls

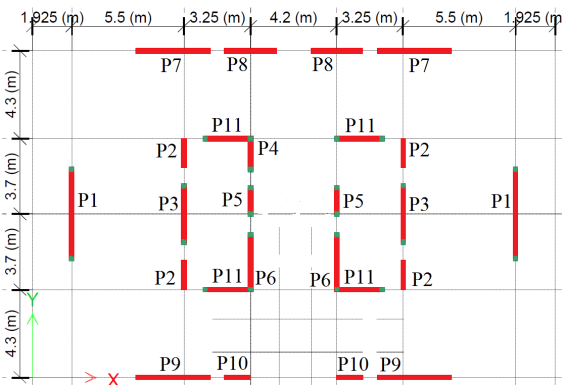


Figure 4: Masonry walls

Figure 4 contains the masonry non-bearing walls red and the slender columns green. The walls names used in design are also written. The software used for analysis is ETABS 2016.

2.2. Materials Properties

Materials used here are concrete C30/37 with elasticity modulus $E_C=33000\text{N/mm}^2$ and full bricks $240 \cdot 115 \cdot 63$ (mm) with standard strength $f_b = 12.5\text{N/mm}^2$, mortar strength $f_m=7.5\text{N/mm}^2$ and masonry elasticity modulus is $E_M= 1000 \cdot f_k = 4700\text{N/mm}^2$ in the elastic analysis [12]. Reinforcement bars are Bst 500. Steel elasticity modulus $E_S=210000\text{N/mm}^2$ [16]. Design strengths for

concrete and steel are calculated using the characteristic values (f_{ck} and f_{yk}) [12]. The walls stresses analyzed are: σ_x , σ_z and τ_{xz} .

$$f_{cd}=f_{ck}/\gamma_M=30/1.5= 20\text{N/mm}^2 \quad (1)$$

$$f_{yd}=f_{yk}/\gamma_M=500/1.15=435 \text{ N/mm}^2 \quad (2)$$

They are compared to masonry strengths in (4), (5) and (6). Design compression strengths on the horizontal (f_{hd}) and on the vertical direction (f_d) are determined from the characteristic masonry compression strengths (f_{hk} and f_k) using the insurance factor (γ_M) [12]. In (3) $K=0.55$ for full bricks [12]. Design shear strengths for horizontal and inclined direction ($f_{vd,0}$ and $f_{vd,i}$) are calculated by using the characteristic strengths ($f_{vk,0}$ and $f_{vk,i}$). $f_{bt}=0.035 \cdot f_b$ is the masonry characteristic tension strength [12].

$$f_k = 0.8 \cdot K \cdot f_b^{0.7} \cdot f_m^{0.3} \text{ N/mm}^2 \quad (3)$$

$$f_{hd} = f_{hk}/\gamma_M=1.91/1.9=1.0 \text{ N/mm}^2 \quad (4)$$

$$f_d = f_k/\gamma_M=4.7/1.9= 2.47 \text{ N/mm}^2 \quad (5)$$

$$f_{vd,0}=f_{vk,0}/\gamma_M = 0.3/1.9= 0.158 \text{ N/mm}^2 \quad (6)$$

$$f_{xd1}= f_{xk1}/ \gamma_M =0,24/1,9= 0,126 \text{ N/mm}^2 \quad (7)$$

$$f_{xd2}= f_{xk1}/ \gamma_M =0,48/1,9 = 0,25 \text{ N/mm}^2 \quad (8)$$

$$f_{vd,i}=f_{vk,i}/\gamma_M= 0.22 \cdot f_{bt} \cdot (1+5 \cdot \sigma_{0d}/f_{bt})/\gamma_M= 0.07 \text{ N/mm}^2 \quad (9)$$

σ_{0d} in (9) is the unitary pressure stress perpendicular to the shear stress direction. Design strength for horizontal and vertical stresses perpendicular to the wall (f_{xd1} and f_{xd2}) are calculated in (7) and (8) using the characteristic strengths (f_{xk1} and f_{xk2}) [12]. The concrete used is C16/20 ($E_C=29000\text{N/mm}^2$) for the slender columns connected to the masonry walls and the reinforcement bars steel is S355 ($E_S=210000\text{N/mm}^2$). These materials are not as strong as C30/37 and Bst 500, so they work better together with masonry. The strengths for C16/20 and S355 used for slender columns are seen in (10) and (11).

$$f_{cd}=f_{ck}/\gamma_M=16/1.5= 10.6\text{N/mm}^2 \quad (10)$$

$$f_{yd}=f_{yk}/\gamma_M=355/1.15=309 \text{ N/mm}^2 \quad (11)$$

3. Design Code Theory

3.1. Masonry Walls Bending Moments

Bearing bending moments perpendicular to the wall M_{Rxd1} (horizontal) and M_{Rxd2} (vertical) are calculated according to the design strengths perpendicular to the wall [12]. These values will be compared to the design bending moments values, M_{Exd1} (horizontal) and M_{Exd2} (vertical) [12] calculated from the model.

$$M_{Rxd1} = W_w \cdot (f_{xd1} + \sigma_{dw}) \quad (12)$$

$$M_{Rxd2} = W_w \cdot f_{xd2} \quad (13)$$

$W_w = 1000 \cdot t^2/6$ is the wall resistance modulus (in mm^3/m), t is the wall thickness, σ_{dw} is the compression stress at the wall's middle height section [12], $\gamma_{mas} = 18\text{kN}/\text{m}^3$ is masonry weight per cubic meter and $H_w = 2.5\text{m}$ is masonry walls height. The material properties written in this chapter are used in the elastic analysis together with the poisson's ratio $\nu=0.2$ for concrete and $\nu=0.3$ for masonry and steel.

$$\sigma_{dw} = \gamma_{mas} \cdot H_w/2 = 18 \cdot 2.5/2 = 0.0225 \text{ N}/\text{mm}^2 \quad (14)$$

$$W_w = 1000 \cdot t^2/6 = 1000 \cdot 250^2/6 = 10416666 \text{ mm}^3/\text{m} \quad (15)$$

$$M_{Rxd1} = 10416666.67 \cdot (0.126 + 0.0225) = 1.546 \text{ kNm}/\text{m} \quad (16)$$

$$M_{Rxd2} = 10416666.67 \cdot 0.25 = 2.604 \text{ kNm}/\text{m} \quad (17)$$

3.2. Seismic Action Evaluation

The base seismic force is calculated using: the building's importance-exposure coefficient $\gamma_{I,e} = 1.2$, the elastic spectrum maximum value $\beta_0 = 2.5$ and the structure's behavior factor $q = 3 \cdot k_w \cdot \alpha_u / \alpha_1 = 3 \cdot 1 \cdot 1.15 = 3.45$ [18]. α_u / α_1 is the base shear force value for the failing mechanism/the base shear force value for the first plastic hinge, m = building's mass. $\lambda = 0.85$ for buildings higher than 3 stories, $a_g = 0.30g$ [18], G = building's weight. This is a medium ductility structure: DCM [18].

$$F_b = \gamma_{I,e} \cdot \beta_0 \cdot a_g \cdot m \cdot \lambda / q = c_s \cdot G = 0.24 \cdot G \text{ [kN]} \quad (18)$$

3.3. Seismic Force Perpendicular to the Masonry Walls

$$F_{NBW}(z) = \gamma_{I,e} \cdot a_g \cdot \beta_{NBW} \cdot k_z \cdot m_{NBW} / q_{NBW} = 1.88 \text{ kN}/\text{m}^2 \quad (19)$$

The force is considered uniformly distributed, perpendicular to the non-bearing walls [18]. $\beta_{NBW} = 1$ is the non-bearing walls amplification factor, k_z is a coefficient according to the non-bearing wall's level (the distance to the building's base), z is the non-bearing wall's level and H is the building height [18].

$$k_z = 1 + 2 \cdot z/H \quad (20)$$

$$k_z = (k_{z1} + k_{z2})/2 = (1 + 2 \cdot z_1/H + 1 + 2 \cdot z_2/H)/2 = (1 + 2 \cdot 33/33 + 1 + 2 \cdot 30/33)/2 = 2.91 \quad (21)$$

k_{z1} and k_{z2} are coefficients that refer to the highest and lowest points of the wall. Of course, the greatest value for k_z is calculated at the top building story. $q_{NBW} = 2.5$ is the behavior factor for non-bearing walls. $m_{NBW} = \gamma_{mas} \cdot t = 18 \cdot 0.25 = 4.5\text{kN}/\text{m}^2$ is the wall mass/ m^2 . Seismic force value F_{NBW} is limited as (22) shows [18].

$$0.75 \cdot \gamma_{I,e} \cdot a_g \cdot m_{NBW} \leq F_{NBW} \leq 4 \cdot \gamma_{I,e} \cdot a_g \cdot m_{NBW} \quad (22)$$

$$0.75 \cdot 1.2 \cdot 0.30 \cdot 4.5\text{kN}/\text{m}^2 \leq 1.57 \text{ kN}/\text{m}^2 \leq 4 \cdot 1.2 \cdot 0.30 \cdot 4.5 \text{ kN}/\text{m}^2$$

$$1.215 \text{ kN}/\text{m}^2 \leq 1.57 \text{ kN}/\text{m}^2 \leq 6.48 \text{ kN}/\text{m}^2$$

3.4. Masonry Walls Design

The masonry walls are not loaded as much as bearing walls would be, but they are subjected to sectional efforts: axial force N_{Ed} , bending moment M_{Ed} and shear force V_{Ed} .

$$M_{Rd} = M_{Rd(M)} + M_{Rd(Aas)} \text{ [kNm]} \quad (23)$$

$M_{Rd(M)}$ and $M_{Rd(Aas)}$ are the bending moments the masonry area and slender columns can take. The wall compressed length is l_C . z is the distance from the wall's weight center to the compressed masonry area center.

$$I_C = N_{Ed} / (0.85 \cdot t \cdot f_d) \text{ [mm}^2] \quad (24)$$

$$M_{Rd(M)} = N_{Ed} \cdot z \text{ [kNm]} \quad (25)$$

$$M_{Rd(Aas)} = d_s \cdot A_s \cdot f_{yd} \text{ [kNm]} \quad (26)$$

d_s is the distance between the slender columns at the wall's margins centers. A_s is the slender columns reinforcement area. $t \cdot l_C = C_A$ is the walls compressed area. z_{WC} is the compressed area center. The wall section center is W_C . E is the earthquake action. V_{Rd} is the confined masonry walls bearing shear force and V_{Rd1}^* is the bearing shear force taken by the masonry panel [12]. V_{Ed} is the horizontal shear force from seismic loads.

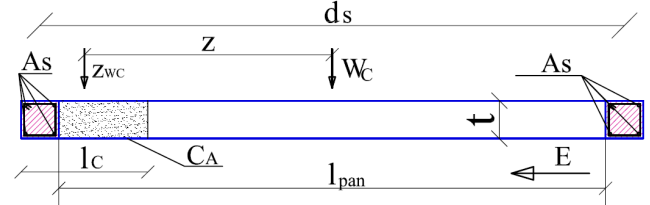


Figure 5 Confined masonry wall section

V_{Rd2} is the bearing horizontal shear force from the slender column reinforcement at the walls compressed edge. A_s is the reinforcement area in the slender column at that walls edge. λ_c is the reinforcement participation factor [12]. h_{pan} is the masonry wall height (2.5m). The load combination used to design the structure is 1.0·permanent loads+0.4·variable loads+1.0·seismic loads.

$$V_{Rd} = V_{Rd1}^* + V_{Rd2} \text{ [kN]} \quad (27)$$

$$V_{Rd1}^* = 0.4 \cdot N_{Ed} + 0.8 \cdot V_{Ed} \cdot h_{pan} / I_{pan} \text{ [kN]} \quad (28)$$

$$V_{Ed} \leq I_{pan} \cdot t \cdot f_{vd,0} \quad (29)$$

$$V_{Rd2} = \lambda_c \cdot A_s \cdot f_{yd} \text{ [kN]} \quad (30)$$

4. Elastic Stage Results

4.1. Masonry Nonbearing Walls Stresses

Piers P1, P3, P4, P5, P6 and P11 need slender columns reinforced with A_s consisting of 4 bars of 14mm diameter, to withstand the bending moment and shear force they are subjected to. The piers names are shown in Figure 4. Efforts in piers vary slightly from one story to another. Theoretically, non-bearing walls are subjected only to their own weight and the seismic force

according to that weight. However, those non-bearing walls are placed in different places and on different stories in a high structure. The position of each wall in the structure influences the efforts they are subjected to.

Table 1 Masonry walls design and bearing efforts

Wall (pier)	Wall length [mm]	N _{Ed} [kN]	M _{Ed} [kNm]	M _{Rd} [kNm]	A _s [mm ²]	V _{Ed} [kN]	V _{Rd} [kN]
P1	4350	68	85	1021	616	12	97
P2	1425	24	15	16	0	4	41
P3	2750	46	89	634	616	8	88
P4	1525	29	40	360	616	4	81
P5	1275	26	20	307	616	7	80
P6	2675	59	179	634	616	4	93
P7	3650	76	60	133	0	10	62
P8	2550	50	26	61	0	7	51
P9	3650	76	32	133	0	10	62
P10	1275	32	11	19	0	11	44
P11	2225	65	47	525	616	47	89

4.2. Natural Vibration Periods

The influence of non-bearing masonry walls to the structure’s stiffness is evaluated first in the elastic stage, by the natural vibration periods. The building with non-bearing masonry walls building shows values reduced to 88%.

Table 2 Natural vibration periods

	Natural vibration periods for reinforced building with non-bearing masonry walls	Natural vibration periods for reinforced building without non-bearing masonry walls
Mode 1	0.581s	0.659s
Mode 2	0.471s	0.545s
Mode 3	0.382s	0.417s

4.3. Story Displacements

The story displacements diagram shows very similar values for X and Y at all stories (story height=3m) if masonry nonbearing walls are present. Story displacements are greater if there are no stiff partitioning walls. The highest values are reached on Y. This may be because there are 4 long concrete walls on direction X, that provide stiffness to the structure. It is interesting that the nonbearing walls balance the building stiffness, at least for the elastic stage.

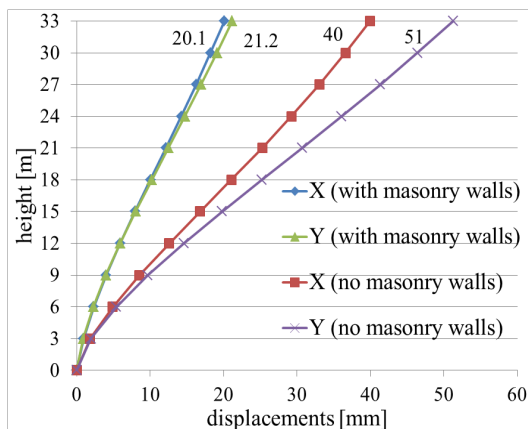


Figure 6 Story displacements

5. Plastic Stage Behavior

Four pushover cases (PX, PY, PX1 and PY1) are used for the building’s nonlinear analysis. PX and PY diagrams resulted from pushover cases on directions X and Y when the masonry non-bearing walls were taken into consideration in the structure’s analysis. PX1 and PY1 diagrams resulted from pushover cases on directions X and Y without the masonry non-bearing walls in the analysis. It is considered the analysis ends when the chosen displacement (500mm) is reached. The study only highlights the beams plastic hinges development.

5.1. Nonlinear Hinges Development

Figures 7 to 10 show the development stages for plastic hinges in beams, when the chosen displacement is reached.

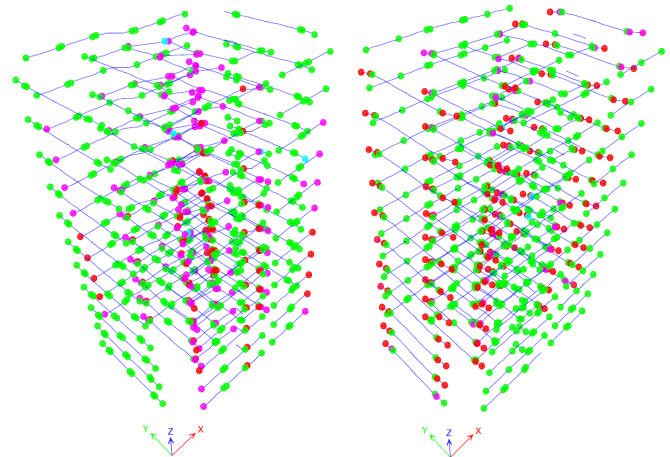


Figure 7 PX at step 90

Figure 8 PY at step 98

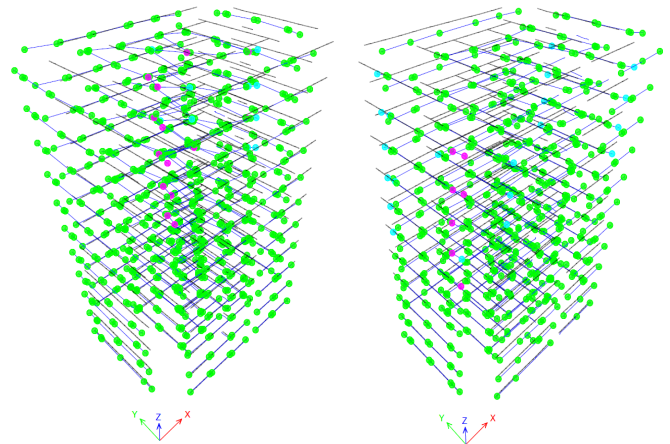


Figure 9 PX1 at step 151

Figure 10 PY1 at step 78

This value is great enough for important damages to occur to the structure, so the plastic mechanism is already formed at this stage. Directions X and Y are shown in each graphic by the red and green arrows. The color code is the following: B (green) means the plastic hinge is formed, C (blue) means the plastic hinge reaches the limit and the element gives out, D (pink) means the load was redistributed and E (red) means collapse. These colors are seen in Figures 9 and 10. For pushover cases PX and PY, plastic hinges reach stage E mainly in beams on the same direction as the case. There are more red hinges for PY, because they mainly develop in the short beams. There are no short beams on X. For cases PX1

and PY1 the hinges in stage D are located more to the edge short beams on direction Y, at the analysis end. Shorter beams are loaded more by the seismic action. For cases PX1 and PY1 only a few hinges reach stage D. Stage C is seen in more hinges for case PY1. They are all in short beams at the structure sides on direction Y.

The hinges developed to the highest stages are distributed more evenly for direction Y cases. This may be because the building is more flexible. Hinges do not reach stage E for PX1 and PY1. The building does turn into a plastic mechanism for case PX1, before displacement 50cm is reached. This means the building behaves more ductile for these cases. Elasticity modulus values in the plastic stage are E_{Cpl} in (31) for concrete, E_{Spl} in (32) for reinforcement bars [16] and E_{Mpl} in (33) for masonry [12]. f_{cm} is the concrete medium pressure strength, ϵ_{c1} is the strain reached for the maximum stress in the plastic stage, $k=f_t/f_y$ =tension strength/elasticity limit strength and ϵ_{uk} is the maximum strain reached in the plastic stage [16]. ϵ_{m1} is the strain reached for the maximum stress value in the plastic stage [12].

$$E_{Cpl} = 0.8 \cdot f_{cm} / (0.5 \cdot \epsilon_{c1}) \quad (31)$$

$$E_{Spl} = k \cdot f_{yk} / \epsilon_{uk} \quad (32)$$

$$E_{Mpl} = f_k \cdot 1000 / \epsilon_{m1} \quad (33)$$

5.2. Pushover Diagrams

Figure 11 shows the pushover curves for all 4 cases. The pushover diagrams PX and PY show a rigid behavior compared to the slower decrease of stiffness for PX1 and PY1.

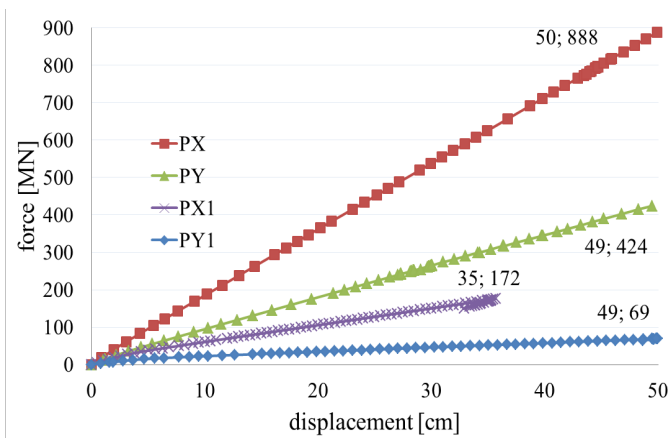


Figure 11 Pushover diagrams

The structure maintains the same stiffness for PX until the required displacement is reached. PY and PX1 show a stiffness decrease at the analysis beginning. For PX1 the structure gives in at an earlier stage. Diagram PY1 displays a low rigidity from the beginning that drops to 0 towards the end. The structure is clearly stiffer on direction X. This is expected, as there are 4 long concrete walls on direction X. It is also clear the masonry non-bearing walls have a great influence on the structure's stiffness. Maximum base shear forces reached for PX1 and PY1 are 2 to 3 times smaller than for cases PX and PY, for the same displacement value.

5.3. Masonry Walls Stresses

The maximum value for each stress is written after the analysis step number. Stresses σ_x values surpass the strength f_{hd} at step 8 for PX and step 12 for PY. The highest values are reached at the walls bottoms perpendicular to each pushover case direction. The stress values generally decrease to the walls tops. There is an increase in stress for walls developed on both directions. This may be explained as they are stiffer. Stresses σ_z clearly surpass the strengths f_d at steps 8 for PX and 9 for PY. The walls perpendicular to the stress direction are the most affected.

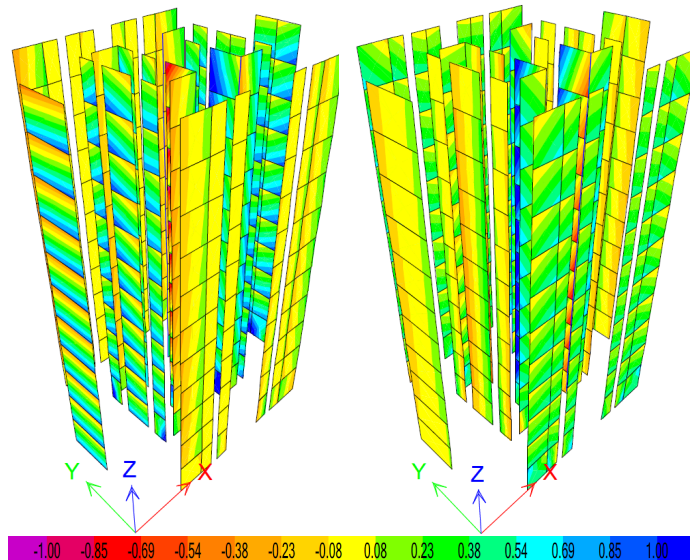


Figure 12 PX step 8 $\sigma_x=1$ N/mm²

Figure 13 PY step 12 $\sigma_x=1$ N/mm²

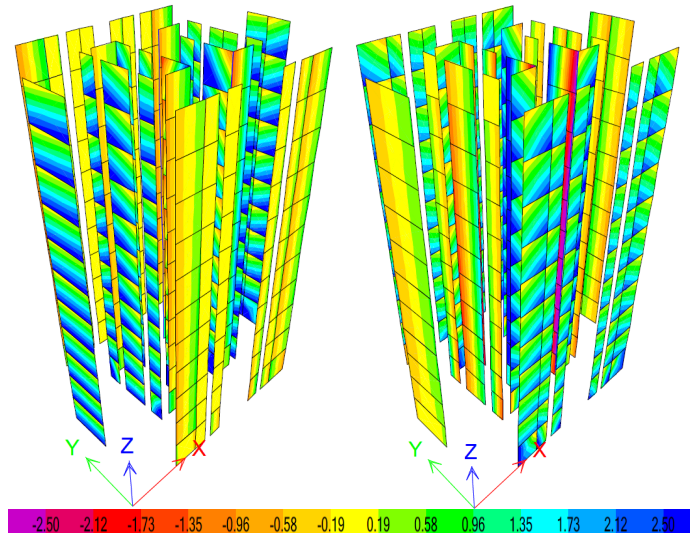


Figure 14 PX step 8 $\sigma_z=2.5$ N/mm²

Figure 15 PY step 9 $\sigma_z=2.5$ N/mm²

The stress pattern shows high values at the bottoms, as for consoles. This is expected because non-bearing walls are not stiffly connected to the structure at the top, otherwise they would be loaded as bearing walls are. The walls developed on both directions show higher stress values on the same direction as the pushover case. There is a pattern of compression and tension in the walls parallel to the stress, but the values do not increase at the bottom stories, because those are non-bearing walls. For both pushover cases, stresses τ_{xz} are greater than $f_{vd,1}$ from step 1 of the

analysis. τ_{xz} is greater in walls on the pushover case direction and at the bottoms of walls perpendicular to the stress. The stress reaches the highest values in the stiffest walls, developed on both directions. Stress increases at the top stories, as if the walls are crushed at one side and stretched at the other. This may be caused by the structure's slenderness.

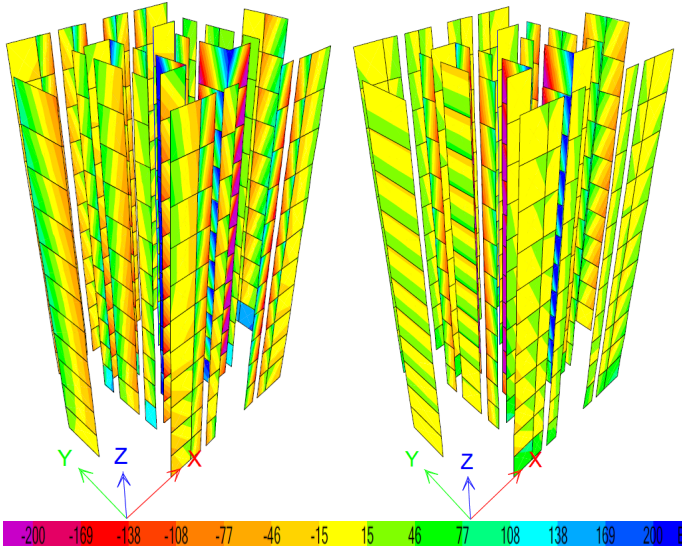


Figure 16 PX step 1 $\tau_{xz}=0.2 \text{ N/mm}^2$ Figure 17 PY step 1 $\tau_{yz}=0.2 \text{ N/mm}^2$

5.4. Bending Moments Perpendicular to the Walls

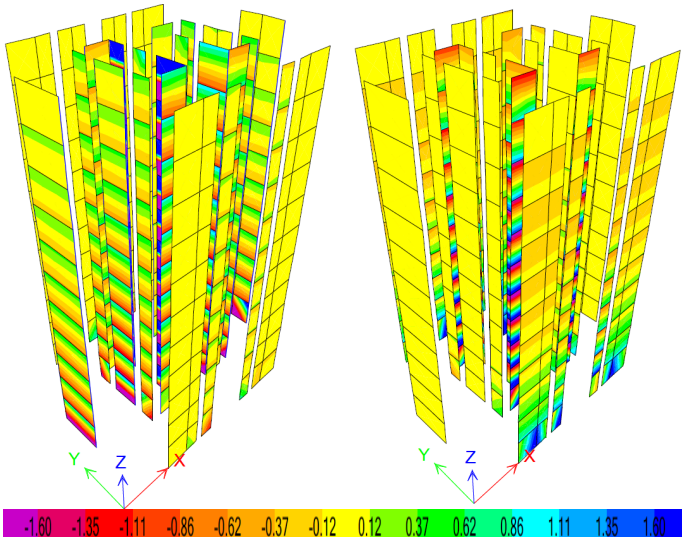


Figure 18 PX step 6 $M_{Exd1}=1.6 \text{ kNm/m}$ Figure 19 PY step 10 $M_{Eyd1}=1.6 \text{ kNm/m}$

Nonlinear analysis shows the bearing bending moments perpendicular to the wall M_{Rxd1} (horizontal) and M_{Rxd2} (vertical) are surpassed by the design bending moments M_{Edx1} and M_{Edx2} in the nonlinear stage, at step 6 for PX and 10 for PY. It was expected that the bearing moment values would not be exceeded in the elastic stage.

5.5. Seismic Forces Perpendicular to the Walls

The seismic force perpendicular to the walls determined by the design code is reached at step 7 of the nonlinear analysis. The highest values are in walls perpendicular to the pushover direction.

This is expected, as it is the seismic force perpendicular to the walls. These values are reached at the lower stories. This can be explained as the structure's lower part is stiffer and the non-bearing walls can also be loaded more, because they do interact with the structure. For PX, there are 2 walls developed on both X and Y directions, that take most of the perpendicular force. This force reaches the same maximum value from the bottom to the top walls. This is seen on direction Y. The building is less stiff on Y. This may cause non-bearing walls to be loaded more.

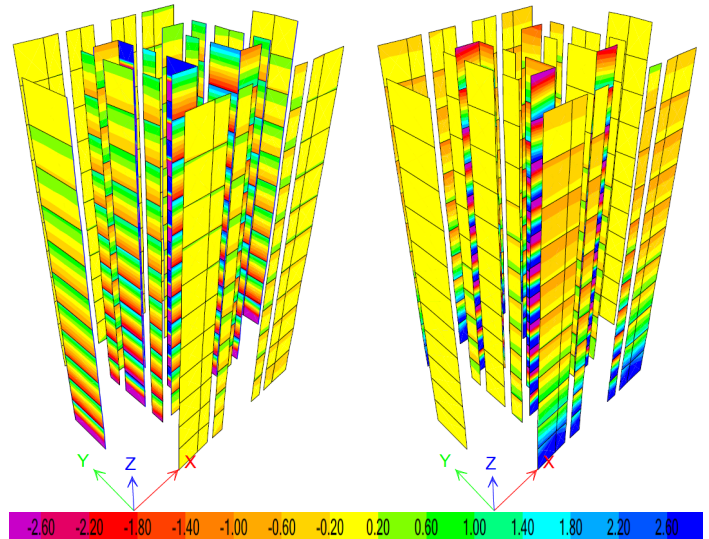


Figure 20 PX step 4 $M_{Exd2}=2.6 \text{ kNm/m}$ Figure 21 PY step 8 $M_{Eyd2}=2.6 \text{ kNm/m}$

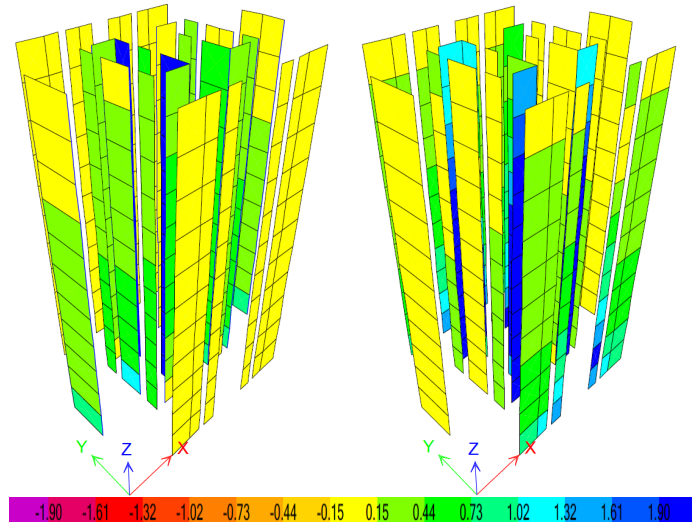


Figure 22 PX step 7 $F_{NBW}=1.9 \text{ kN/m}^2$ Figure 23 PY step 7 $F_{NBW}=1.9 \text{ kN/m}^2$
experimental seismic force experimental seismic force

5.6. Mesh Discretization Importance

Mesh elements dimensions have an impact on the analysis results. To study this impact, a 2D elevation in the building in study was used. It is seen in Figure 24. Figure 25 shows the pushover diagrams for different mesh discretization. The base force reached subsides as the mesh is smaller. The base force drops to 0.89 times the value as the mesh dimensions decrease 8 times.

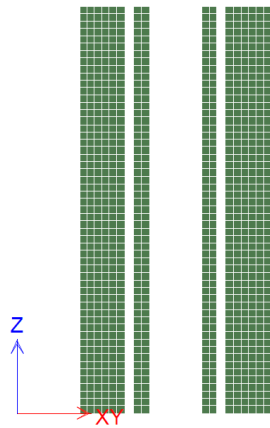


Figure 24 2D elevation

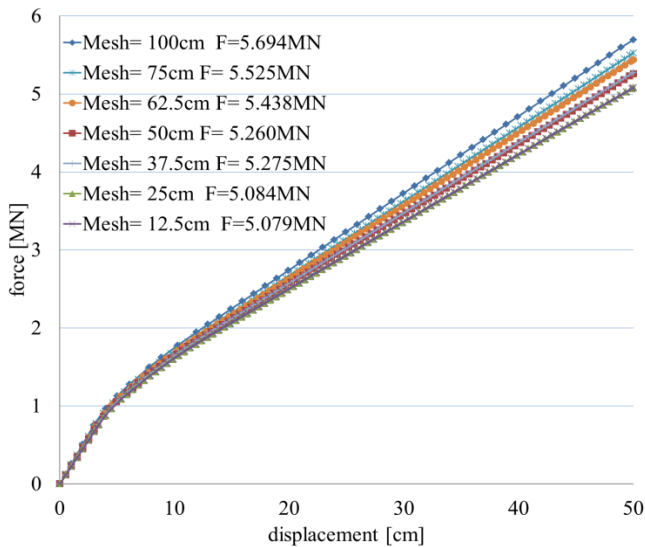


Figure 25 Pushover diagrams for different mesh discretization

6. Conclusions

Non-bearing masonry walls are able to bear the perpendicular and parallel forces and also the bending moments they are subjected to in the elastic stage. They reduce the building's natural vibration periods values. The pushover curves show a stiff behavior when the masonry walls are present. The maximum base shear force increases 2 times if non-bearing walls are present. When designing a high building with partitioning masonry walls, it is important to take the non-bearing masonry walls into account as elements, not only as loads.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] A. Madan, A. K. Hashimi, "Performance Based Design of Masonry Infilled Reinforced Concrete Frames for Near-Field Earthquakes Using Energy Methods" WASET, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering Vol:8, No:6, p 689-695, 2014. <https://pdfs.semanticscholar.org/bbdd/eb245adce7d0b1604f474ad249e5a57c1397.pdf>
- [2] M. Vokal, M. Drahorad "Non-linear analysis of slender masonry beam" Transactions of the VSB- Technical University of Ostrava Civil Engineering Series, Vol. 17, No. 2 p 151-160, 2017. <http://dx.doi.org/10.1515/tvsb->

2017-0039

- [3] C. Cornado, J.R. Rosell, J. Leiva, C. Diaz, "Experimental study of brick masonry walls subjected to excentric and axial load" International RILEM Conference on Materials, Systems and Structures in Civil Engineering Conference segment on Historical Masonry Technical University of Denmark, Lyngby, Denmark p 33-40, 2016. www.rilem.net/publications/proceedings-500218
- [4] M. Teguh, "Experimental evaluation of masonry infill walls of RC frame buildings subjected to cyclic loads" Elsevier Procedia Engineering 171 Sustainable Civil Engineering Structures and Construction Materials 2016 p 191-200, 2017. doi:10.1016/j.proeng.2017.01.326
- [5] P. Lourenco, "Design of large size non-loadbearing masonry walls: case study in Portugal. Technical and economic benefits" in 13th International Brick and Block Conference Amsterdam, 2004. www.hms.civil.uminho.pt/ibmac/2004/
- [6] M. Dhanasekar, "Shear in reinforced and unreinforced masonry: response, design and construction" in The Twelfth East Asia Pacific Conference on Structural Engineering and Construction, Elsevier Procedia Engineering Vol 14, p 2069- 2076, 2011. doi:10.1016/j.proeng.2011.07.260
- [7] M. Kaluza, "Analysis of in plane deformation of walls made using AAC blocks strenghten by GFRP mesh" International Conference on Analytical Models and New Concepts and Masonry Structures AMCM Elsevier Procedia Engineering Vol 14, p 393-400, 2017. doi:10/1016/j.proeng.2017.06.229
- [8] X. Lu, X. Lin, Y. Ma, Y. Li, L. Ye, "Numerical Simulation for the Progressive Collapse of Concrete Building due to Earthquake" in Proc. The 14th World Conference on Earthquake Engineering 2008, October 12-17 Beijing, China http://www.iitk.ac.in/nicee/wcee/article/14_14-0044.PDF
- [9] Z. Huang, T. Liao, B. Huang, B. Huang, Z. Li, B. Zhang, J. Pan, W. Qi, X. Li, J. Wang, "Investigation on unified model of constitutive relations for masonry" in 3dr International Conference on Energy Materials and Environment Engineering IOP Conf. Series: Earth and Environmental Science 61, 2017. doi:10.1088/1755-1315/61/1/012125
- [10] P. Naik, S. Annigeri, "Performance evaluation of 9 story RC building located in North Goa" in 11th International Symposium on plasticity and Impact Mechanics, Implast 2016 Elsevier Procedia Engineering 173, p 1841 -1846, 2017. doi:10.1016/j.proeng.2016.12.231
- [11] B. R. Patel, "Progressive Collapse Analysis of RC Buildings Using Non Linear Static and Non-Linear Dynamic Method" in IJETAE 2014; ISSN 2250-2459, ISO 9001:2008 Certified Journal. <https://pdfs.semanticscholar.org/3716/2c9411ca6362c448779d5af0f09f38f67de9.pdf>
- [12] CEN EN 1996-1-1-2006 Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures, 2006.
- [13] CEN EN 1991-1-1-2004 Eurocode 1: Actions on structures - Part 1-1: General actions- Densities, self-weight, imposed loads for buildings, 2004.
- [14] CEN EN 1990-2004 Eurocode 0: Basics of structural design, 2004
- [15] CEN EN 1991-1-3-2005 Eurocode 1: Actions on structures - Part 1-3: General actions- Snow loads, 2005
- [16] CEN EN 1992-1-1-2004 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, 2004.
- [17] CEN EN 1998-1-2004 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, 2004.
- [18] P100-1/2013 Seismic design code – Part 1- General rules for buildings, 2013

Slender Confined Masonry Buildings in High Seismic Areas

Sorina Constantinescu*

Technical University of Construction Bucharest, Department of Civil Engineering, ZIP Code 011711, Romania

ARTICLE INFO

Article history:

Received: 28 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

ABSTRACT

This is a study on a confined masonry slender walls building in a high seismic area. The structure also contains frames, but the walls bear most of the gravity and seismic loads. The building will be used as a school. It will be built in Bucharest, Romania. It contains a ground floor and 2 stories above it. Story height is 4m. The structure is interesting as it is not common practice to use slender masonry walls for buildings with large bays. Such a structure is allowed by the design codes in force but this solution is not often used. The building will be studied in the elastic state, as the structure bears important gravity and seismic loads, then in the plastic state to establish the walls stresses development and the failure mechanism. It is interesting to see how the walls behave in the nonlinear stage, as they are slender, but masonry is a stiff material.

1. Introduction

The paper presents the behavior of slender walls confined masonry buildings, in high seismic areas. The literature points out that slender masonry walls may experience buckling [1] and the axial bearing capacity may be calculated on the deformed shape [2]. Slender walls have less shear force bearing capacity [3]. Confined masonry walls behave well under seismic loading, [4], as the concrete elements increase the energy dissipation, ductility and cracking pattern [5]. Reinforcement bars connecting tie columns to the masonry panels help them work together better [6]. Confining clamps at tie columns tops and bottoms are very important for preventing their shear failure [7]. Masonry walls can behave well in the nonlinear state, as their ductility may be improved by using proper confining reinforced concrete elements [8]. Three dimensional analyses are important as they are able to capture the collapse mechanisms well [9]. Laboratory tests show that confined masonry walls main failing mechanism is diagonal cracking, sliding on the bed joint of brick mortar [10]. It is adequate to analyze a structure's behavior in the plastic stage, to predict its possible failure mechanism, and the maximum base force it can bear. The moment-curvature diagram may show the ductility or stiffness of the building in study [11]. The structure analyzed here is mostly composed of confined masonry walls, but there are also frames due to the illumination condition. Confined masonry walls are the main bearing elements. The building in study is composed of a ground floor and 2 stories above it. Story height is 4m, and the bays spans are up to 9m, so the masonry

walls are very slender. This is a special structure, as nowadays it is most common to use masonry structures with bays up to 5m and story heights up to 3.2m, because masonry is a fragile material and does not behave well under seismic loading. The design codes in force used to design the structure are: [12–18]. The building is a school. It will be built in Bucharest, Romania. This is considered a high seismic area according to the seismic code in force, as the seismic acceleration is 0.30g (g is the gravity acceleration) [18]. The building's behavior is studied for both elastic and plastic state.

2. Building Description

The structure in study is composed of 3 stories: a ground floor and 2 floors above it. The bays are up to 9m and the story height is 4m. This is regarded as a large bay confined masonry walls building. Confined masonry means that the brick walls are framed by reinforced concrete elements: tie beams (belts) at each story level, placed horizontally and slender columns placed vertically. Those reinforced concrete elements work together with the brick panels. Walls are 30cm thick so the slender columns and the belts need to be 30cm wide. It is allowed to design such buildings in high seismic areas [18]. For confined masonry buildings, in the design code it is recommended to use less strong concrete, class C12/15 or C16/20 [12, 16], because it interacts better with masonry. In Figures 1 and 2 the walls are red, slender columns are green, belts and beams are blue, short beams are white and slabs are grey. This building measures 35m on direction X and 22.25m on Y. Distances between load bearing elements are given in meters. The software used for analysis is ETABS 2016.

*Sorina Constantinescu, 0742265890, sorina.constantinescu@yahoo.com

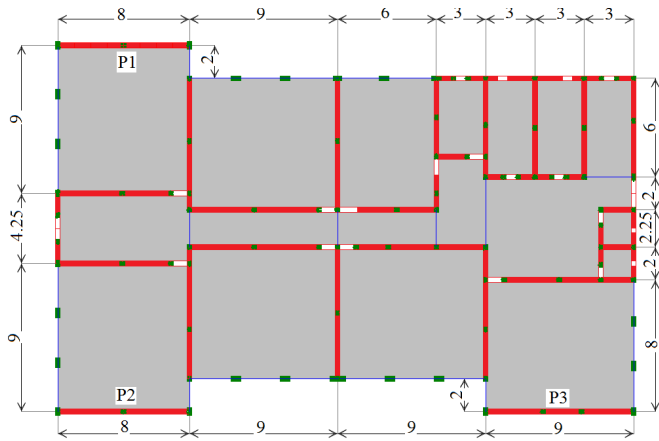


Figure 1 Floor plan

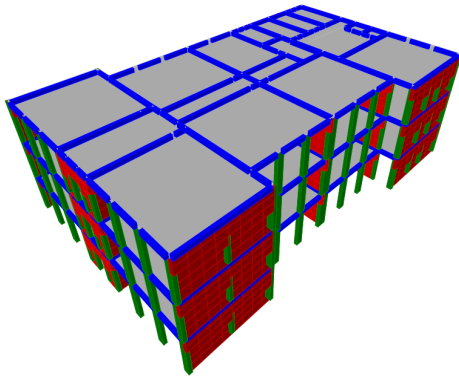


Figure 2 Building in 3D

3. Structure Design Theory

3.1. Materials Used

Materials used here are: concrete C16/20 with $E_c=29000\text{N/mm}^2$ (E is elasticity modulus) [16] and vertical perforated bricks $290 \cdot 140 \cdot 88$ (mm) with $E_M=4500\text{N/mm}^2$ [12] with standard strength $f_b=10\text{N/mm}^2$ and mortar M10 [12]. Reinforcement bars are S345 $E_s=210000\text{N/mm}^2$ [16]. The walls stresses analyzed are: $\sigma_x, \sigma_z, \tau_{xz}, \tau_{xy}$ and τ_{yz} . They are compared to the masonry strengths (1), (2), (3), (4) and (5) [12]. Horizontal design compression strength is f_{hd} and the vertical one is f_d .

$$f_{hd} = f_{hk}/\gamma_M = 0.98/2.2 = 0.445 \text{ N/mm}^2 \quad (1)$$

$$f_d = f_k/\gamma_M = 4.5/2.2 = 2.04 \text{ N/mm}^2 \quad (2)$$

f_{hk} and f_k are the characteristic masonry compression strengths. γ_M is the insurance factor [12]. Design shear strength for horizontal direction $f_{vd,1}$ (3) is calculated by using the characteristic strength ($f_{vk,0}$) and the unitary vertical stress (σ_d) [12].

$$f_{vd,1} = f_{vk,0}/\gamma_M + 0.4 \cdot \sigma_d = 0.3/2.2 + 0.4 \cdot 0.1 = 0.18 \text{ N/mm}^2 \quad (3)$$

Design strength for horizontal and vertical stresses perpendicular to the wall (f_{xd1} and f_{xd2}) are calculated in (4) and (5) using the characteristic strengths (f_{xk1} and f_{xk2}) [12]. Concrete and steel reinforcement bars design strengths are determined in (6) and (7) using the characteristic values (f_{ck} and f_{yk}) [16].

$$f_{xd1} = f_{xk1}/\gamma_M = 0.24/2.2 = 0.11 \text{ N/mm}^2 \quad (4)$$

$$f_{xd2} = f_{xk2}/\gamma_M = 0.48/2.2 = 0.22 \text{ N/mm}^2 \quad (5)$$

$$f_{cd} = f_{ck}/\gamma_M = 16/1.5 = 10.6 \text{ N/mm}^2 \quad (6)$$

$$f_{yd} = f_{yk}/\gamma_M = 345/1.15 = 300 \text{ N/mm}^2 \quad (7)$$

3.2. Seismic Action Evaluation

The coefficient c_s is calculated using the base force F_b in (8) according to [18]. $\gamma_{1,e} = 1.2$ is the building's importance-exposure coefficient, $\beta_0 = 2.5$ is the elastic spectrum maximum value, q is the structure's behavior factor, $q = 2.25 \cdot \alpha_w/\alpha_1 = 2.25 \cdot 1.25$ [18], α_w/α_1 = the base shear force value for the failing mechanism/the base shear force value for the first plastic hinge, m = building's mass, $\eta = 0.88$ is the reduction factor according to ξ (damping ratio for masonry) = 8% [16]. $\lambda = 0.85$ for 3 stories buildings, $a_g = 0.30g$ [18] and G = building's weight.

$$F_b = \gamma_{1,e} \cdot \beta_0 \cdot a_g \cdot m \cdot \lambda \cdot \eta / q = c_s \cdot G = 0.24 \cdot G \text{ [kN]} \quad (8)$$

3.3. Walls Bearing Axial Force

The bearing axial force N_{Rd} for masonry walls is calculated using [12]. N_{Rd} = wall's bearing axial force at the bottom story, $\Phi_{i(m)} = \min(\Phi_i; \Phi_m)$ wall strength reduction factor (buckling factor), Φ_i = wall strength reduction factor at the wall's top and bottom, Φ_m = wall strength reduction factor at the wall's center, t = wall thickness, e_i = wall eccentricity at the top and bottom and A = wall section area [12]. The load combination used to determine N_{Ed} here is: 1.35·permanent loads+1.5·variable loads.

$$N_{Rd} = \Phi_{i(m)} \cdot A \cdot f_d \quad (9)$$

$$\Phi_i = 1 - 2 \cdot e_i / t \quad (10)$$

$$e_i = e_{0i} + e_{hi} + e_a \geq 0.05 \cdot t \quad (11)$$

$e_{0i} = 0$, vertical loads eccentricity, $e_{hm(i)}$ = eccentricity from forces perpendicular to the wall, N_1 = axial load from the upper story wall and N_2 = load from the slabs above the story [12]. $M_{hm(i)}$ = bending moment from forces perpendicular to the wall, e_a = accidental eccentricity and h_s = story height [12].

$$e_{hm(i)} = M_{hm(i)} / (N_1 + \Sigma N_2) \quad (12)$$

$$e_a = \max(t/30; h_s/300; 1\text{cm}) \quad (13)$$

$$e_m = 2/3 \cdot e_{0i} + e_{hm} + e_a \quad (14)$$

Table 1 Walls strength reduction factor at middle story height (Φ_m)

slenderness (h_s/t) _{max}	masonry type	relative eccentricity e_m/t					
		0.05	0.10	0.15	0.20	0.25	0.30
15	CM	0.75	0.64	0.53	0.42	0.32	0.22

In Table 1, CM means confined masonry. The maximum walls slenderness allowed is 15 [12]. Φ_m is determined from Table 1, according to h_s/t [12], e_m = wall eccentricity at middle story height

and e_{hm} = eccentricity from forces perpendicular to the wall, at middle story height [12].

3.4. Walls Bearing Bending Moment

The bearing bending moment M_{Rd} for masonry walls associated to the design axial force N_{Ed} [12] is calculated in (15). Walls compressed area is A_c . The bearing bending moment borne by the masonry area is $M_{Rd(M)}$ (17). y_c is the distance from the wall's weight center to the compressed masonry area center [12]. The bearing bending moment borne by the slender columns reinforcement at the wall ends is $M_{Rd(As)}$ (18). The load combination used to determine N_{Ed} and M_{Ed} here is: 1.0·permanent loads+0.4·variable loads+1.0·seismic loads.

$$M_{Rd} = M_{Rd(M)} + M_{Rd(As)} \text{ [kNm]} \quad (15)$$

$$A_c = N_{Ed} / (0.85 \cdot f_d) \text{ [mm}^2\text{]} \quad (16)$$

$$M_{Rd(M)} = N_{Ed} \cdot y_c \text{ [kNm]} \quad (17)$$

$$M_{Rd(As)} = I_s \cdot A_s \cdot f_{yd} \text{ [kNm]} \quad (18)$$

$f_{yd} = 300 \text{ N/mm}^2$ is the reinforcement bars design strength. I_{as} is the distance between slender columns at the walls ends. A_s is the slender columns longitudinal reinforcement area. C_c is the compressed area gravity center. C is the wall section gravity center. S is the seismic action [12]. $b = n \cdot t$ and $n = f_{cd} / f_d$.

$$b = t \cdot f_{cd} / f_d \quad (19)$$

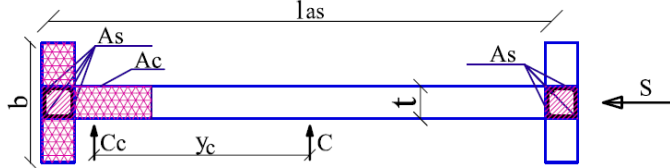


Figure 3 Wall horizontal section

4. Building Design Results

The results shown here are story displacements, steel reinforcements in concrete elements and efforts in the walls most vulnerable to seismic action. They are vulnerable because they are slender and not connected to other walls on the other direction, so they are susceptible to fogging.

4.1. Building Story Displacements

The story displacements will be checked for the elastic state, for both directions X and Y. In the elastic state, the maximum relative story displacements are $dr = 0.07 \text{ cm}$ on direction X and $dr = 0.15 \text{ cm}$ on direction Y. Maximum displacements are calculated with formulas (20) and (21), considering the building has fragile non-bearing elements, are checked for direction X ($0.098 \text{ cm} < 2 \text{ cm}$) in (20) and for direction Y ($0.21 \text{ cm} < 2 \text{ cm}$) in (21). $h_s = 4 \text{ m}$ is the story height, $v = 0.5$ is a reduction factor for class I and II importance buildings and $q = 2.81$ is the structure's behavior factor [18].

$$dr \cdot v \cdot q = 0.10 \text{ cm} \leq 0.005 \cdot h_s = 2 \text{ cm} \quad (20)$$

$$dr \cdot v \cdot q = 0.21 \text{ cm} \leq 0.005 \cdot h_s = 2 \text{ cm} \quad (21)$$

4.2. Reinforcement for Concrete Elements

According to the design results [16], the reinforced concrete elements dimensions and reinforcements are described in Table 2. A_s is the longitudinal reinforcement in concrete slender columns, belts, columns and beams [16]. The bars are seen in each figure as black circles and the diameter (Φ) of bars is 16mm.

Table 2 Concrete elements reinforcement

Column 30x50 As → 8Φ16	Slender column 30x30 As → 4Φ16	Column 30x40 As → 6Φ16	Column 30x60 As → 12Φ16	Beams and belts 30x40 As → 8Φ16

4.3. Confined Masonry Walls Susceptible to Fogging

The most susceptible walls to loss of axial bearing capacity due to buckling are P1, P2 and P3 seen in Figure 1. They are situated on direction X. Those walls are not connected to other walls on the perpendicular direction.

Table 3 Confined masonry walls with high fogging probability

	P1			P2			P3		
	N_{Ed} [kN]	$\Phi_{i(m)}$	N_{Rd} [kN]	N_{Ed} [kN]	$\Phi_{i(m)}$	N_{Rd} [kN]	N_{Ed} [kN]	$\Phi_{i(m)}$	N_{Rd} [kN]
story 3	640	0.22	1368	637	0.23	1422	745	0.28	1931
story 2	1289	0.62	3857	1238	0.62	3835	1498	0.64	4413
story 1	1918	0.66	4106	1909	0.86	5319	2231	0.7	4827

5. Building Nonlinear Analysis

Two pushover cases (PX and PY) are used for the building's nonlinear analysis. Each case determines the building's stresses and plastic hinges development on an orthogonal direction.

5.1. Plastic Hinges in Final Stages

Figures 4 and 5 show the final stages of development for plastic hinges on both directions. The color code is the following: B (pink) means the plastic hinge is formed, IO (blue) is for immediate occupancy, LS (light blue) is for life safety, CP (green) is for collapse prevention, C (yellow) means the plastic hinge reaches the limit and the element gives out, D (brown) means the load was redistributed and E (red) means collapse. Those colors are seen in the figures that show the pushover analysis last steps. For both pushover cases, the analysis ends when plastic hinges reach stage D. For PX, there are fewer hinges in stage D. Hinges in stage D are formed at the ends of walls coupling beams. Figures 4 and 5 show the plastic hinges but also the masonry walls as translucent light grey. This is done in order for the figures to be more complete and the masonry walls not to obstruct the plastic hinges visibility.

5.2. Pushover Diagrams

The pushover diagrams show a stiff behavior on both directions. There are 24 steps for the nonlinear analysis on direction X and

21 steps on direction Y. Some steps are far from each other and some are very close, so they look overlapped. The rigidities remain the same until the structure becomes a mechanism. The structure's rigidity (base force/top displacement ratio) on direction X is 3806501kN/m, greater than 2269242kN/m for direction Y. The building reaches a higher displacement (143 mm) and a higher base force (545091kN) for direction X.

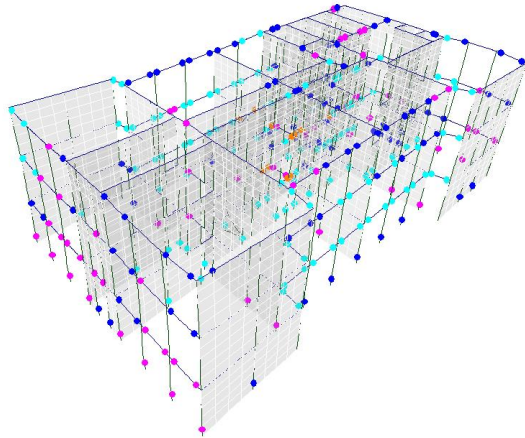


Figure 4 Plastic hinges for case PX at step 24

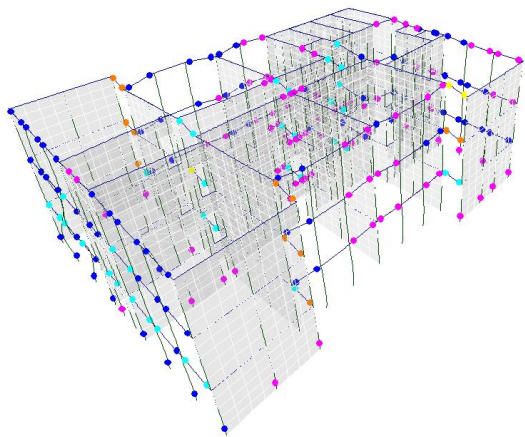


Figure 5 Plastic hinges for case PY at step 21

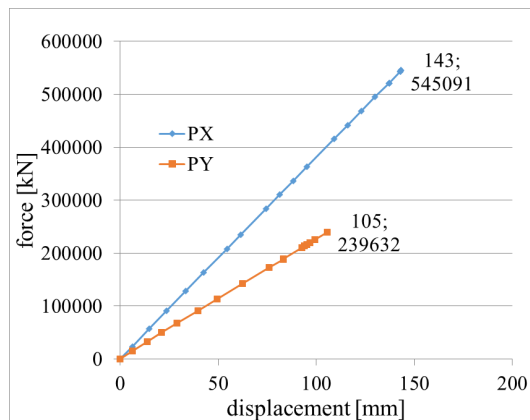


Figure 6 Pushover curves for PX and PY

5.3. Walls Stresses

The maximum value for each stress is written after the analysis step number. The highest values are reached at joints between

beams and walls as the stress is transmitted between those elements. There are also increased stress values at the walls corners on the same direction as the pushover case, one corner is crushed and the other is stretched. The maximum value for each stress is written after the analysis step number.

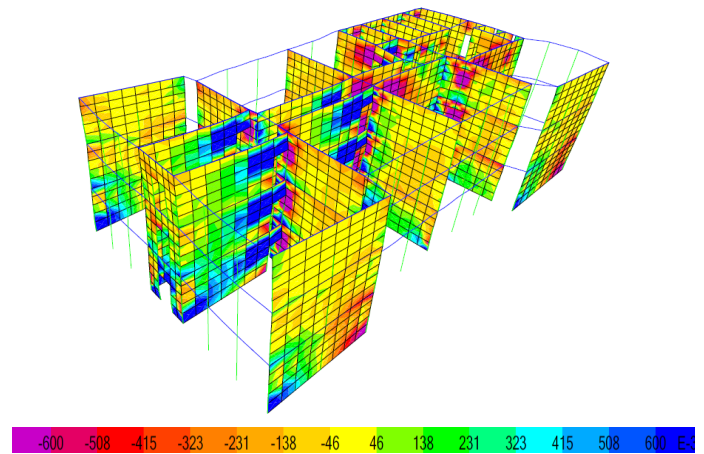


Figure 7 PX step 4 $\sigma_x=0.6\text{N/mm}^2$

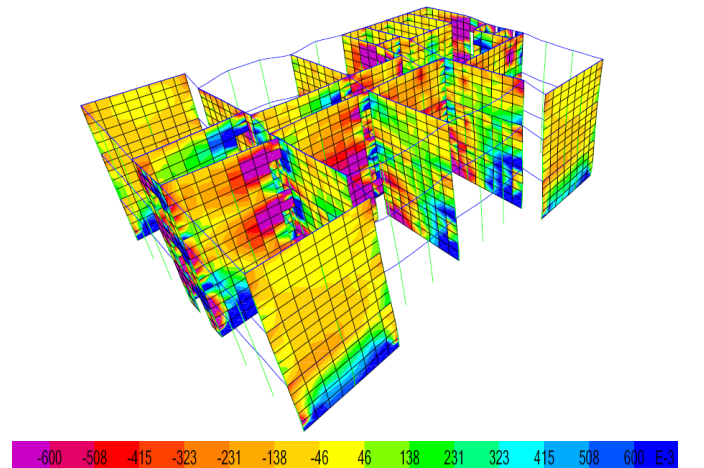


Figure 8 PY step 6 $\sigma_x=0.6\text{N/mm}^2$

For walls perpendicular to the pushover case there is an increase of stress at the walls bottoms. This means that the walls are subjected to fogging, as they are very slender.

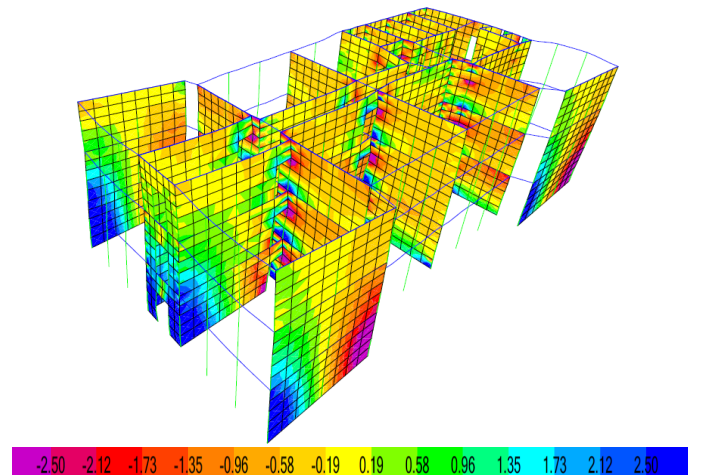


Figure 9 PX step 4 $\sigma_z=2.5\text{N/mm}^2$

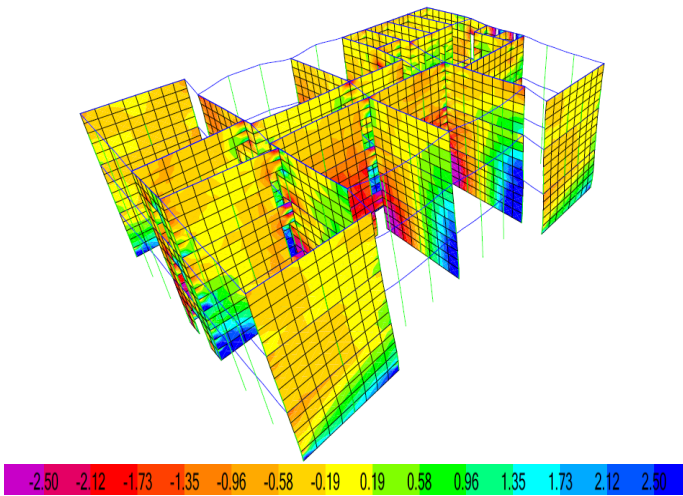


Figure 10 PY step 5 $\sigma_z=2.5\text{N/mm}^2$

Stresses σ_z clearly surpass the strengths f_{td} at steps 4 and 5. The crushing and stretching at the walls bottoms is more visible for σ_z . The wall fogging is also present here.

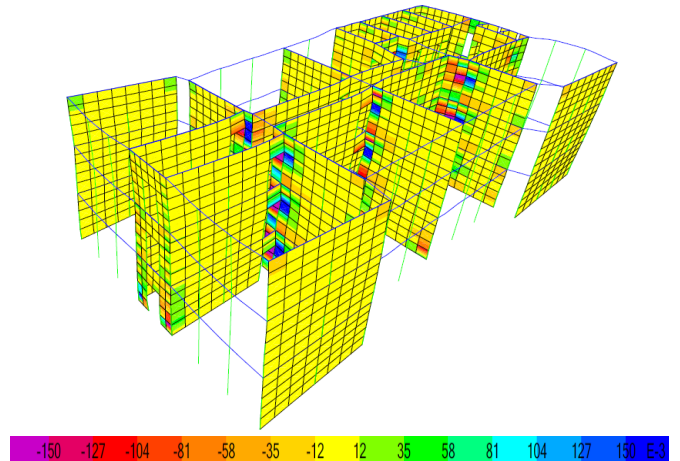


Figure 14 PY step 12 $\tau_{xy}=0.15\text{N/mm}^2$

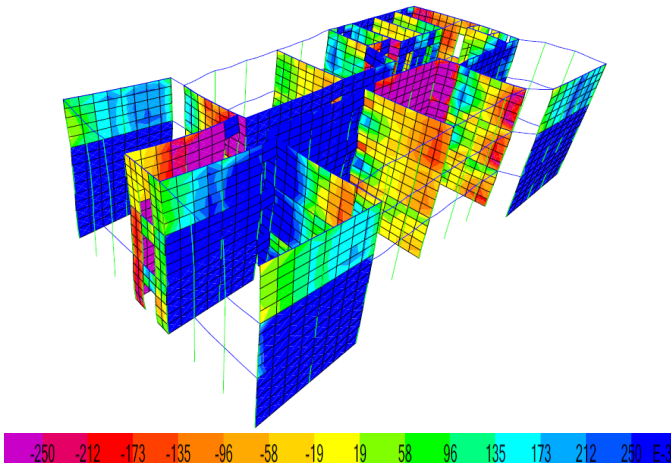


Figure 11 PX step 1 $\tau_{xz}=0.25\text{N/mm}^2$

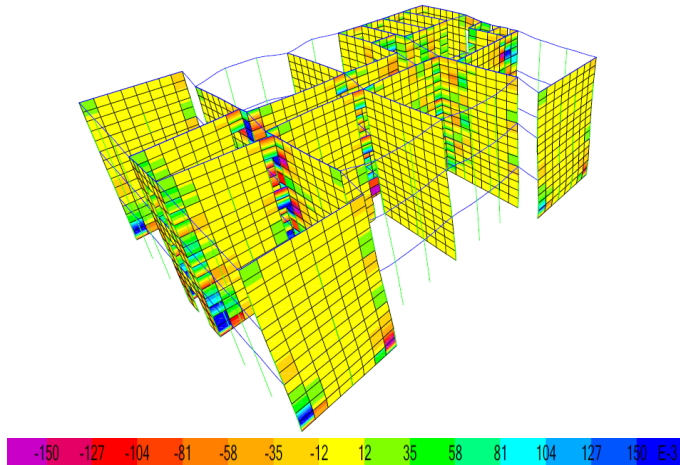


Figure 15 PX step 5 $\tau_{yz}=0.26\text{N/mm}^2$

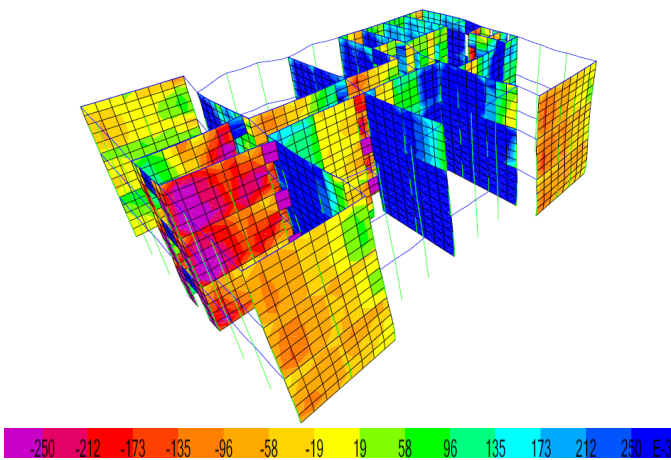
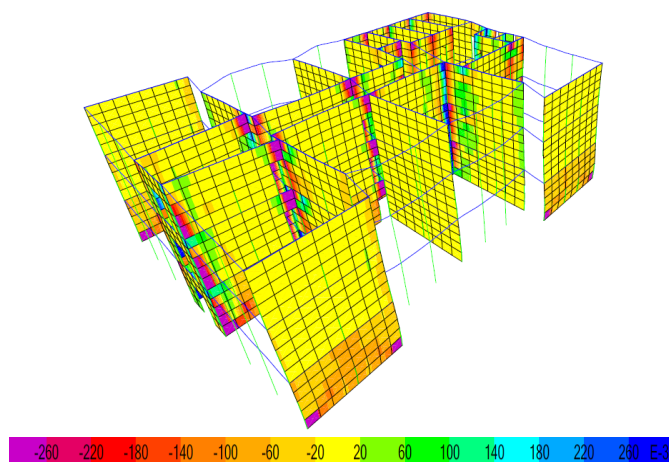


Figure 12 PY step 1 $\tau_{xz}=0.25\text{N/mm}^2$

Stresses τ_{xz} are greater than $f_{vd,1}$ from step 1 of the analysis. τ_{xz} is taken mostly by the walls on the stress direction. The stress is also transmitted to the walls connected perpendicular to

Figure 16 PY step 13 $\tau_{yz}=0.26\text{N/mm}^2$

6. Discussions

The building has a slender structure and the walls are made of a low strength material. There are 3 walls susceptible to fogging as they are developed on one direction only. The reinforced concrete elements are placed in accordance with the masonry code in force [18]. Additional reinforced concrete bands or lintels and slender columns may be provided in walls. In this latter case, walls would behave less slender. The walls fogging phenomenon would be less visible, as the confined masonry walls would behave more like concrete walls.

The building's lateral displacements in the elastic stage are lower than the maximum values accepted by [18]. The reinforcement needed for beams, columns, bands and slender columns is within the range demanded by [18]. This means the efforts in concrete elements are within the desired range for the elastic state.

The nonlinear analysis ends when plastic hinges reach stage D for both pushover cases PX and PY. There are more hinges in stage D for PY. Hinges in stage D are seen at the ends of walls coupling beams. The pushover diagrams show a higher rigidity on direction X. For case PX, the graphic has a higher slope. This slope is calculated as the base force/ top displacement ratio. This ratio is called rigidity. Both diagrams show the same rigidity, as both their slopes remain unchanged until the analysis ends. The analysis ends when the structure is turned into a mechanism by the plastic hinges giving in. The structure gives in at a lower base force and displacement on direction Y. This is explained by the 3 walls susceptible to buckling on that direction. All masonry walls crack before the plastic mechanism is reached. The fogging phenomenon is seen very clearly for stresses σ_x and σ_z . All masonry walls get local damage from crushing at the bottoms before the structure turns into a mechanism.

7. Conclusions

The building can bear both gravity and seismic loads, although it has a slender structure and the walls are made of a low strength material.

Lateral displacements generated by the seismic combination are allowed by the code in force.

Confined masonry walls developed only on one direction are susceptible of fogging.

In the nonlinear stage the building remains stiff until it turns into a plastic mechanism.

Both pushover diagrams PX and PY are straight lines, so the structure's rigidity is maintained the same until it collapses.

The structure reaches lower stress values in walls on direction Y.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] C. Cornado, J.R. Rosell, J. Leiva, C. Diaz, "Experimental study of brick masonry walls subjected to eccentric and axial load" RILEM International Conference on Materials, Systems and Structures in Civil Engineering Conference segment on Historical Masonry Technical University of Denmark, Lyngby, Denmark p 33- 40, 2016. www.rilem.net/publications/proceedings-500218
- [2] C. Glock, C. A. Graubner, "Design of slender unreinforced masonry walls" 13th International Brick and Block Conference Amsterdam, 2004. No 3 www.hms.civil.uminho.pt/ibmac/2004/
- [3] J.J. Perez-Gavilan, L.E. Flores, A. Manzano, "A new shear strength design formula for confined masonry walls: proposal to the Mexican code" Tenth U.S. National Conference of Earthquake Engineering Frontiers of Earthquake Engineering Anchorage, Alaska, 2014. July 21-25 <https://www.eeri.org/products-page/national-conference-on-earthquake-engineering/10th-u-s-national-conference-on-earthquake-engineering-frontiers-of-earthquake-engineering-proceedings-thumb-drive/>
- [4] M. Dhanasekar, "Shear in reinforced and unreinforced masonry: response, design and construction" The Twelfth East Asia Pacific Conference on Structural Engineering and Construction , Elsevier Procedia Engineering Vol 14. p 2069- 2076, 2011. doi:10.1016/j.proeng.2011.07.260
- [5] A. Marinilli, E. Castilla, "Experimental evaluation of confined masonry walls with several confining columns" 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada paper No. 2129, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [6] K. Yoshimura, K. Kikuchi, M. Kuroki, H. Nōkana, K. Tae Kim, R. Wangdi, A. Oshikata, "Experimental study for developing higher seismic performance of brick masonry walls" 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada; 2004 paper No. 1597, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [7] A. Chourasia, S.K. Bhattacharyya, P K. Bhargava, N. M. Bhandari "Influential aspects on seismic performance of confined masonry construction" Scientific research An academic publisher Natural science Vol. 5. Nr. 8. A1, 2013. <http://dx.doi.org/10.4236/ns.2013.58A1007> p56-62.
- [8] D. Liu, M. Wang, "Masonry structures confined with concrete beams and columns" 12th World conference on Earthquake Engineering, Auckland, New Zealand, 2000. <http://www.worldcat.org/title/12wcee-2000-12th-world-conference-on-earthquake-engineering-auckland-new-zealand-sunday-30-january-friday-4-february-2000/>
- [9] A. Alexandris, E. Protopapa, I. Psycharis "Collapse Mechanisms of masonry buildings derived by the distinct element method" 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada paper No. 548, 2004. www.researchgate.net/publication/264850141_Collapse_mechanisms_of_masonry_buildings_derived_by_the_distinct_element_method
- [10] W. Wijaya, D. Kusumastuti, M. Suarjana, R. Rildova, K. Pribadi "Experimental Study on Wall-Frame Connection of Confined Masonry Wall" The Twelfth East Asia -Pacific Conference on Structural Engineering and Construction Elsevier Procedia Engineering 14 p 2094-2102, 2011 doi:10.1016/j.proeng.2011.07.263
- [11] P. Naik, S. Annigeri, "Performance evaluation of 9 story RC building located in North Goa" 11th International Symposium on plasticity and Impact Mechanics, Implast 2016 Elsevier Procedia Engineering 173, p 1841-1846, 2017. doi:10.1016/j.proeng.2016.12.231
- [12] CEN EN 1996-1-1-2006 Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures, 2006.
- [13] CEN EN 1991-1-1-2004 Eurocode 1: Actions on structures - Part 1-1: General actions- Densities, self-weight, imposed loads for buildings, 2004.
- [14] CEN EN 1990-2004 Eurocode 0: Basics of structural design, 2004
- [15] CEN EN 1991-1-3-2005 Eurocode 1: Actions on structures - Part 1-3: General actions- Snow loads, 2005
- [16] CEN EN 1992-1-1-2004 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, 2004.
- [17] CEN EN 1998-1-2004 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, 2004.
- [18] P100-1/2013 Seismic design code – Part 1- General rules for buildings, 2013.

Masonry Walls Behavior in Predominant Frames Structures

Sorina Constantinescu*

Technical University of Construction Bucharest, Department of Civil Engineering, ZIP Code 01171, Romania

ARTICLE INFO

Article history:

Received: 28 July, 2018

Accepted: 10 August, 2018

Online: 14 November, 2018

Keywords :

Rigidity center

Masonry stresses

Plastic mechanism

ABSTRACT

This was a study on the behavior of a confined masonry bearing wall in a medium height dual building. This wall had to be placed at one corner of the building. It had to be a masonry wall, not to be too stiff and drag the rigidity center too far from the building's center. The structure's stiffness was also to be analyzed by using a concrete wall instead of the masonry one, as an alternative solution. This showed the importance of using a masonry wall. The dual structure contained only one other wall, made of reinforced concrete. The 2 bearing walls bore most of the shear force from seismic loads, because they were the stiffest load bearing elements in the structure. It was interesting to see if the masonry wall could bear these loads. The structure was unusual, as it contained frames, a concrete and a masonry wall. These elements behave differently. The structure was analyzed for both the elastic and plastic stage. The loadbearing elements stiffness, the stresses development and structure failure mechanism were studied for both solutions. The results showed it is appropriate to use a masonry wall at the corner. This wall can bear the loads it is subjected to.

1. Introduction

The present paper studies the behavior of a dual medium height building. It will be built in Bucharest, Romania. This structure contains mainly frames. There are also two bearing walls. One is placed close to the center and the other is in one corner. The first is a reinforced concrete wall, while the latter can either be made of confined masonry or reinforced concrete. The first solution has the advantage of keeping the center of rigidity closer to the building's center. A confined masonry wall does not have such a great rigidity. Although it is placed in a corner, it can't drag the rigidity center too far from the reinforced concrete wall. The second solution has the advantage of having two reinforced concrete walls that can bear the horizontal loads easier, but it is possible for the stiffness center to be shifted more towards the corner. Both solutions will be studied here. If both walls and frames are present, the lateral-force resisting system is normally provided by the walls, since they are much stiffer than the column frames [1]. Load bearing capacity of masonry panels is determined mainly by the stress distribution shape. The load-deformation pattern depends on the material properties of masonry bricks and mortar [2] and also masonry strength increases as the bricks dimensions decrease [3]. Laboratory tests prove that both confined masonry walls and masonry infill panels

show diagonal cracks when subjected to horizontal loads. For confined masonry, the cracks are more evenly distributed on the masonry panel [4]. Confined masonry, known to have performed well in moderate earthquakes, can be regarded as a form of a partially reinforced masonry. Masonry walls are thought to fail through diagonal shear [5]. Confined masonry structures show greater lateral strength and ductility compared to plain masonry structures. In these systems the majority of gravity and shear loads are taken by the masonry panels [6]. If masonry walls reinforcement is used, it helps the masonry to work together with the confining elements [7]. The stiffness for columns and walls will be calculated to see the difference between them. It is interesting to see if the masonry walls, that are only 0.3% of the building area can bear the loads, particularly the lateral loads they are subjected to. It is also important to see how the structure's failing mechanism occurs. Medium rise reinforced concrete walls show a good seismic behavior for different earthquake patterns. Buildings with slender walls may also show important ductility. Plastic hinges mostly develop at the beams ends [8]. Lack of symmetry may cause undesired seismic behavior for a structure [9]. For framed buildings, the beams and columns bending provides the resistance to lateral forces. Nonlinear static pushover analysis is useful to evaluate the real strength for buildings [10]. It is important to study a building's global seismic response in terms of capacity curve and plastic hinges location and

*Sorina Constantinescu, 0742265890, Email: sorina.constantinescu@yahoo.com

development [11]. The codes in force used to design the building are [12–18].

2. Structure Description

The floor plan is presented in Figure 1. The 3D building image is seen in Figure 2. The beams are blue, columns are green, reinforced concrete walls are dark gray, confined masonry walls are red and slabs are light gray. The concrete wall is composed of 2 piers (walls) on direction X (P4 and P6) and one on direction Y (P5).

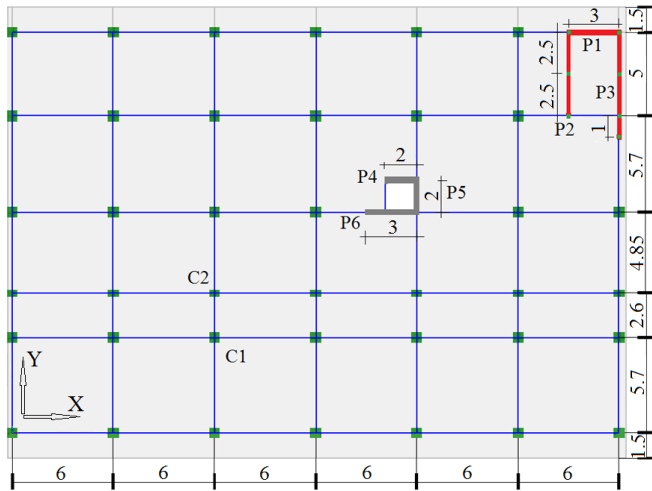


Figure 1 Story plan (all dimensions are in m)

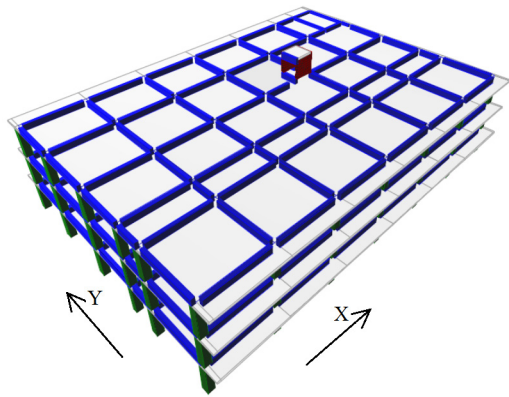


Figure 2 Building 3D image

The confined masonry wall contains one pier (wall) (P1) developed on direction X and 2 piers (walls) developed on direction Y (P2 and P3). The bearing capacity and stresses will be evaluated for all 3 piers (P1, P2 and P3) separately, for the plastic state. The confined masonry wall has a greater stiffness compared to the columns, so it will bear a higher amount of seismic force. Confined masonry walls can be designed with more confining slender columns and more vertical and horizontal reinforcement, so that they can bear higher loads. This can be done up to a point, as the slender columns reinforcement can't surpass a certain limit, and the horizontal bars can only be placed in the horizontal gaps between the brick rows. It is not desired to use too many slender columns as this would make the masonry wall behave more like a concrete wall and move the mass and stiffness center too close to the building's edge. Walls P1, P2, P3, P4, P5 and P6 are seen in Figure 3. The hatched areas are slender columns for P1, P2 and

P3. P4, P5 and P6 are completely hatched because they are made of reinforced concrete. The walls are 3 stories high (9m). The dimensions in Figure 3 are in cm.

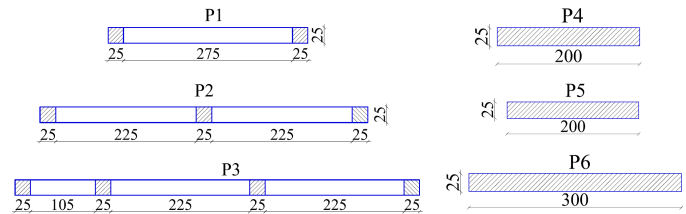


Figure 3 Cross sections for walls P1, P2, P3, P4, P5 and P6

The building will have an important seismic load to bear, as it is built in a high seismic area ($a_g=0.30g$, g is the gravity acceleration). The software used for analysis is ETABS 2016.

3. Theory Elements Used in Paper

3.1. General Material Characteristics

To study the behavior of the confined masonry walls, they will have to be designed to bear the vertical and horizontal loads they are subjected to. This design will be done using the seismic load combination: 1.0·permanent loads +0.4·variable loads+ 1.0·seismic loads. The concrete used is C20/25, with elasticity modulus $E_c=30000\text{N/mm}^2$ and reinforcement bars are S355 with elasticity modulus $E_s=210000\text{N/mm}^2$ [16]. The bricks for the masonry wall are full bricks 240·115·63 (mm) with $E_M=4400\text{N/mm}^2$, standard strength $f_b=10\text{N/mm}^2$ and mortar M10 [12]. The walls stresses analyzed are: σ_x , σ_z , τ_{xz} , τ_{xy} and τ_{yz} . They are compared to the design masonry strengths that are as follows: design horizontal (f_{dh}) and vertical (f_d) compression strengths, design shear strength for horizontal direction ($f_{vd,1}$) and design strengths for horizontal and vertical stresses perpendicular to the wall (f_{xd1} and f_{xd2}) [12]. They are calculated using their corresponding characteristic masonry strengths f_{kh} , f_k , $f_{vk,0}$, f_{xk1} and f_{xk2} , the insurance factor γ_M and the unitary vertical stress σ_d [12]. Concrete design compression strength is f_{cd} , calculated using the characteristic strength f_{ck} and steel design strength f_{yd} is calculated from the characteristic value f_{yk} [16].

$$f_{dh} = f_{kh}/\gamma_M = 2.09/1.9 = 1.1 \text{ N/mm}^2 \quad (1)$$

$$f_d = f_k/\gamma_M = 4.4/1.9 = 2.31 \text{ N/mm}^2 \quad (2)$$

$$f_{vd,1} = f_{vk,0}/\gamma_M + 0.4 \cdot \sigma_d = 0.3/1.9 + 0.4 \cdot 0.1 = 0.2 \text{ N/mm}^2 \quad (3)$$

$$f_{xd1} = f_{xk1}/\gamma_M = 0.24/1.9 = 0.126 \text{ N/mm}^2 \quad (4)$$

$$f_{xd2} = f_{xk2}/\gamma_M = 0.48/1.9 = 0.252 \text{ N/mm}^2 \quad (5)$$

$$f_{cd} = f_{ck}/\gamma_M = 20/1.5 = 13.3 \text{ N/mm}^2 \quad (6)$$

$$f_{yd} = f_{yk}/\gamma_M = 355/1.15 = 308 \text{ N/mm}^2 \quad (7)$$

3.2. Seismic Force Evaluation

The seismic action is introduced by the seismic coefficient c_s . The base force F_b is evaluated according to [17, 18]. $\gamma_{1,e} = 1.2$ is the building's importance-exposure coefficient, $\beta_0 = 2.5$ is the

maximum value of the elastic spectrum and q is the structure's behavior factor, $q=3.5 \cdot \alpha_u/\alpha_1=3.5 \cdot 1.35$, α_u/α_1 = the base shear force value for the failing mechanism/the base shear force value for the first plastic hinge, m = building's mass. $\lambda = 0.85$ as this is a 3 stories building, $a_g = 0.30g$ (because of the building's location), G = building's weight.

$$F_b = \gamma_{1,c} \cdot \beta_0 \cdot a_g/q \cdot m \cdot \lambda = c_s \cdot G = 0.17 \cdot G \quad [kN] \quad (8)$$

3.3. Confined Masonry Wall Design Theory Elements

M_{Rd} associated to N_{Ed} is the walls bearing bending moment associated to design axial force N_{Ed} . M_{Rd} is calculated using [12]. Wall's cross section compressed area is named A_c . l_s is the distance between the edge slender columns centers. y_c is the distance between the wall's weight center and A_c weight center. A_s is the reinforcement area in the slender columns. $b = t \cdot f_{cd}/f_d$. $t = 25cm$ is the wall's thickness.

$$M_{Rd} = M_{Rd(M)} + M_{Rd(As)} \quad [kNm] \quad (9)$$

$$A_c = N_{Ed}/(0.85 \cdot f_d) \quad [mm^2] \quad (10)$$

$$M_{Rd} = N_{Ed} \cdot y_c + A_s \cdot l_s \quad [kNm] \quad (11)$$

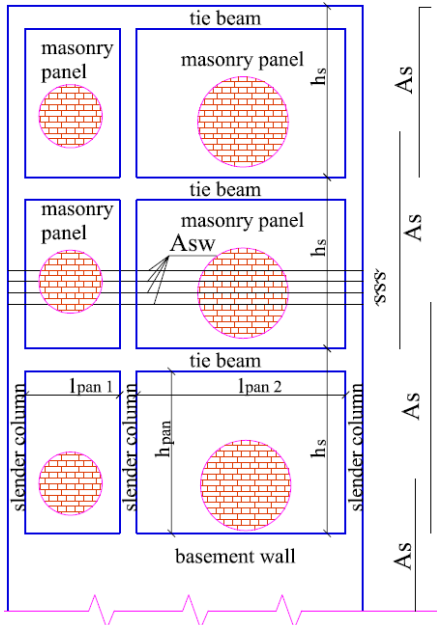


Figure 4 Confined masonry wall view

The confined masonry walls bearing shear force is named V_{Rd} . It is calculated using [12]. V_{Rd1} is the bearing shear force taken by the masonry panels. V_{Rd2} is the bearing horizontal shear force taken by the slender column at the compressed wall edge. V_{Rd3} is the shear bearing force taken by the horizontal reinforcement bars A_{sw} .

$$V_{Rd} = V_{Rd1} + V_{Rd2} + V_{Rd3} \quad [kN] \quad (12)$$

$$V_{Rd1} = 0.4 \cdot N_{Ed} + 0.8 \cdot V_{Ed} \cdot h_{pan}/l_{pan} \quad [kN] \quad (13)$$

$$f_{v,d,0} = f_{vk,0}/\gamma_M = 0.30/1.9 = 0.158 \text{ N/mm}^2 \quad (14)$$

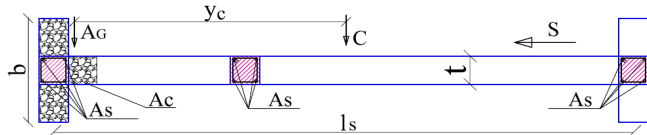


Figure 5 Confined masonry wall horizontal cross section

Figures 4 and 5 show a confined masonry wall's view and cross section. S shows the seismic action direction. l_s is the distance between the edge slender columns centers.

$$V_{Rd2} = \lambda_c \cdot A_s \cdot f_{yd} \quad [kN] \quad (15)$$

$$V_{Rd3} = 0.8 \cdot l_w \cdot A_{sw} \cdot f_{yd}/s \quad [kN] \quad (16)$$

A_{sw} is the reinforcement area in the horizontal bricks joints. s is the vertical distance between two horizontal reinforced joints. λ_c is the reinforcement participation factor.

3.4. Frames Design Theory Elements

Bending reinforcement in beams is designed according to M_{Ed} (bending moment from the seismic load combination) [16, 18].

$$M_{Ed} = b \cdot \lambda_x \cdot f_{cd} \cdot (d - \lambda_x/2) \quad [kNm] \quad (17)$$

$$M_{Ed} = A_s \cdot f_{yd} \cdot z \quad [kNm] \quad (18)$$

$$m = M_{Ed}/(b \cdot d^2 \cdot f_{cd}) \quad (19)$$

$$z = d - \lambda_x/2 = d - d \cdot (1 - (1 - 2m)^{0.5})/2 \quad [mm] \quad (20)$$

$$A_{s,min} = \min\{0.26 \cdot f_{ctm}/f_{yk} \cdot b \cdot d; 0.0013 \cdot b \cdot d\} \quad (21)$$

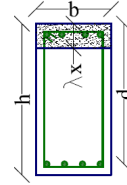


Figure 6 Beam section

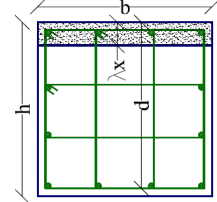


Figure 7 Column section

λ_x is the beam section compressed area height [16, 18]. d is the distance seen in Figures 6 and 7. The minimum reinforcement area A_s for beams is $A_{s,min}$. $f_{ctm} = 2.6 \text{ N/mm}^2$ is the concrete tensile strength medium value for C20/25. M_{Rc} is the bearing bending moment in columns. It is calculated according to [16, 18]. $\gamma_{Rd} = 1.2$ is the steel stiffening factor for DCM (medium ductility buildings), ΣM_{Rc} and ΣM_{Rb} are the sums of bearing bending moments in columns and beams near a frame joint. For columns, the longitudinal reinforcement coefficient minimum value is $\rho_{min} = 0.01$ and the maximum value is $\rho_{max} = 0.04$. N_{Ed} is the axial force in columns. A_s will be determined from (25) if $\lambda_x < 2 \cdot a_s$, and from (26), if $\lambda_x \geq 2 \cdot a_s$. Here $a_s = 45 \text{ mm}$.

$$\Sigma M_{Rc} \geq \gamma_{Rd} \cdot \Sigma M_{Rb} \quad [kNm] \quad (22)$$

$$\rho = A_s/(b \cdot d) \quad (23)$$

$$x = N_{Ed}/(b \cdot \lambda \cdot f_{cd}) \quad [mm] \quad (24)$$

$$A_s = [M_{Ed} - N_{Ed}(d - a_s)]/[f_{yd} \cdot (d - a_s)] \quad [mm^2] \quad (25)$$

$$A_s = [M_{Ed} + N_{Ed}(d - a_s) - 2 \cdot b \cdot \lambda_x \cdot f_{cd}(d - \lambda_x/2)]/[f_{yd}(d - a_s)] \quad (26)$$

4. Elastic Analysis Results

4.1. Walls and Frames Efforts

For direction X, the bending moments in columns show increased values at the base on the structural lines where no walls are present and at the upper stories on the lines where there are walls. The

effect is increased if they are reinforced concrete walls. The high bending moments at the columns base are generated by the seismic force that is to be taken by the frames only. The high bending moments at the upper stories are created by the energy dissipation mechanism that is done by the frames. On direction Y, bending moments visibly reach the highest values in columns farthest from the walls. This happens because in those areas frames have to withstand the horizontal loads without help from the walls. These observations are made according to Figures 8 and 9.

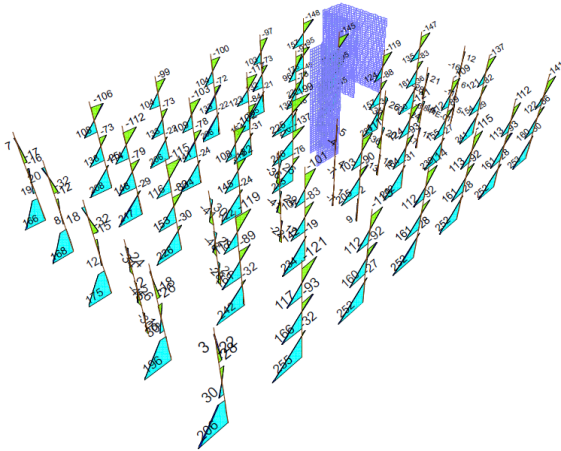


Figure 8 M_{Ed} from the seismic load combination on X direction

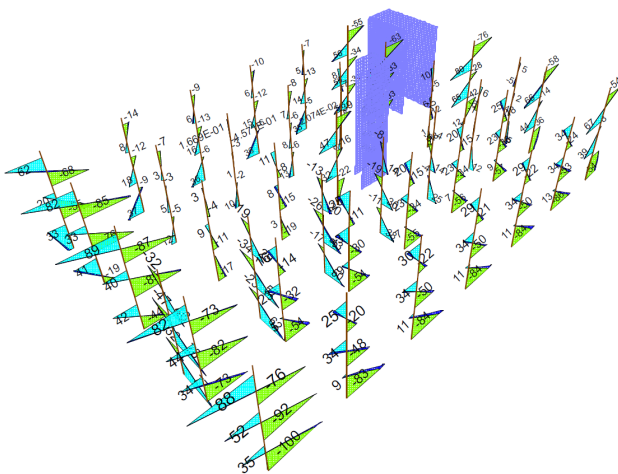


Figure 9 M_{Ed} from the seismic load combination on Y direction

According to the design results, the reinforced concrete elements dimensions and reinforcements are described in Table 1. As is the longitudinal reinforcement area. The bars are seen in each figure as black discs and the diameter (Φ) of bars is in mm. Walls P1, P2, P3, P4 P5 and P6 are seen in detail in Figure 3. For the bearing efforts analysis, P1, P2 and P3 will be considered working separately. P1 is working on X direction, while P2 and P3 are working on Y direction.

When determining the walls rigidities, P1, P2 and P3 are considered working together as one wall. This wall is developed on both X and Y directions. This wall is named **P1-3**.

Table 1 Concrete elements dimensions (in cm) and reinforcements

Beam 30x60 As →4 Φ 20 up and down	Beam 25x50 As →4 Φ 20 up and down	Tie beam 25x30 As →4 Φ 16
Column C1 60x60 As →12 Φ 22	Column C2 40x60 As →10 Φ 22	Slender column 25x25 As →4 Φ 16

The efforts in masonry walls are written in Table 2. P3 does not need horizontal reinforcement in the masonry panels. P1 behaves like a slender wall, as the design shear force is greater at the second story.

Table 2 Masonry walls efforts

	P1		
	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]
story 3	170	307	997
story 2	424	977	1332
story 1	718	1676	1679
	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw} []
story 3	383	464	2 Φ 8/30
story 2	563	675	2 Φ 10/30
story 1	471	683	2 Φ 8/30
	P2		
	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]
story 3	326	572	1915
story 2	673	1539	1246
story 1	1021	2555	3072
	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw} []
story 3	511	526	2 Φ 8/30
story 2	902	914	2 Φ 8/15
story 1	944	1130	2 Φ 8/15
	P3		
	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]
story 3	360	676	2757
story 2	710	2267	4575
story 1	1038	4364	5228
	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw} []
story 3	265	340	-
story 2	406	480	-
story 1	580	611	-

4.2. Walls Rigidities Values and Structure's Rigidity Center

The building's rigidity center will be determined according to the rigidities of all vertical load bearing elements. When determining the walls rigidities, P1, P2 and P3 are considered working together as one wall. This wall is developed on both X and Y directions. This wall is named **P1-3**, seen in Figure 10. The rigidity center for wall P1-3 is named RC P1-3 in Figure 12. When determining the walls rigidities, P4, P5 and P6 are considered working together as one wall. This wall is developed on both X and Y directions. This wall is named **P4-6**, seen in Figure 11. The rigidity center for wall

P4-6 is named RC P4-6 in Figure 12. If P1-3 is a masonry wall, the building's rigidity center is RC1, seen in Figure 12. If P1-3 is a concrete wall, the building's rigidity center is RC2, seen in Figure 12. Equation (27) is used to calculate walls rigidities [12].

$$R = 1 / [H^3 / (3 \cdot E_{CM} \cdot I) + k \cdot H / (G_{CM} \cdot A)] \quad [\text{kN/cm}] \quad (27)$$

$$G_{CM} = 0.4 \cdot E_{CM} \quad (28)$$

$$E_{CM} = (E_M \cdot I_M + E_C \cdot I_C) / (I_M + I_C) \quad [\text{kN/m}^2] \quad (29)$$

R is the wall's rigidity, H=9m is the total wall height, k=1.2 is a coefficient according to the wall's horizontal cross section shape. E_{CM} and G_{CM} are the longitudinal and respectively transversal elasticity modulus for confined masonry walls [12]. I is the wall's cross section moment of inertia and A is the wall's cross section area. I_M and I_C are the moments of inertia for the masonry cross section areas (white in Figure 10) and respectively concrete cross section areas (hatched in Figures 10 and 11). E_M and E_C are the longitudinal elasticity modulus for masonry and respectively concrete [12].

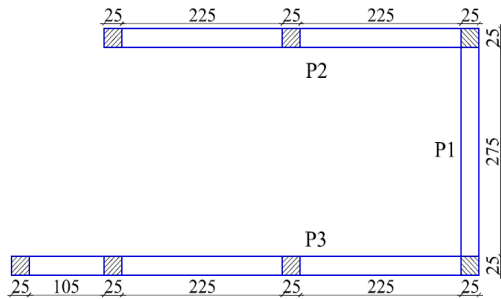


Figure 10 Confined masonry wall P1-3 (cross section)

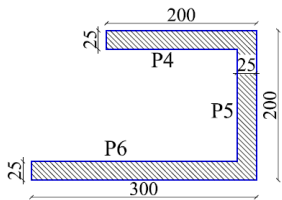


Figure 11 Reinforced concrete wall P4-6 (cross section)

Both reinforced concrete and masonry walls are 25cm thick, and their lengths are seen in Figure 10 for wall P1-3 made of confined masonry (M), and in Figure 11 for wall P4-6 made of reinforced concrete. Dimensions in Figures 10 and 11 are in cm.

Figure 12 shows the elements used in determining the structure's rigidity. The distances from the origin (0;0) to each center of the load bearing vertical element are drawn with blue arrows for direction X and red for Y. RC 1 is the structure's rigidity center if P1-3 is a confined masonry wall. RC 2 is the structure's rigidity center if P1-3 is a reinforced concrete wall. It is shown the way the rigidity center is shifted to the corner for the second case. There is an important impact on the rigidity center position given by the material used for P1-3. The rigidity center for the whole building is calculated with (30) for both orthogonal directions X and Y. In (30), i is the load bearing element number in the sum and d_i is the distance from the origin (0;0) to the element's rigidity center. Those distances are shown in Figure 12 (blue for direction X and red for direction Y).

$$R_b = (\sum R_i \cdot d_i) / \sum R_i \quad [\text{kN/cm}] \quad (30)$$

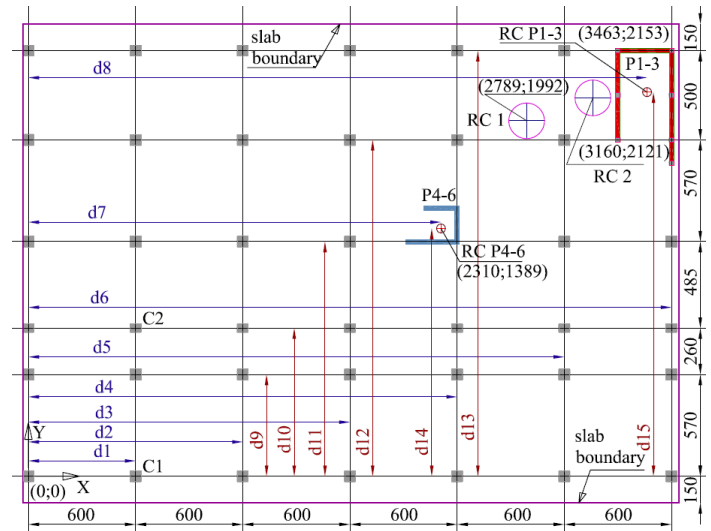


Figure 12 Building rigidity centers (dimensions are in cm)

The building's rigidity values on X and Y are comparable if wall P1-3 is made of confined masonry. If P1-3 is a reinforced concrete wall, the building's rigidity is increased less on X and more on Y. This is explainable because P1-3 is more developed on Y. This is valid only for the elastic stage.

Table 3 Walls and columns rigidities values

Wall	P1-3 on X (M)	P1-3 on Y (M)	P1-3 on X (C)	P1-3 on Y (C)	P4-6 on X	P4-6 on Y
R [kN/cm]	1887.9 3	3656.49	7137.6	11938. 97	1971.6	654.5
Column	C1 60x60 on X and Y		C2 40x60 on X		C2 40x60 on Y	
R [kN/cm]	13.289		8.859		3.945	
ΣR on X [kN/cm]	4271 if P1-3 is M			9520 if P1-3 is C		
ΣR on Y [kN/cm]	4670 if P1-3 is M			12953 if P1-3 is C		

In Table 3 **P1-3(M)** means P1-3 made of confined masonry. **P1-3(C)** means P1-3 made of reinforced concrete. **ΣR on X** is the whole building's rigidity on direction X. **ΣR on Y** is the whole building's rigidity on direction Y.

5. Elastic Stage Results

5.1. Pushover Diagrams

The 4 pushover cases used for the building's nonlinear analysis are PX, PY, PXC and PYC. PX and PY are used for the first solution (a concrete and a masonry wall), while PXC and PYC are used for the second solution (2 concrete walls). The pushover diagrams in Figure 13 are drawn for each case. The maximum base force reached is greater for PX. The maximum displacement values reached are close. The building is stiffer on direction X. This can be explained by the reinforced concrete wall extended more on direction X. Diagrams for cases PY and PXC overlap each other up to a point. This means the building has the same rigidity for those cases. Displacements reached for cases PXC and

PYC are smaller than those for PX and PY. This is because the building becomes stiffer if 2 concrete walls are used.

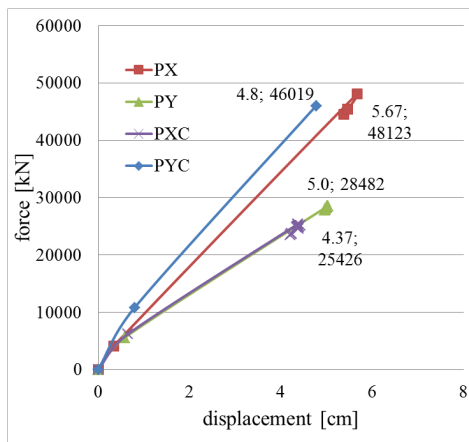


Figure 13 Pushover diagrams

5.2. Plastic Mechanism

Figures 14 and 15 show the final stages of plastic hinges development on both directions. For both cases, the last steps show the collapse of plastic hinges at columns bottoms.

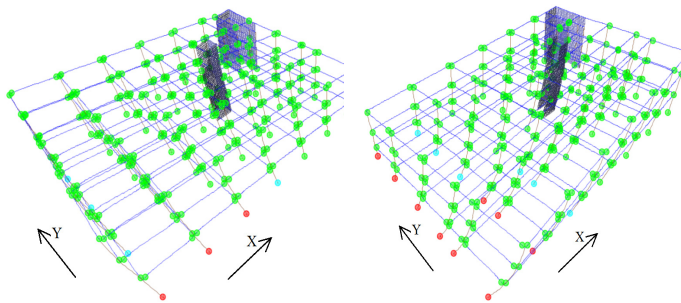


Figure 14 PX step 9

Figure 15 PY step 16

There are also hinges formed at the beams ends. Most developed hinges are present in columns placed farther from the walls area, particularly the hinges that reach the collapse stage. Columns closer to the walls have less horizontal loads to bear, as they are taken mostly by the stiffer elements. For case PY there are more plastic hinges that reach collapse and a large buckling phenomenon for the edge columns and the reinforced concrete wall. The structure has a lower stiffness for case PY in the plastic stage, as it is seen in Figure 13. The color code in Figures 14 and 15 is the following: **green** means the plastic hinge is formed, **light blue** means the plastic hinge reaches the limit and the element gives out, **pink** means the load was redistributed and **red** means collapse.

5.3. Masonry walls stresses

Walls P1, P2 and P3 are shown separately, to see the stress patterns more clearly. Stresses in each wall are shown for the case that creates the highest values. The case names are written in brackets. For P1 stresses σ_x show increased and alternating values at the intersections with beams, as seen in Figure 16.

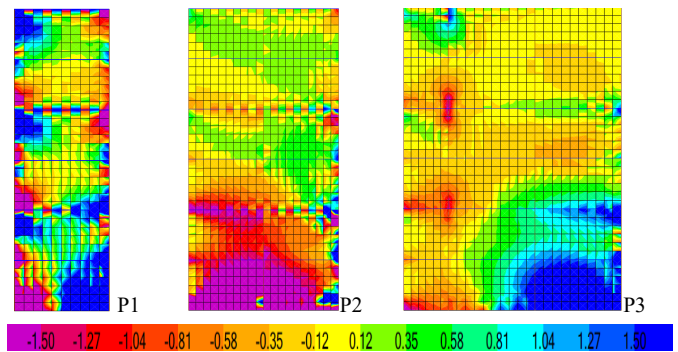


Figure 16 σ_x values at step 2 (case PX for P1 and case PY for P2 and P3)

This means the masonry is stretched below the tie beams (blue areas) and crushed above them and at the wall's bottom.

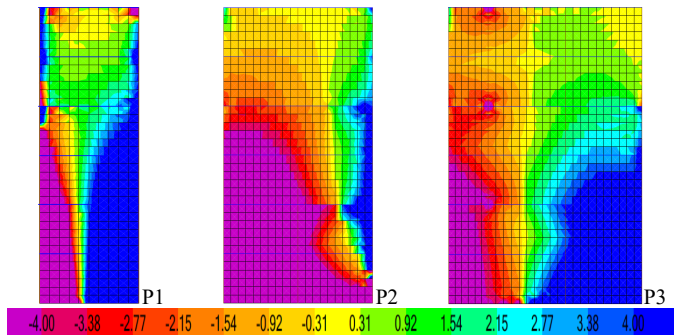


Figure 17 σ_z values at step 2 (case PX for P1 and case PY for P2 and P3)

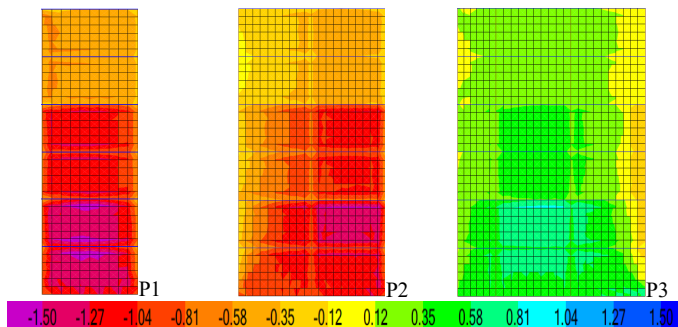


Figure 18 τ_{xz} values at step 1 (case PX for P1 and case PY for P2 and P3)

For P2 and P3 those stresses show increased values at the wall's bottoms, as seen in Figure 16. Also, for walls, the stresses increase around the tie beams, where the floors connect to the walls. Floors may transmit horizontal stresses to walls. Stresses σ_z reach the highest values at the walls bottoms and at the walls corners, as this is a vertical stress, but the walls are also subjected to horizontal loads. In Figure 17 it is seen that one corner is crushed and the other is stretched. Stress τ_{xz} surpass the masonry strengths from the first step of the analysis as seen in Figure 18. Stress values in walls P1 and P2 are smaller than in P3. Another observation is that the maximum stresses have different signs for P2 and P3. P2 is crushed and P3 is stretched. P3 is less affected, as it is longer.

Stresses τ_{xy} , in Figure 19, and τ_{yz} , in Figure 20, surpass the strengths from the second step of the analysis. The values are higher at the walls intersections, as these areas are stiffer. Stress values also increase at the beams and tie beams intersections with perpendicular walls. This is particularly seen for P3.

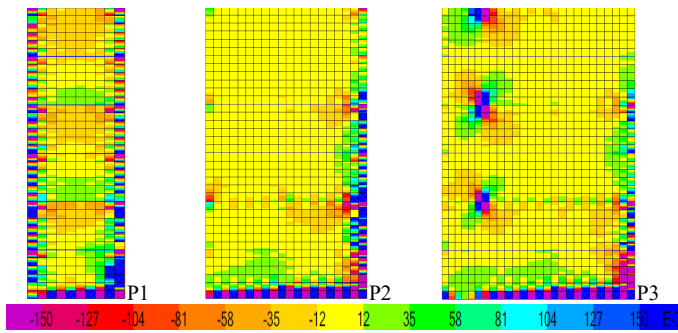


Figure 19 τ_{xy} values at step 2 (case PY for P1, P2 and P3)

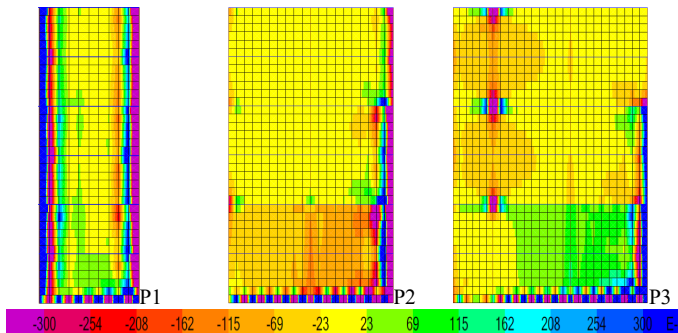


Figure 20 τ_{yz} values at step 2 (case PX for P1 and case PY for P2 and P3)

6. Conclusions

The confined masonry wall can bear the loads it is subjected to. The building's rigidity center is moved closer to the corner if a concrete wall is used instead of the masonry one. The failure mechanism is reached when plastic hinges reach the collapse stage at the columns bottoms. This solution is suitable for medium height buildings with dual structure, when it is necessary to place walls at one corner. It is mandatory, to check if the masonry walls can bear the loads they are subjected to.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] K. Leng, C. Chintanapakdee, T. Hayashikawa, "Seismic Shear Forces in Shear Walls of a medium Rise Building By Response Spectrum Analysis". Engineering Journal Volume 18 Issue 4, 2014. <http://dx.doi.org/10.4186/ej.2014.18.4.73>
- [2] C. Glock, C. A. Graubner, "Design of slender unreinforced masonry walls" 13th International Brick and Block Masonry Conference Amsterdam, 2004. www.hms.civil.uminho.pt/ibmac/2004/
- [3] A. Jaber, "Effect of Masonry Units Type and Concrete Grouting on Compressive Strength of Prisms" Eng. & Tech Journal.2010; Vol.28, No.13, 2010. <https://www.iasj.net/https://www.iasj.net/iasj?func=fulltext&aid=27774>
- [4] L. M. Abel-Hafez, A. E.Y. Abouelezz, E.F. Elzefary, "Behavior of masonry strengthened infilled reinforced concrete frames under in-plane load." HBRC Journal production and hosting by Elsevier 2015 11, p 213-223, 2015. <http://dx.doi.org/10.1016/j.hbrcj.2014.06.005>
- [5] M. Dhanasekar, "Shear in reinforced and unreinforced masonry: response, design and construction" in The 12th East Asia-Pacific Conference on Structural Engineering and Construction. Elsevier Procedia Engineering 14 p 2069-2076, 2011. doi:10.1016/j.proeng.2011.07.260
- [6] Y. Ouyang, H. J. Pam, S. H. Lo, Y. L. Wong, J. Li, "Preliminary study of Masonry -RC Hybrid Structure Behavior under Earthquake Loading" 15 World Conferences on Earthquake Engineering (WCEE) Lisboa, 2012. https://www.iitk.ac.in/nicee/wcee/fifteenth_conf_purtgal/
- [7] K. Yoshimura, K. Kikuchi, M. Kuroki, H. Nokana, K. Tae Kim, R. Wangdi, A. Oshikata, "Experimental study for developing higher seismic performance

- of brick masonry walls" in 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada, paper No. 1597, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [8] H. Akiyama, M. Teshigawara, H. Kuramoto, F. Kumazawa, Y. Inoue, K. Watanabe, "Development and structural design guideline for medium/high rise RC wall-frame structures with flat beams" in 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada, paper No. 2354, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [9] A. W. Hameed, "Failure modes for different structural types during an earthquake" IJCE Vol. 2, Issue 1, p47-56, 2013file:///C:/Users/User1/Documents/Articol/New%20folder/2-11-1360748911-6.%20IJCE%20-%20Failures%20Modes%20-%20Alaa%20W.%20Hameed.pdf
- [10] S. A. El-Betar, "Seismic performance of existing R.C. framed buildings" Elsevier Housing and Building National research center (HBRC) 13, p 171-180, 2015. <http://dx.doi.org/10.1016/j.hbrcj.2015.06.001>
- [11] P. Naik, S. Annigeri, "Performance evaluation of 9 story RC building located in North Goa" 11th International Symposium on plasticity and Impact Mechanics, Implast 2016 Elsevier Procedia Engineering 173, p 1841 -1846, 2016. doi:10.1016/j.proeng.2016.12.231
- [12] CEN EN 1996-1-1-2006 Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures, 2006.
- [13] CEN EN 1991-1-1-2004 Eurocode 1: Actions on structures - Part 1-1: General actions- Densities, self-weight, imposed loads for buildings, 2004.
- [14] CEN EN 1990-2004 Eurocode 0: Basics of structural design, 2004
- [15] CEN EN 1991-1-3-2005 Eurocode 1: Actions on structures - Part 1-3: General actions- Snow loads, 2005
- [16] CEN EN 1992-1-1-2004 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, 2004.
- [17] CEN EN 1998-1-2004 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, 2004.
- [18] P100-1/2013 Seismic design code – Part 1- General rules for buildings, 2013.

Impacts of Synchronous Generator Capability Curve on Systems Locational Marginal Price through a Convex Optimal Power Flow

Italo Fernandes*

IEEE Member, Electrical Engineering Department, ISL Wyden International College, 65071-380, Brazil

ARTICLE INFO

Article history:

Received: 31 July, 2018

Accepted: 10 August, 2018

Online: 14 November, 2018

Keywords :

Optimal power flow

Convex Relaxation

Second Order Cone

Programming Synchronous

Generator Capability Curve

ABSTRACT

This paper deals about an application of optimal power flows (OPF) constrained with synchronous generator capability curve for power market analysis (SGCC). OPF main features gather in its mathematical formulation non-convexity, non-linearity, and it shows to be a hard to solve optimization problem. In some operational scenarios, SGCC can limit power flows bringing the theoretical OPF situation to not be applied for real context on power plants. Thus, for electric market analysis solutions without those constraints could provide results that are not exact for systems Locational Marginal Prices (LMPs), messing with the power cost estimation. For this work SGCC includes limits of current and power, preserving the generator against overheating and immoderate mechanical stress. Properties of convexification brings to mathematical problems in general a good performance if compared to original problem and a simpler way of its formulation and solution. Thereby, OPF formulation is given as a second order cone programming (SOCP) approach handled by techniques of convex relaxation, with active power generators cost objective function. Numerical results are obtained in MATLAB® environment and applied in IEEE 14-bus test system. OPF results shows the good performance of the proposed methodology, whereas solutions will not violate SGCC limits constraints.

1. Introduction

The optimal power flow (FPO) was introduced by J. Carpentier, but it took a long time to become a useful algorithm that could be used as an improvement tool for electrical power systems. For planners and operators, power flow corresponds only to a snapshot of the current state of the network. Planning and operating requirements often require adjustments to the electrical parameters of the system according to a particular criterion [1].

The relaxation techniques allow modeling the problem into a simpler one, facilitating the solution and formulation. In this way, convex relaxation methods have attracted several researches with proposals to simplify and improve the performance of the FPO algorithms in the search for optimal solutions. Relaxation consists on using mathematical transformations to eliminate the power flow terms that introduce the nonconvex characteristic into the optimization problem. The inherent properties of the network are maintained, the output variables of the FPO must also be maintained (however they can be transformed), but the problem becomes more intelligible.

The set of limitations that compound synchronous generator realistic operational bounds is called SGCC. Synchronous machine working as generator base its main features in terminal voltage, field and armature current, power factor and efficiency. Those constraint are rarely take into consideration in OPF formulation, what is extremely dangerous for real operation scenarios leading the machine to operate on overheating conditions and providing inaccurate LMPs information for electric market dealers. Those real limits are extremely useful for network manager, as it will be possible to avoid situations that could offer risks to generator [2].

This work will present a market analysis in locational marginal prices through a non-conventional OPF method using a convex approach from SOCP relaxation, including in it constraints SGCC limits. Text is structured as follows. In the second section will be shown the nonlinear optimal power flow (ACOPF) formulation and it also introduce SOCP relaxation summed up. Section 3 involves the study and equations of SGCC. Simulation and analysis will be found in Section 4. Simulations were held in MATLAB® using its optimization tools for solving both: convex and non-linear OPF.

* Italo Fernandes Email: italo.fernandes@ieee.org

2. ACOPF Formulation and Relaxation

Optimal power flow is an important tool for planning and operation of system that in its main objective intends to minimize a function of interest referred as objective function and at the same time look for a feasible operation point for power flow equation. It is basically a constrained optimization problem. In its formulation OPF can include also some special devices connected to the grid, using its parameters as control variables, consequently increasing the options for optimization [2].

2.1. ACOPF Formulation

For original formulation, so called ACOPF, it should be considered a system with the following features:

- A number of lines NL ;
- A number of buses NB in wich NPV generators buses, NPQ load buses and normally one slack bus;
- Bus voltage, angle, active and reactive power respectively V_k , θ_k , P_k and Q_k
- Elements of Y admittance matrix

Minimizing a generation cost objection objective function $f(V, \theta)$ (used for numerical results), as: [2]

$$\min \quad f(V, \theta) \quad (1)$$

$$s. t \quad P_k^{sched} - P_k = 0 \quad (2)$$

$$Q_k^{sched} - Q_k = 0 \quad (3)$$

$$P_{min} \leq P_G \leq P_{max} \quad (4)$$

$$Q_{min} \leq Q_G \leq Q_{max} \quad (5)$$

$$V_{min} \leq V \leq V_{max} \quad (6)$$

$$\theta_{km}^{min} \leq \theta_{km} \leq \theta_{km}^{max} \quad (7)$$

$$S_{km} \leq S_{km}^{max} \quad (8)$$

where,

$$P_k = V_k \sum_{m \in k} V_m (G_{km} \cos \theta_{km} + B_{km} \sin \theta_{km}) \quad (9)$$

$$Q_k = V_k \sum_{m \in k} V_m (G_{km} \sin \theta_{km} - B_{km} \cos \theta_{km}) \quad (10)$$

For this formulation $V\theta$ bus voltage ($V_{sl} = 1 p.u$) and angle ($\theta_{sl} = 0^\circ$) constraints should be concerned. Network constraints are specified in (2) and (3) and its main purpose are find a feasible operation point for power flow. From (4) to (8) are handled systems operational limits, preventing OPF to find a solution that could be not safe for the system. Thus, OPF can minimize a the main function and in parallel return a result for network variables through power flow equations [3].

2.2. SOCP OPF Relaxation

Non-convexity source in ACOPF model lies on (9) and (10). Therefore, techniques of convexification can be used to relax those equations. References [4,5] show that a simple variable changing in voltage magnitude and phase will make the problem a convex one, as shown in (11)-(13).

$$R_{km} = V_k V_m \cos(\theta_k - \theta_m) \quad (11)$$

$$T_{km} = V_k V_m \sin(\theta_k - \theta_m) \quad (12)$$

$$u_k = V_k^2 \quad (13)$$

Changing R_{km} and T_{km} in (9) and (10), the follow expressions are reached:

$$P_k = G_{kk} u_k + \sum_{m \in k} (G_{km} R_{km} + B_{km} T_{km}) \quad (14)$$

$$Q_k = -G_{kk} u_k - \sum_{m \in k} (B_{km} R_{km} - G_{km} T_{km}) \quad (15)$$

Using trigonometric identity and manipulation on (11) and (12), it could be said that:

$$u_k u_m = R_{km}^2 + T_{km}^2 \quad (16)$$

$$\theta_k - \theta_m = \tan^{-1} \left(\frac{T_{km}}{R_{km}} \right) \quad (17)$$

Constraint (6) becomes:

$$u_{min} \leq u_k \leq u_{max} \quad (18)$$

where, $u_{min} = V_{min}^2$ and $u_{max} = V_{max}^2$.

Taking (11) and (12) it can be noticed that $R_{km} = R_{mk}$ and $T_{km} = -T_{mk}$.

Hermitian matrix W can be defined as [6,7]:

$$W = \begin{cases} W_{kk} = u_k^2, \forall k \in NB \\ W_{km} = R_{km} + jT_{km}, \forall km \in NL \\ W_{mk} = R_{km} - jT_{km}, \forall km \in NL \\ W_{km} = 0, \forall km \notin NL \end{cases} \quad (19)$$

Decomposing Hermitian matrix into NL submatrix (20) referred as W_{NL} ($2x2$) and making all of them to be semi-definite positive ($W_{NL} \geq 0$), it could be said that the problem is relaxed as a SOCP problem.

$$W_{NL} = \begin{bmatrix} W_{kk} & W_{km} \\ W_{mk} & W_{mm} \end{bmatrix} \quad (20)$$

3. Synchronous Generator Capability Curve

The capacitance curve (capacity curve or D-curve in other literature) is a selection of important curves for the actual operation of the synchronous generator, with respect to the steady-state analysis. It defines the region of practical operation of the machine, preventing it from operating under overload conditions [8], [9].

The most explored limits in the bibliographies cover only limits of armature current and field. However, other limits are also important for the operation of the synchronous machine, such as mechanical turbine power limits, permanent stability limit and minimum excitation current threshold. [10,11,12]

3.1. Prime Mover Mechanical Power Limits

Mechanical power limits is verify due the maximum stress that prime-mover can stand without suffer damage in its structure. This

constraint will be defined through a constant value on P axis in the diagram. Given a mechanical nominal power S_{gNOM} for generators turbine the constraints will be:

$$P - k_{max} S_{gNOM} \leq 0 \tag{21}$$

$$k_{min} S_{gNOM} - P \leq 0 \tag{22}$$

where, k_{max} and k_{min} are constants that weighted mechanical power bounds.

3.2. Armature Current Limit

This limit can be modeled based on the relation of apparent, active and reactive power. Basically, when machines operate in a constant terminal voltage value, its armature current limits is established through its winding thermal limits. So, its noticeable that this constraint is a circumference centered in the origin [2]:

$$P^2 + Q^2 \leq (VI_{max})^2 \tag{23}$$

3.3. Field Current Limit

In the same way when current field is maximum (due its winding thermal limits) and consequently voltage field is maximum, the constraint becomes a circumference centered in $-V^2/X_s$: [2]

$$P^2 + \left(Q + \frac{V^2}{X_s}\right)^2 \leq \left(\frac{VE_{max}}{X_s}\right)^2 \tag{24}$$

3.4. SOCP-OPF with SGCC Constraints Formulation

Finally including SGCC in SOCP-OPF formulation, concerning that generator is a PV bus injecting active and reactive power, but now with its real constraints, formulation becomes:

Minimize (1)

Subjected to (2), (3), (8), (16 – 18) and (21) – (24)

The number of constraints for ACOPF is $(2NB + NPQ + 2NG + 2NL + 1)$ constraints and for SOCP-OPF with SGCC will be $(2NB + NPQ + 4NG + 3NL - 1)$. Number of constrains increase in more than two times.

4. Locational Marginal Prices

The new electricity market model adopted around the world had brought a series of advantages regarding services quality, reliability and security for power systems. Indeed, when the market is private and deregulated the competitiveness is much larger enhancing consumer’s electrical energy attendance. Connected systems truly operate more economically and for that reason, sellers and buyers agreed upon a price for a certain number of MWs. Although, each area on the system has its own price, and more specifically each generator has its own one. The concept of LMP allows determining the calculation of nodal price. It can be defined that, LMP for a specific bus is the energy cost needed to supply a 1 MW increment of load attending the operational constraints established for the system [9].

In practice LMPs corresponds to Lagrange multipliers on Karush-Kuhn-Tucker optimization conditions, for real power equality constraints on OPF.

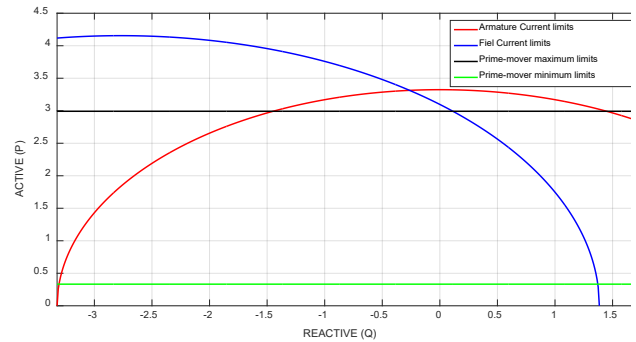


Figure 1-Synchronous Generator Capability Curve

5. Numerical Results

The proposed method was simulate using IEEE-14 bechmark system [12]. For these four situations were held: ACOPF including or not SGCC, and SOCP-OPF including or not SGCC. For all cases five scenarios of loading were take varying both, active and reactive power. All of them had its main results registered in Table 4. Although just 50%, case base and 150% of loading are specifically analyzed in Table 1-3. Tables register active and reactive powers and locational marginal prices for several situations regarding to a complete analysis.

Table 1- IEEE-14 generators power for Case-base

Generator N (bus)	ACOPF						SOCP-OPF					
	Classic			SGCC			Classic			SGCC		
POWER	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)
1 (1)	194.43	0	-	191.36	-16.11	-	192.99	0	-	190.37	-12.32	-
2 (2)	36.74	11.19	38.3705	36.19	25.25	38.0944	36.63	17.48	38.3132	36.16	28.86	38.0798
3 (3)	28.88	21.58	40.5775	20.6	22.9	40.4121	28.76	23.87	40.5752	21.93	25.1	40.4386
4 (6)	0	14.7	40.1957	10	15.45	39.7982	0	17.49	40.1977	10	16.67	39.871
5 (8)	8.19	19.55	39.6718	10	16.64	39.2565	10.14	13.25	39.6526	10	11.26	39.2879

Simulations were held in an Intel (R) Core (TM) i7-4770S 3.1 GHz, 4-Core and 8GB of memory CPU through a in MATLAB R2015a (8.5.0.197613) using optimization tools for solving the problem

The SGCC for generator connected to systems bus 1 in the 150% scenario is registered in Figure 1. For these constraints, concerned reactive bounds are larger than classical ones.

Optimality gap measures the quality in terms of result for a relaxation process and can be set as [13]:

$$\frac{AC\ Heuristic - Relaxation}{AC\ Heuristic}$$

For classical constrains Optimality gap is around 0.072 and 0.316%. Including SGCC constraint in OPF, values are in a range of 0.16 to 7.5%. In Table 3 computational running time its found to be much greater for SOCP-OPF if compared to ACOFP, when SGCC is included. This implies that solver technology stil needs to be improved for time equality in the process. Comparing just constraints modifications, computational time is not a problem.

Those modifications and adaptations are suggested in [14], [15], and [16].

Possible to make an important note from Table 3:

- When system load is low, and the system is lightly loaded OPF including SGCC gives a high cost.

- For heavy loading situations classical constraint gives a lower cost.

This can simply be explained for the fact that reactive limits are larger and minimum mechanical power is greater when taking SGCC approach. In high loading levels the relaxed method has to work hardly to find a feasible solution, and sometimes cannot reach convergence.

LMPs are shown in Tables 1 to 3 for both OPF and for models including or not SGCC constraints. Note that for case-base and 150% loading LMPs practically do not change. Although, for 50% loading case LMPs are lower when OPF includes SGCC.

6. Conclusion

This work exposed a relaxed formulation for Optimal Power Flow including Synchronous Generator Capability Curve constraints, that modeled machines practical bounds. Simulations showed that optimality gap is very short for both formulations. It was shown that for heavy loading levels, relaxed OPF works harder to find a feasible solution. SGCC gives directions for systems operators and planners in respect to availability of active and reactive power plant. Scenarios that SGCC are not concerned could either takes generator to fail in its operations due protection system actuating or damage windings when protection is note involved. Besides, obviously the model that uses SGCC will gives a high LMP, which implies that when it is not considered could

Table 2- IEEE-14 generators power for 50% case

Generator N (bus)	ACOPF						SOCP-OPF					
	Classic			SGCC			Classic			SGCC		
POWER	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)
1 (1)	111.89	0	-	85.58	-7.65	-	111.94	0	-	85.61	-3.95	-
2 (2)	20.76	-3.18	30.3807	15.8	1.11	27.8999	20.84	1.4	30.4211	15.86	3.72	27.9276
3 (3)	0	8.49	31.4963	10	7.35	28.6747	0	9.52	31.5674	10	8.32	28.7251
4 (6)	0	-1.1	31.1546	10	-1.78	28.4146	0	-2.42	31.2298	10	-5.6	28.4676
5 (8)	0	-1.23	30.9135	10	-2.08	28.2352	0	5.41	30.9698	10	4.01	28.2692

Table 3- IEEE-14 generators power for 150% case

Generator N (bus)	ACOPF						SOCP-OPF					
	Classic			SGCC			Classic			SGCC		
POWER	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)	P (MW)	Q (Mvar)	λ (\$/h)
1 (1)	194.43	0	-	191.36	-16.11	-	192.99	0	-	190.37	-12.32	-
2 (2)	36.74	11.19	38.3705	36.19	25.25	38.0944	36.63	17.48	38.3132	36.16	28.86	38.0798
3 (3)	28.88	21.58	40.5775	20.6	22.9	40.4121	28.76	23.87	40.5752	21.93	25.1	40.4386
4 (6)	0	14.7	40.1957	10	15.45	39.7982	0	17.49	40.1977	10	16.67	39.871
5 (8)	8.19	19.55	39.6718	10	16.64	39.2565	10.14	13.25	39.6526	10	11.26	39.2879

offer a fake price for generators power. Results for OPF including SGCC constraint can be analyzed as follow:

- Draw a situation dealing with overprice in light loading scenarios, or;
- Drives the result to a down price in heavily load conditions.

Table 4- ACOFP and SOCPFP results and computational time

Loading	Parameters	ACOPF		SOCOPF	
		Classic	SGCC	Classic	SGCC
50%	Cost	3299,60 \$/h	3608,20 \$/h	3303,50 \$/h	3610,50 \$/h
	Time	0,30 s	0,35 s	4,55 s	3,96 s
	Iterations	20	23	78	125
75%	Cost	5522,40 \$/h	5646,30 \$/h	5533,50 \$/h	5653,90 \$/h
	Time	0,30 s	0,26 s	1,85 s	2,07 s
	Iterations	19	18	71	81
BASE	Cost	8079,20 \$/h	8084,30 \$/h	8095,50 \$/h	8100,70 \$/h
	Time	0,29 s	0,36 s	1,80 s	1,94 s
	Iterations	18	22	73	76
125%	Cost	10704,00 \$/h	10703,00 \$/h	10723,00 \$/h	10722,00 \$/h
	Time	0,29 s	0,33 s	3,13 s	1,95 s
	Iterations	18	21	107	77
150%	Cost	13364,00 \$/h	13360,00 \$/h	13438,00 \$/h	13383,00 \$/h
	Time	0,31 s	0,35 s	1,78 s	1,91 s
	Iterations	20	23	74	77

Simulation time need to be improved, but this just could be done adjusting the solver technology. That is why it is much larger for the relaxation and still greater when SGCC is included. In these specific cases, LMPs are lower just when loading is 50% of case-base loading, but this situation depends on optimal power flow power limits data.

References

[1] M. Farivar and S. H. Low, "Branch Flow Model: Relaxations and Convexification—Part I," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2554-2564, Aug. 2013. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[2] I. G. Fernandes, V. L. Paucar and O. R. Saavedra, "Optimal power flow solution including the synchronous generator capability curve constraints with a convex relaxation method," *2017 IEEE URUCON*, Montevideo, 2017, pp. 1-4. doi: 10.1109/URUCON.2017.8171891

[3] L. S. Vargas, V. H. Quintana and A. Vannelli, "A tutorial description of an interior point method and its applications to security-constrained economic dispatch," *IEEE Transactions on Power Systems*, vol. 8, no. 3, pp. 1315-1324, Aug 1993.

[4] R. A. Jabr, "Optimal Power Flow Using an Extended Conic Quadratic Formulation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1000-1008, Aug. 2008.

[5] R. A. Jabr, "A Conic Quadratic Format for the Load Flow Equations of Meshed Networks," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 2285-2286, Nov. 2007.

[6] H. Hijazi, C. Coffrin and P. Van Hentenryck, "Polynomial SDP cuts for Optimal Power Flow," *2016 Power Systems Computation Conference (PSCC)*, Genoa, 2016, pp. 1-7.

[7] C. Coffrin, H. L. Hijazi and P. Van Hentenryck, "The QC Relaxation: A Theoretical and Computational Study on Optimal Power Flow," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3008-3018, July 2016.

[8] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill Professional, 1994.

[9] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. John Wiley & Sons, 2012

[10] UW-Madison, "Documentation on capability curves," May 2015, https://neos-guide.org/sites/default/files/capability_curves.pdf

[11] H. Zein and Y. Sabri, "Involving generator capability curves in optimal power flow," *2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, 2015, pp. 347-351.

[12] University of Washington, Dept. of Electrical Engineering. Power systems test case archive. Published online at <http://www.ee.washington.edu/research/pstca/>, 1999. Accessed: 30/04/2012.

[13] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Network flow and copper plate relaxations for AC transmission systems," *CoRR* vol. abs/1506.05202, 2015 [Online]. Available: <http://arxiv.org/abs/1506.05202>

[14] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *Int. J. Elect. Power Energy Syst.*, vol. 30, no. 6, pp. 383-392, 2008.

[15] R. A. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458-1459, Aug. 2006.

[16] Z. Miao; L. Fan; H. Ghassempour Aghamolki; B. Zeng, "Least Square Estimation-Based SDP Cuts for SOCP Relaxation of AC OPF," *IEEE Transactions on Automatic Control*, vol. PP, no. 99, pp. 1-1

A Development of Agility Mode in Cardiopulmonary Resuscitation Learning Support System Visualized by Augmented Reality

Keisuke Fukagawa*, Yuima Kanamori, Akinori Minaduki

Medical Informatics Laboratory, Kushiro Public University, 085-8585, Japan

ARTICLE INFO

Article history:

Received: 16 August, 2018

Accepted: 17 September, 2018

Online: 14 November, 2018

Keywords :

Cardiopulmonary Resuscitation
Visualization
Augmented Reality

ABSTRACT

This paper showed visualization of technics about cardiopulmonary resuscitation contributed to acquiring the skills and understanding them. Especially, this system is focused on the individuality in each object (men, women, and babies) There are problems of a general cardiopulmonary resuscitation (CPR) training because learners are generally taught by instruction based on a subjective judgment. Because of this way, they are difficult to get a knack about the CPR themselves.

Our system solved these problems by using Kinect (Augmented Reality) and Wii Balance Board (Weight Scale) when it calculates pressure and posture. They can understand the CPR training as fixed information. The system also expresses the features of compression as results whether the posture is extension or bending. However, it cannot evaluate a process of the compression. In the inspection, 84 people wrote a questionnaire which focused on impression of before and after. This questionnaire expressed two things. The one is that general public don't know the presence of CPR for infants. The other is that visualization was effective and enjoyable. As future works, Another function is going to be added to the present system to evaluate CPR in a process.

1. Introduction

Cardiopulmonary resuscitation (hereinafter referred to CPR), which maintains opportunities for life support by performing chest compressions on victims who are in a state of cardiopulmonary arrest, is a technique of life saving performed before and after using AED. This technique is well known. For example, some instructors teach the skills to learner in a general CPR training. In this case, learners can understand CPR. However, that tends to be subjective because the instructors teach them without using objective indicators. Because of this way, there are some problems in which learners push mannequins excessively. In addition, there are regulations about CPR in each object (men, women, infants, older people) but most people who participate the training don't learn them.

This new system is a visualized learning application of CPR which can calculate the pressure of compression and the angle of elbows with Augmented Reality (hereinafter referred to AR). It is possible to evaluate CPR in 1 minute. Learners of this system can

also get the objective feedback after the training. The system is also constructed another function (it is called agility mode). By this function, learners can understand individuality of CPR in each object.

In a research about this paper with experienced or clarified nursery teachers and medical personnel, the questionnaire verified the usefulness of this system and concentrated on two things. The first one is how much ratio of general public know the presence of CPR for infants, the other one is whether the design of this system is useful or not.

2. Related Researches

There are regulations and previous researches on CPR. It should be done as soon as possible by proper pressure of compression and tempo. Especially, the criterion on pressure of compression is different in each object (men, women, infants, older people) Our research is also based on the criteria in previous ones of World Health Organization (hereinafter referred to WHO), American Heart Association (hereinafter referred to AHA), and the Japan Resuscitation Council and the Japan Foundation for Emergency Medicine.

*Keisuke Fukagawa, Public University, 4-1-1 Ashino Kushiro City Hokkaido, Japan, +81-80-1275-1936, keiry.happy.life2525@gmail.com

www.astesj.com

<https://dx.doi.org/10.25046/aj030616>

2.1. The Situation of CPR in Japan

Types of commonly performed CPR in Japan are Basic Life Support (BLS)/Advanced Cardiovascular Life Support (ACLS) by AHA and BLS/ALS (Advanced Life Support) by the Japan Resuscitation Council and the Japan Foundation for Emergency Medicine. According to the Guideline Update 2015[1], as the situation clearly differs between in-hospital cardiac arrest (IHCA) and out-of-hospital cardiac arrest (OHCA), patients are divided into OHCA and IHCA cases and appropriate treatment methods are specified for both. This paper focuses on increasing the quality of lay rescuers CPR and describes a system development for OHCA.

2.2. The probability of Survival

The CPR is an efficient way to supply oxygen to brains when someone is in a cardiac arrest. If a person who in this situation, it will occur permanent damage for him/her. Dr. Drinker made the named Drinker curve on the influence in 1966 and was reported by WHO about one. According to previous researches, the positive effect of CPR is stronger if it is started in the cardiac arrest as soon as possible. The probability of survival also increases compared with nothing as times go [Fig.1].

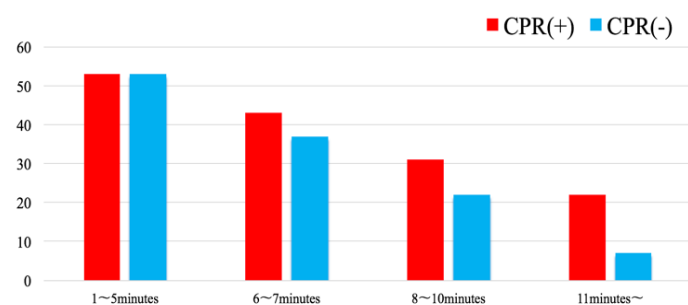


Figure. 1 This graph shows the effect of CPR in each phase. The vertical axis means the probability of survival and the horizontal axis does elapsed time after a cardiac arrest. Also, CPR(+) means a case of this resuscitation done, and CPR(-) does that CPR was not done.

2.3. Regulations of CPR

In the CPR, learners have to acquire skills and obey regulations to do it properly. First of all, the compression depth is at least 5cm but no more than 6cm. Secondly, learners should perform at 100-120 repetitions per a minute [2], [3]. Decompression isn't done properly if the number of the compression is more than 120.

The decompression plays an important part to do the CPR. Third, they do the CPR in an unceasing performance with minimal interruption time within 10 seconds. This is because the probability of survival de-creases. Forth, they have to do it with being perpendicular from the elbow jointing the base of the palm to do the CPR in a proper pressure. Fifth, the pressure of compression is between 40Kg and 50Kg (It is for men.) in order not to break breastbones. In general CPR, breaking breastbones

aren't bad to save a life. However, this breaking make patients take a long period to recover.

2.4. The Individuality of CPR for Infants

In the JRC resuscitation guideline summarized by the Japan Resuscitation Council, the two-fingers compression method is recommended as a CPR method for infants. The method is to compress by two fingers in the middle of the chest and press the sternum with one hand [4]. CPR to adults is also different from using only one hand. In consideration of this difference, the system is constructed so that the KINECT sensor can detect only one arm of the learner. Regarding pressure criteria for infants, the pressure is measured when the same portion of a mannequin similar to the actual infant's rib cage anterior-posterior diameter strength developed by Laerdal pressed to sink one-third of the whole. Besides, the value measured by this was set as the pressure detection standard of this system.

3. System Outline

This system is developed by 6 components and Microsoft Visual Studio 2017 and .NET Framework 4.5 with the C# language. There are a Kinect for windows version2 (1), a Wii Balance Board (2), a Communication Monitor (HDD) (3), a Mini Anne (CPR/ AED learning tool kit) (4), a Bluetooth USB adaptor (5), and Control PC (Windows OS) (6) in our system[Fig.2].

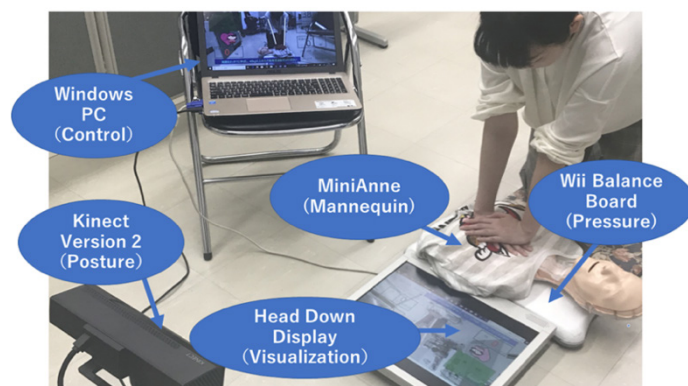


Figure 2 This figure shows the arrangement of our system. Learners can objectively understand the pressure of compression and the posture. They can do the CPR training with seeing the HDD.

3.1. Data Flow

Learners can experience a real CPR training themselves with using our system. The data flow is four steps [Fig.3]. The first one is choosing a target in men, women, babies (They have to choose right or left because the CPR for babies are done by only one hand).

Next, the system sets the level of sensitivity on the pressure of compression. The criteria on pressure compression and shape of men, women, and babies are different, so the system should change in each object. As third step, the AR of Kinect reads their body, and learners can start doing the CPR training. Lastly, the system counts whether their compression is extending or bent position pressure. Each compression of learners is counted.

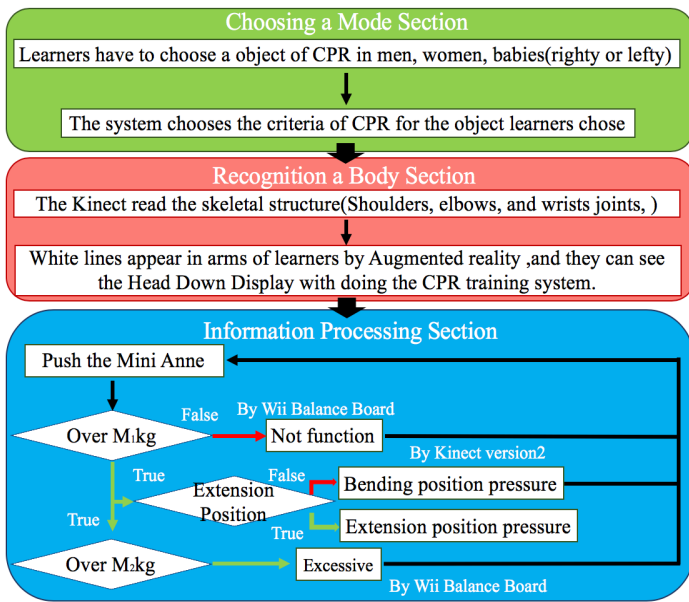


Figure 3 This chart shows the outline of our system. In the third section, M1 means a minimum and M2 does a maximum. In case of the CPR for infants, the criteria of minimum are 5 kg, and the one of maximum is 10 kg. This regulation is based on a research of Laerdal.

3.2. Feedback System

After finishing the training, our system reports their compression features. The results are classified by frequency of learner's compression and posture into five types. The posture of compression are counted in extension or bending position pressure [Fig.4].



Figure 4 The results are determined by whether the counts of extension position pressure are enough or not.

3.3. Motion Capture

This system can calculate whether learner's posture is extension or bending by AR. Kinect reads 26 joints in human body in initial setting, but it is controlled [Fig.5, Fig.6]. Learners can see the angles from shoulders to elbows and elbows to wrists in 6 points. Also, learners should obey regulations which was found in a process of the demonstrations below to use the system properly.

1. Tie your hair if you have long one.
2. Don't use in front of white walls.

3. Don't use the system outside.
4. Start to use the system from the pose of T.
5. Make your elbows straight (in an extension position).

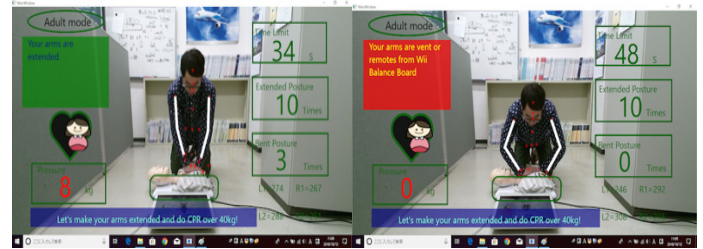


Figure 5 This left picture shows the situation of extension position pressure. This heart mark is also usually black. However, the heart mark becomes pink if the learners do compression over criteria. This right one shows the situation in the bending position pressure. If so, it is hard to press the Mini Anne in an appropriate compression force. This is because the bending position pressure needs power to do one properly more than the extended position one.

3.4. Agility Mode(Individuality)

An agility mode function was developed so that training suitable for the individuality in each object can be selected by expanding the system by the target age groups and gender classification. As a result, the learner can select CPR targets to be learned on the home screen [Fig. 6]. The subjects are classified into, men, women, and infants, and are displayed on the home screen accordingly. In particular, when choosing CPR training for infants, considering the difference in domestic hand for each learner, the interface was arranged whether it is for right-handed or left-handed, with classification about the individuality of each target.



Figure 6 This shows the start screen of our system. Learners can select the object they want to experience.

4. Results

In this verification, CPR training was experienced with 84 people (33men, 51women) and they answered closed questions. The questionnaire was mainly surveyed whether the people know the CPR for infants or not and the effect of visualization, and the impression of the CPR training before and after (Table.1).

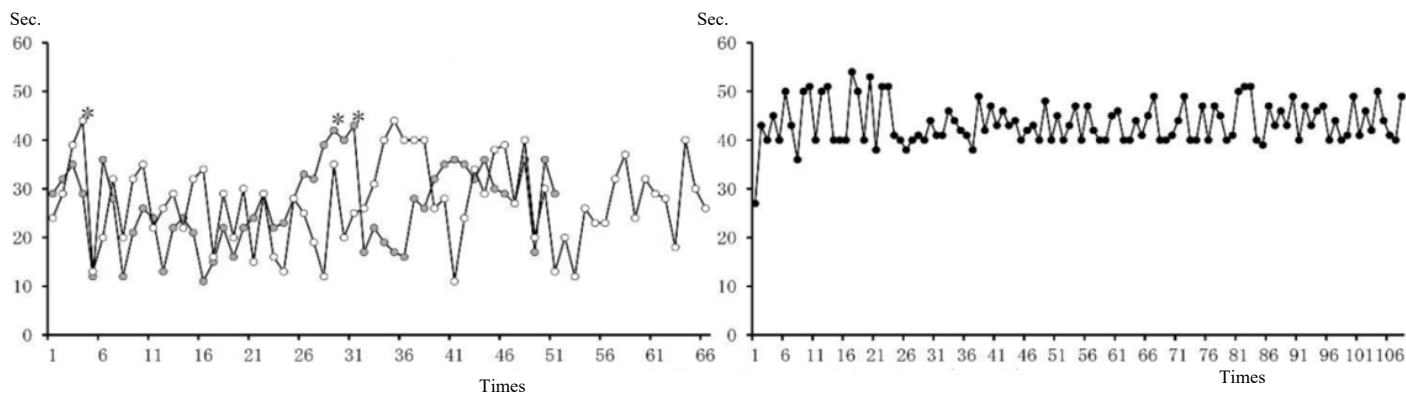


Figure 7 This graph (the object was men in this case) is that there is difference between before the CPR training (the left one) and after one (the right one) by one research subject. The vertical axis is a pressure of compression (kg) and, the horizontal axis means the number of compression(times). The white line is the first, and the gray one is the second. Also, the black one is the third. This verification clearly represents an effect by using the CPR training system because the pressure of compression of the research subject became more stable in an appropriate level, and the number of compression increased. In addition, this asterisk is the case of bent position pressure.

Table.1 The questionnaire of our system

Question Items
Q1 Have you ever participated in the lecture class of CPR training?
Q2 Did you know a specific way of CPR for infants before participating the lecture? (For Childcare Worker, General Public)
Q3 Do you mind doing the CPR for the opposite sex?
Q4 Was the CPR training System useful?
Q5 Were you able to study the CPR training with enjoying?

5. Discussion

According to this survey[Table.2], this system contributed two things. First of all, it showed that about 88% of general public didn't know the CPR for infants even if 75% of them have participated in the lecture. Second, about 92% of respondents could feel enjoyable and think it was useful. The system was actually focused on the design by using a whale because we wanted everyone to study with enjoying. However, the number of people with whom we trained the CPR was not enough to prevail it about infants and women.

Table 2 The Results of the questionnaire

Question	YES	NO
Q1	75.00%	25.00%
Q2(All)	45.31%	54.69%
Q2(Only General Public)	12.12%	87.88%
Q2(Only Childcare Worker)	93.10%	6.900%
Q3(Men)	52.17%	47.83%
Q3(Women)	19.51%	80.49%
Q4	98.44%	1.56%
Q5	92.19%	7.81%

6. Conclusion

In this paper, objects (men, women, babies) of this system could be classified as different ones. This progress is a first step to prevail the presence of CPR for general public. Another function is going to be added to the present system to evaluate CPR in a process.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This research was started by Medical Informatics Laboratory of Kushiro Public University in 2009. As a verification, Kushiro Kojinkai Nursing School, Ashino Nursery School, Kushiro City Childcare Support Center Eastern Branch helped us. We appreciate these organization.

References

- [1] Japan Resuscitation Association, "JRC Resuscitation Guidelines 2015 Online Version, Chapter 2 Adult Secondary Life Support Measures (ALS)" (2015)
- [2] Kralj E, Podbregar M, Kejzar N, Balažic J, "Frequency and number of resuscitation related rib and sternum fractures are higher than generally considered.Resuscitation".2015;93:136141.doi:10.1016/j.resuscitation.2015.02.034
- [3] A.E.Tomlinson, J.Nysaether, J.KramerJohansen,P.A.Steen,E.Dorph, "Compression force-depth relationship during out-of-hospital cardiopulmonary resuscitation", doi: 10.1016/j.resuscitation.2006.07.017
- [4] Japan Resuscitation Association, "JRC Resuscitation Guidelines 2015 Online Version, Chapter 3 Child Resuscitation (PLS)"

Analysis Refactoring with Tools

Zhala Sarkawt Othman*

Department of Software Engineering, Firat University, 23000, Elazig, Turkey

ARTICLE INFO

Article history:

Received: 30 September, 2018

Accepted: 14 October, 2018

Online: 14 November, 2018

Keywords:

Refactoring tools

Reverse engineering

Refactoring in VB

Analysis

ABSTRACT

The drive for this report is to inaugurate the innumerable techniques espoused by the refactoring tools in coding development. The software product is a very complex and time-consuming process of development. Difficulty understanding and maintaining poorly designed software systems Software maintenance can take up to 50% of total development costs for software production. As a modus operandi, the refactoring tools purpose ultimately to amend the basis codes into an easier and more comprehensible way.

Moreover, refactoring succors to check the trifle of the coding procedure. This is apparent through having deliberation on the program catalog, precision and the use of the deconstruct trees. Refactoring tools are convenient for innumerable observes done by the human beings. Software refactoring has a direct impact on reducing the cost of software maintenance by changing the internal structure of the code without changing its external behavior. So the time taken to process as well as doing a critical analysis of complex codes is reduced.

This report proposes to have a precarious scrutiny on the various use including the pluses of using refactoring tools.

1. Introduction

Refactoring is a technique of fluctuating the now prevailing basis codes using an unconventional as well as the incremental system. Moreover, the universal comportment of the code is not reformed through refactoring. Refactoring is advantageous as its reprocess results in either the refactored code being definitely employed in a more advanced way or for an additional tenacity.

Principally, refactoring is deliberated to be a nonspecific software design system which is neither tied to any defined solicitation language. For more accomplishment, the refactoring web site by Martin Fowler makes available a podium for use of a number of refactoring practices. Having a refactoring tool arranges for a much indispensable sustenance in scrutinizing a code rather than doing it manually. This is because manual refactoring consumes a lot of time.

Sanctioning the program writer to refactor his or her code without having them too personally retest the database is the significant driver for a refactoring tool [1]. This eradicates period lost when the process is either done physically or computerized. Moreover, the program catalog is critical and must be conserved repetitively in a securely cohesive atmosphere. This consents the

computer operator to scan and detect cross-references. Furthermore, it disregards the actuality of vibrant collation arising from the codes.

Moreover, refactoring encompasses guidance of the portion of the arrangement that is present below the average level. This comprises references to database elements which are being altered. This leads to update prepared on the references, therefore, management of the structure of the method.

The origination of a construe tree will portray the existing interior structure for the method itself. This includes:

```
Void hello ()  
  
{  
    System.out.println ("Hello World/n");  
}
```

Nonetheless, the refactoring process must be in route with the common comportment of the programs set in place. This is because the whole activities preservation for a program is unbearable to accomplish. This makes refactoring tools to be implemented to support improve the executions by most of the programs. This brands the programmers to either commend the usage of the refactoring techniques or fix glitches arising in their

*Zhala Sarkawt OTHMAN, Elazig 23119 Turkey, Phone: 05366836869;
E-mail: zhala.sarkawt@gmail.com

databases which cannot be done by the refactoring tools physically.

Moreover, the refactoring tool is eligible for backing the conduct of human. This involves, promptness: the breakdown, as well as the renovation necessary to complete refactoring, is time-consuming in cases where the manner is depicted to be erudite [2]. Several considerations are taken. This includes the cost of time plus the level of accuracy. In the case where refactoring consumes a lot of time, a programmer is restricted to the convention of the computerized refactoring but is accountable for manual refactoring as well as bear the consequences. This makes the speed of refactoring to be an important aspect.

The program writer is tasked with the choice of providing an eminence work at a cognizant time interval. Much of the statistics is already known to the programmer. However, this scheme can be perilously leading to the programmer making a lot of blunders while producing the prerequisite information. This fact countenances for most people to use refactoring tools as a foundation to avoid making errors plus a search platform for finding important information.

2. Method and Materials

A project analyzer is a refactoring tool that is essential to refactoring already in place Visual basic code. This tool has substantiated to be useful in various ways. The project analyzer will succor in the documentation of codes that which are openly connected resulting in its aiding from the refactoring techniques used.

As a refactoring tool, the scheme analyzer also completes robotically a splinter of the refactoring. Structures on the project analyzer which take account of the auto-fix assists in the computerized amendment of the basic code to be in line with the now predefined guidelines. This is also trailed by the programmed encapsulation of session variables as property Get/Set [3].

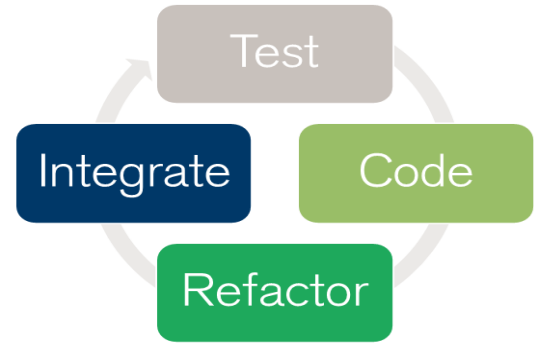
A number of vagaries in the codes entail mortal consideration, therefore, making a refactoring to be completed by hand. However, on monotonous tasks, mechanical changes are required. Visiting is mostly used in the refactoring process. This is done by the displaying a flow chart for a given piece of executable code. The codes can be in various forms including Visual Basic, Java, Pascal or Ruby. However, codes containing nested conditions, loops as well as jumps have logic errors [4].

There are some existing refactoring tools for the most widely used modern languages, such as Java, C# and C++, mostly as refactoring browser plug-ins to the most mainstream IDEs (e.g., Microsoft Visual Studio or Eclipse) [5].

The construction of flowchart assists to extant program logic which is easy to apprehend. In addition, the database logic is not related to the fundamental code that it is initially written [6]. It is easy to read the source code with the aid of the flowchart. This sorts the user categorizes logic defects which may be solid to ascertain by a graphical apparatus on the code.

Moreover, the flowchart assists to refactor the codes to systems that are easily decipherable plus increased. The

(flowchart 1) about Refactoring Process. This makes a graphic construal of the code to be more favored since it divulges designs which can be distinguished in the source code.



Flowchart 1, Refactoring Process

The refactoring techniques streamline methods, remove code duplication, and pave the way for future improvements.

In (Table 1) the following are examples of simple refactoring in Visual Studio.

Table 1, Visual Studio Refactoring

Refactoring Technique	Meaning in Life
Extract Method	This allows you to define a new method based on a selection of code statements.
Encapsulate Field	Turns a public field into a private field encapsulated by a .NET property.
Extract Interface	Defines a new interface type based on a set of existing type members.
Reorder Parameters	Provides a way to reorder member arguments.
Remove Parameters	As you would expect, this refactoring removes a given argument from the current list of parameters.
Rename	This allows you to rename a code token (method name, field, local variable, and so on) throughout a project.
Promote Local Variable to Parameter	Moves a local variable to the parameter set of the defining method.

This example of the "Extract Method" Refactoring. We use one of the tools in programming to improve the code in order to make the intent of the code clearer; you will extract the code that collects test cases from base classes into a new method called "printmarks".

1. We will select the following range of code inside the Test (Class):

- After we select three lines of the code from the selection's context menu in the editor; select Refactor > Extract Method or using (Alt+Shift+M) for the refactoring Extract Method:

```

1
2 public class Test {
3
4 public static void main(String[] args) {
5 // TODO Auto-generated method stub
6 int[]marks={1,2,3};
7 for(int mark1 : marks) {
8 System.out.println(mark1);
9 }
10 }
11 }
12 }
13 }
14 }
    
```

Figure 1, Select the following

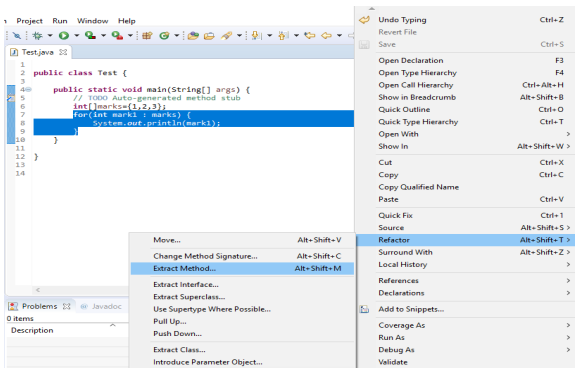


Figure 2, Refactoring Extract Method

- In the **Method Name** field, type *printmarks*:

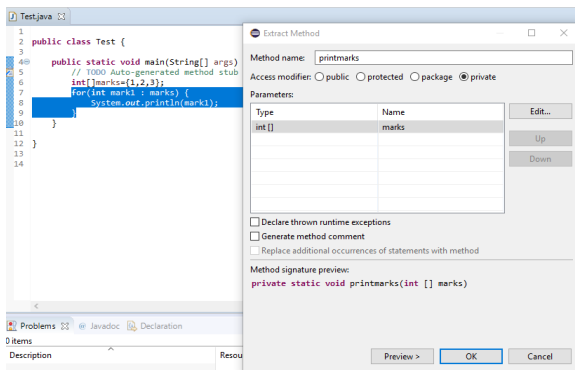


Figure 3, Method Name

- After refactoring, we get a cleaner and more readable code:

```

1
2 public class Test {
3
4 public static void main(String[] args) {
5 // TODO Auto-generated method stub
6 int[]marks={1,2,3};
7 printmarks(marks);
8 }
9
10 private static void printmarks(int[] marks) {
11 for(int mark1 : marks) {
12 System.out.println(mark1);
13 }
14 }
15 }
16 }
    
```

Figure 4, After Refactoring

3. Results

Various changes in the codes require human consideration, therefore, making a refactoring to be done manually. However, on routine tasks, automated changes are required.

The important part of the test is to make sure that your application performs efficiently and responsibly. This is where code analysis and profiling tools and techniques are evaluated: allows you to evaluate your code for errors, bottlenecks, and efficient use of processing and memory resources. Modern code recognizers can direct you to the exact lines of code that need to be resold.

The re-export tools change between IDEs and software programs. Visual Studio contains embedded built-in analysis tools. In addition, there are excellent tools to help you get deeper into your application for performance and optimization testing, project templates that rely on effective accreditation and embedded testing frameworks, and solid tools to integrate automated system analysis and testing into your business structure and workflow [7].

4. Discussion and Conclusion

The use of the hide method by the project analyzer allows the detection of excess method scope. It also suggests how to make the procedure private where it is convenient. Variables which exist outside of their respective classes or module are hidden. This procedure can also be done by making the variable to be local. Moreover, the cases for substitute nested conditional with guard clauses, the project analyzer is able to identify unwarranted conditional nesting. This is essential to understanding the steps taken in its execution. Furthermore, there is a high probability to replace the nesting with guard clauses. This procedure includes a sequential on the non-nested if statement, making it easier to read [8].

The procedure of using the manual as a refactoring tool is long. This is with the inclusion of a similar line of code in various locations. However, with the use of the visiting, there is a high probability for the user to alter the logic available. This enables elimination of the any existing copied lines. In the case where a logical structure exists in more than two procedures which may be as a result of duplication and pasting of the codes. Nonetheless, detections allow for the logic to be restructured into a new function and its access is from other functions [9].

Multifaceted algorithms are divided into various functions. The formations of the flowchart allow easy identification of every block from which the new functions were formed from. In cases where intricate conditional expression exists, it is removed using decompose conditional which also rewrites it into a new function. The multifaceted expressions formed are easily detected in a flow chart. Moreover, it provides a suitable platform for the logic to be rewritten in a simpler manner.

The use of the converse conditional reverses the logic of a conditional statement through the use of the not operator. However, it is not convenient to remove the Not from the code. The flow charts assist to display the usage of the reverse logic which is convenient for refactoring [10].

References

- [1] Murphy-Hill, E. and Black, A.P., 2008. Refactoring tools: Fitness for purpose. *IEEE software*, 25(5).
- [2] Campbell, D. and Miller, M., 2008, October. Designing refactoring tools for developers. In *Proceedings of the 2nd Workshop on Refactoring Tools* (p. 9). ACM.
- [3] Garrido, A. and Johnson, R., 2003, October. Refactoring C with conditional compilation. In *Automated Software Engineering, 2003. Proceedings. 18th IEEE International Conference on* (pp. 323-326). IEEE.
- [4] Murphy-Hill, E., 2006, October. Improving usability of refactoring tools. In *Companion to the 21st ACM SIGPLAN symposium on Object-oriented programming systems, languages, and applications* (pp. 746-747). ACM.
- [5] designs, S., 2016. *Semantic designs*. [Online] Available at: <https://www.semanticdesigns.com>
- [6] Fleming, Scott D., et al. "An information foraging theory perspective on tools for debugging, refactoring, and reuse tasks." *ACM. Transactions on Software Engineering and Methodology (TOSEM)* 22.2 (2013): 14.
- [7] Dorsey, T., 2017. *visual studio magazine*. [Online] Available at: <https://visualstudiomagazine.com>[Accessed 26 10 2017].
- [8] Liebig, Jörg, et al. "Morpheus: Variability-aware refactoring in the wild." *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*. Vol. 1. IEEE, 2015.
- [9] Murphy-Hill, E. and Black, A.P., "Breaking the barriers to successful refactoring: observations and tools for extract method" In *Proceedings of the 30th international conference on Software engineering* (pp. 421-430). ACM, 2008, May.
- [10] Fontana, F.A., Mangiacavalli, M., Pochiero, D. and Zanoni, M. "On experimenting refactoring tools to remove code smells". In *Scientific Workshop Proceedings of the XP2015* (p. 7). ACM. 201

Management Tool for the “Nephele” Data Center Communication Agent

Angelos Kyriakos^{*1,2}, Thomas Tsavalos¹, Dionysios Reisis^{1,2}

¹ National and Kapodistrian University of Athens, Electronics Lab, Physics Dpt, GR-15784, Zografos Greece

² Institute for Communication and Computers (ICCS), National Technical University of Athens, Greece

ARTICLE INFO

Article history:

Received: 23 July, 2018

Accepted: 04 October, 2018

Online: 14 November, 2018

Keywords:

Graphical User Interface

Data Center

Agent

ABSTRACT

Optical switching provided the means for the development of Data Centers with high throughput interconnection networks. A significant contribution to the advanced optical Data Centers designs is the Nephele architecture that employs optical data planes, optical Points of Delivery (PoD) switches and Top of Rack (ToR) switches equipped with 10 Gbps connections to the PoDs and the servers. Nephele follows the Software Defined Network (SDN) paradigm based on the OpenFlow protocol and it employs an Agent communicating the protocol commands to the data plane. The current paper presents a management tool for the Agent. The Agent's management tool is utilized to configure the Agent, create commands, perform step operations and monitor the results and the status. Moreover, as a testing and validation tool, it plays a significant role in the improvement of the Agent's design as well as in the upgrade of the entire data center's organization and performance.

1. Introduction

Currently, the integration of Information Technology (IT) activities and applications takes place in data centers, which also include the necessary devices for communication, high performance computing and data storage. Data centers play an important role in organizations based on IT services, as they provide the means for fast responses to business demands, they facilitate the IT operations and their utilization leads to the reduction of the capital expenditures and the operating costs. Targeting the improvement of data centers, researchers and engineers focus on the use of optical switching due to the bandwidth capabilities that it provides. A significant contribution to this design effort features optical links connected through optical Point of Delivery (PoD) switches to the Top of Rack (ToR) switches, SDN with OpenFlow organization, an Agent connecting the SDN controller and the data plane and an enhanced agent management tool [1], which all integrate in the Nephele [2] data center.

The Nephele is based on a dynamic optical network infrastructure for scale-out, disaggregated datacenters that leverages optical switching with SDN control and orchestration to overcome current datacenter limitations. The Nephele design

follows vertical end-to-end development approach extending from the data center architecture to the overlaying control plane and its interface to the application, in order to deliver a fully-functional networking solution, extending network virtualization to the optical layer. The Nephele design achieves dynamic reconfiguration by utilizing the slotted operation of the network based on the Time-Division Multiple Access (TDMA). Moreover, the SDN control can effectively manage the data plane elements. The OpenFlow protocol communicates the SDN control's messages to the data plane [3]. Nephele uses an Agent to realize the communication between the SDN controller and the data plane. The Agent includes functions filtering the control plane (SDN controller and the Agent) instructions that are transmitted through the OpenFlow messages; the Agent translates these messages and forwards them to the corresponding ToR switch. Although, the Agent can be classified as a back-end process, there is a need for an interactive management tool that allows the interaction of the designers and the future users with the Agent. The need for the above tool appeared in the course of the data center's design and implementation phase, it became more emphatic during the integration and finally the validation and testing phases. Similar interactive tools are reported in the literature as important tools for the management, testing and evaluation of networks [4], [5], [6].

*Dionysios Reisis, +30 210 727 6708/6720 & dreisis@phys.uoa.gr

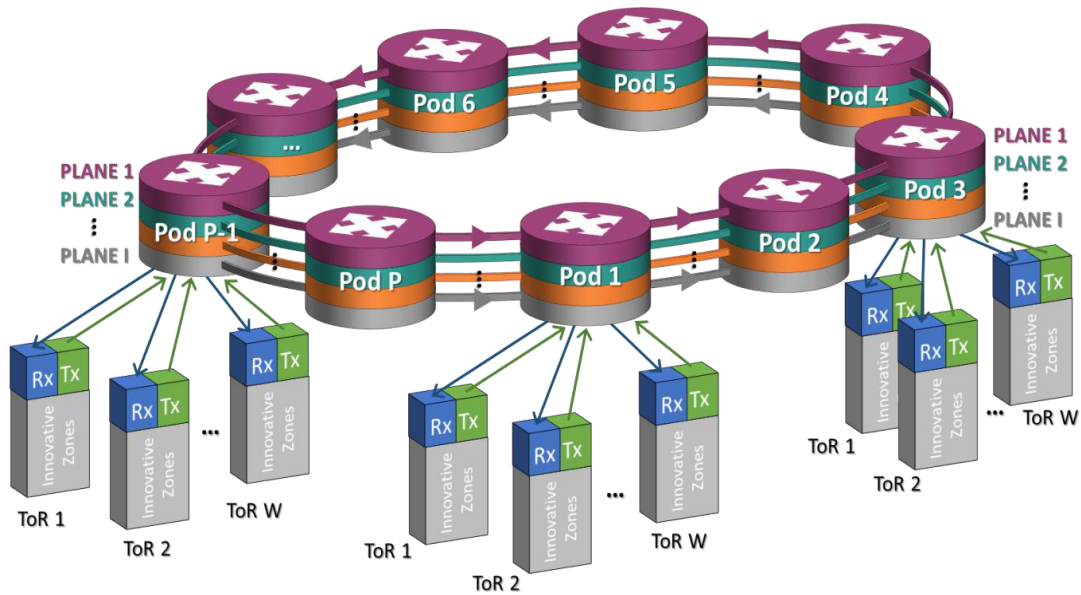


Figure 1: Nephelē Data Center Network Architecture

Focusing on providing an effective tool mainly for advancing, testing and monitoring the Agent’s functionality and performance [7], the current work presents a management tool for the Nephelē Agent. The proposed Agent’s tool is able to access all the information that it is directed to the data plane. Moreover, it can be used to create the commands for the data plane, monitor the commands transmission to the devices and also, the corresponding responses of the devices to the Agent. Furthermore, it provides the ability to request all the information with respect to the status of the devices. The use of the proposed management tool contributed significantly to the development of the entire Nephelē data center and consequently the testing phase. Additionally, it benefits the entire system because it will still be most suitable for effectively monitoring the Agent’s performance during normal operation and also it provides the means for realizing scenarios in the cases of demonstrations and presentations [7].

The paper is organized as follows: Section II highlights the Nephelē data center architecture. Section III presents the Agent’s management tool and Section IV concludes the paper.

2. The Nephelē Data Center

The Nephelē data center involves a slotted hybrid electrical/optical interconnection network that is advantageous with respect to the dynamic allocation of resources. The network includes PoDs of racks that communicate with the so-called innovation zones, which are the devices dedicated for the disaggregated computing, storage and memory resources. The innovation zones are connected to ToR switches [8]. Each innovation zone can communicate to other innovation zones through an all optical or an electro-optical channel. The architecture of the Nephelē data center is depicted in Figure 1.

The Nephelē data center is designed for an operation that includes dynamic and efficient sharing of the optical resources

and a collision free network operation by using Time Division Multiplexing Access (TDMA). The control plane is based on a Software Defined Network (SDN). The SDN controller is divided in two distinct interfaces, namely the Northbound Interface and the Southbound Interface. A high-level view of the Nephelē control plane architecture is presented on Figure 2.

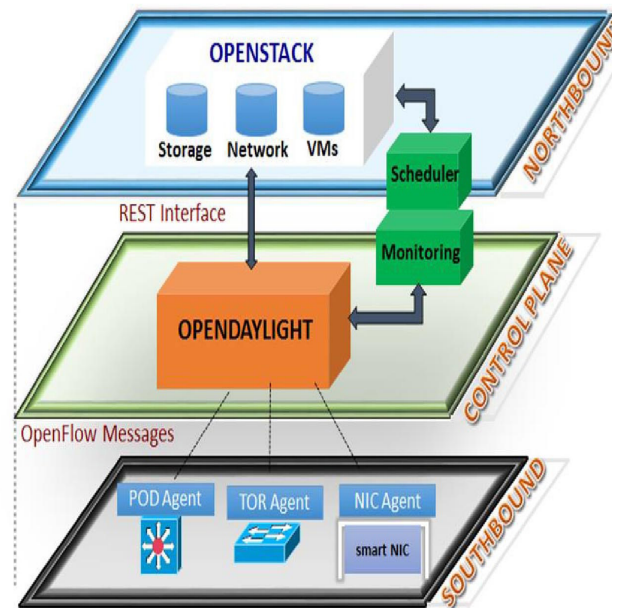


Figure 2: Nephelē SDN Control Plane

The Application to Controller Plane Interface defined by ONF (Open Networking Foundation) in the SDN architecture is realized by the Northbound Interface of the Nephelē SDN controller. This interface allows the interaction between the core services of the Nephelē SDN controller and the upper layer network applications, which implement the logic of the network resource allocation in the data center. The Nephelē’s design follows the approach of an overall centralized architecture. For

this purpose, all the scheduling plans are carried out according to the algorithms that are performed by the central controller's Traffic Offline Scheduling Engine [9]. Considering the optimization of the utilization of the entire network the Offline Scheduling Engine is equipped with mechanisms able to allocate resources of the data center network in the long term.

The data-controller plane interface defined by ONF in the SDN architecture is realized by the Southbound Interface of the Nephela SDN controller. The commonly used in these cases OpenFlow has been chosen as a standardized communication channel for this interface. It executes two main tasks: to command and configure the data plane devices via the device specific Agents. A device specific Agent performs as a proxy for the data plane switching devices. Consequently, the Agent should have two communication interfaces the Agent-Controller interface and the Agent-FPGA interface. The Nephela Agent's is mainly devoted to filter the control plane instructions, that are included in the OpenFlow messages. Additionally, the Agent translates these instructions and then, it forwards them to the corresponding FPGA via a PCI Express interconnection. The Agent is a back-end process. It is activated at the beginning of each Nephela scheduling period and it will communicate the new schedule instructions in order to configure the data plane switches. The instructions come in the form of scheduling tables; the format of these scheduling tables is presented by Figure 3.

Timeslot	Destination	VLAN	Wavelength
1			
80			

Figure 3: The Format of the Schedule Table

3. The Management Tool of the Agent

The present section describes first the graphical user interface (GUI) architecture of the management tool of the Agent; second, the tool's usability and third, the back-end of the management tool.

3.1. The GUI Architecture

The Agent's management tool is implemented by using the JavaFX software platform of the Java programming language; JavaFX consists of a set of graphics and media packages, which provide the means to the developers for the design, creation, testing, debugging, and deployment of rich client applications that operate consistently across diverse platforms. The management tool includes a GUI that presents to the user a Nephela network of smaller size as an image-map. This image-map includes clickable areas, which are illustrated graphics created on a raster graphics editor and enhanced with interactive attributes. This design has led to the implementation of a graphic environment, which, considering the interaction of the user with the management tool, ensures both, optimized usability and user experience, compared

to an environment using the standard widgets, provided by JavaFX.

The user of the management tool sees the data center network, the scheduling table, an explanatory image and a menu, which are brought to her/him as the main scene of the GUI. This main scene is shown in Figure 4. The smaller scale network includes four PoDs residing in the network and connected via four WDM (Wavelength Division Multiplexing) rings. Each of the PoDs includes four PoD elements; these are divided into the disaggregated rack and the ToR switch.

The GUI includes an explanatory image, that is located over the menu in the right top corner. The image presents an enlargement of a PoD element in higher resolution and it is augmented with annotations, so that the user is able to understand what the image portrays.

In order to present the graphic display of the PoD elements three objects of the ImageView class were stacked in a StackPane object [10]. This design has been implemented as follows: they were aligned three image layers one over another (depicted by Figure 5), so that they appear as a single solid object and at the same time the developer can handle each one independently. The ImageView object is a type of Node object in the JavaFX Scene Graph that is used for painting a view; the painting is carried out by using data contained in an Image object. The StackPane is also a type of Node object acting as the layout container and it contains the ImageView objects. The three ImageView objects include the images that represent the ToR switch, the disaggregated rack and a visual effect.

In the GUI, the ToR switches are the interactive parts of the management tool: the user can select by clicking on them and she/he can create the scheduling table of the data center. Each ToR is a clickable area and it can be used by the user as the source and/or the destination in the scheduling table entry. In our case the upper left ToR is chosen by default as the host Agent PC scheduling engine. This is the source ToR and the remaining ToRs are the destinations. The interactive feature is accomplished by registering an event handler on the ImageView object that includes the ToR image. An event handler is an implementation of the EventHandler interface. The handle() method of this interface will let the code filling the entries in the scheduling table to perform if the ToR image is clicked. Upon the cursor click event, all the necessary code is executed to fill in the required fields of a scheduling table's entry. The management tool fills the Destination field with the identity (id) of the ToR switch where the event occurred. The Timeslot field takes the value of the time sequence of the event, which is calculated based on a counter. The Wavelength field is filled with a value selected from a closed interval of integer values. Finally, the VLAN (Virtual LAN) field entry represents the identification number that is assigned to the WDM ring, through which the data transmission will occur. Furthermore, when the ToR is clicked, as depicted in Figure 5, it

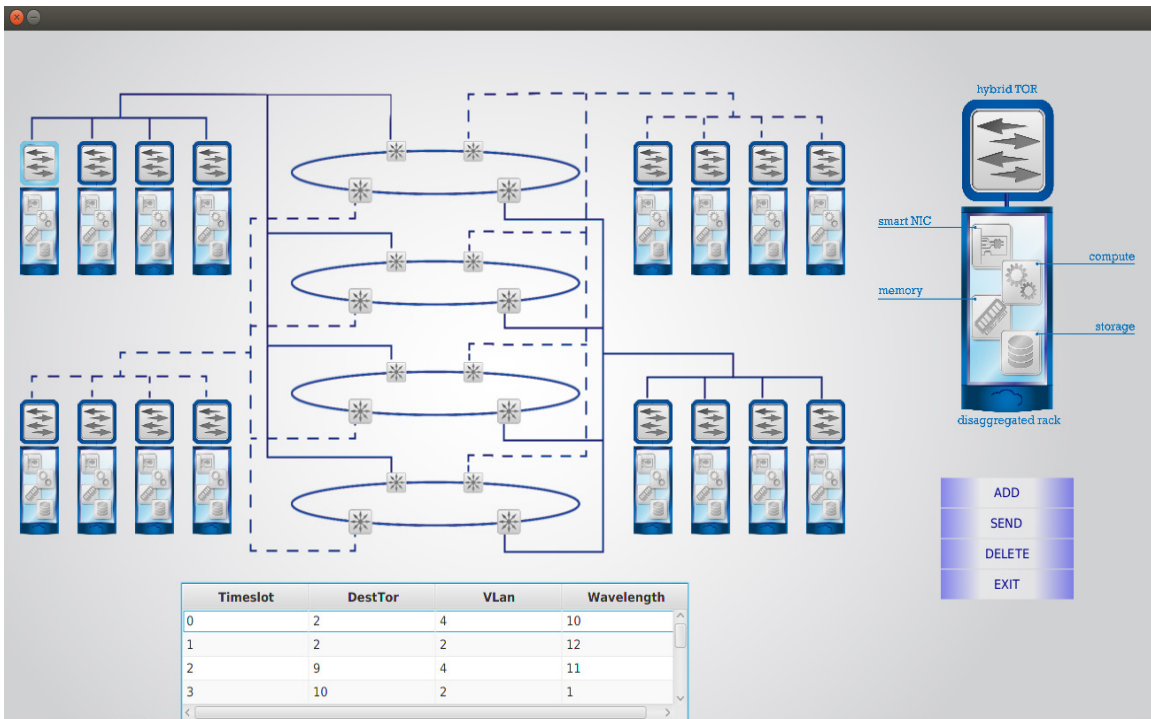


Figure 4: The GUI Main Scene

will trigger the effect displaying that it is the selected ToR. The effect is represented by a brighter image enclosing the ToR switch. The effect is set not to be visible at first, it will be set to full opacity if the ToR is selected and it will return to zero opacity with a two seconds lasting fade transition. The fade transition is an instance of the FadeTransition class, which is a subclass of the JavaFX Animation class and it changes the opacity of a node over a given time. The same effect has been implemented similarly to the WDM rings and it indicates graphically what WDM ring is chosen based on the VLAN field in the scheduling table entry.

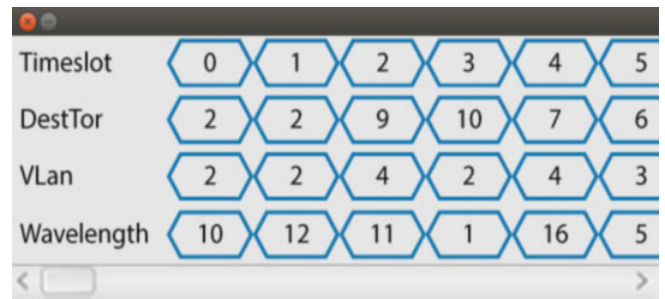


Figure 6: Pop-up Window with the Values sent to FPGA

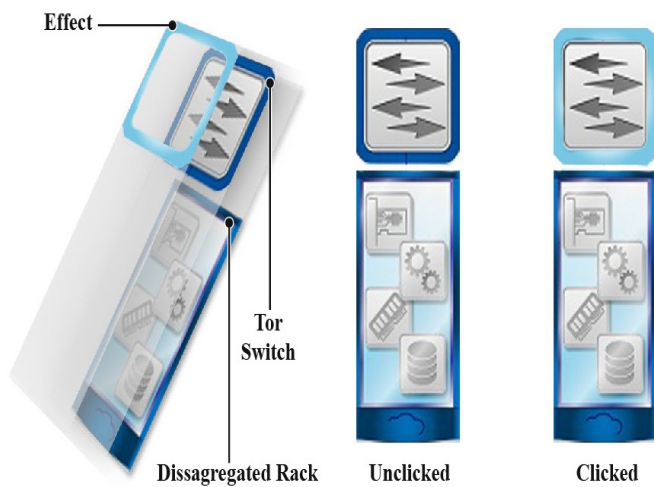


Figure 5: Effect of clicking the ToR

All the aforementioned elements of the tool's design let the user to construct the scheduling table and provide the option of editing it; this operation can be carried out by the use of the menu. The menu consists of four buttons and inherits its attributes from the VBox class, which is a container that sorts its contents into a single vertical column. The menu buttons were created as a separate class. It is distinct from the Button class of JavaFX and is created by stacking a TextField object over a filled Rectangle object. This object's filling is colored by an instance of the LinearGradient class, in order to apply effects that are suitable to the entire design of the GUI and preserve the uniformity to the user eye. These effects are triggered by the events originating from the mouse cursor and their implementation is based on switching the order of the colors in the gradient fill. Each time the user clicks the Add menu button she/he will start a new session of constructing a scheduling table and the source ToR will be automatically selected and indicated. A scheduling period of the Nephel network can accommodate up to eighty entries, as the corresponding allowed time slots. If the user exceeds that ceiling,

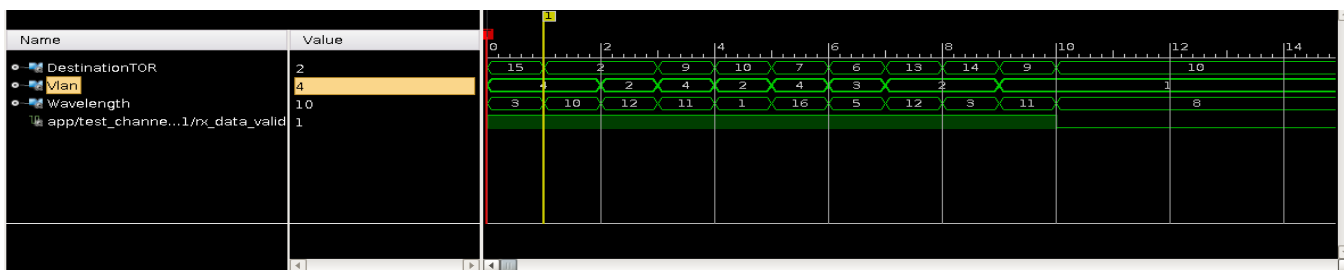


Figure 7: Output of the Logic Analyzer

a pop-up dialog box will emerge with the corresponding message, prompting her/him to stop importing entries. The dialog box prevents the user from interacting with the main application window but it keeps the window visible in the background. When the user has completed the creation of the scheduling table, she/he is able to review it and delete any misplacing entries by using the delete button from the menu. If the key is pressed and no entry is selected or the scheduling table is empty, a pop-up window will be called informing the user of the corresponding case. As a final step the user presses the send button, an action which transmits the scheduling table to the FPGA data plane devices.

The conclusion of the transaction is marked by the appearance of a pop-up window that it will be shown to the user. The window includes all the values that were sent to the FPGA in a format that resembles that of a logic analyzer. The pop-up window is shown in Figure 6 and the corresponding output of the logic analyzer is depicted in Figure 7. The logic analyzer exports the output as a CSV file (Comma-Separated Values); this file can be processed by the management tool and in this case, the file's values will be forwarded to the pop-up window. The pop-up window incorporates the graphical theme of the management tool and is designed to model the layout of the logic analyzer.

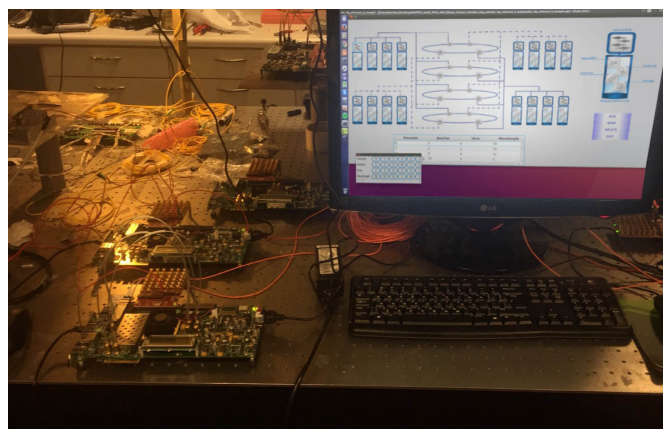


Figure 8: Nephela Data Plane Development

3.2. Usability of the Agent's Management Tool

The use of the management tool is of great importance to the development and operation of the data center, since the users can create their own traffic schedule and then transfer that schedule to the data plane ToR switch. The engineers are able to control the data plane switches, during the development and testing phase of

the physical layer of the data center network as shown in Fig 8. The tool's GUI allows to construct the commands directly in the format of the scheduling tables of the FPGAs (instead of using the OpenFlow protocol). Additionally, it is straightforward to extend the management tool for creating the scheduling tables of a PoD switch in the Nephela network. Given the fact that there is a ToR Agent PC for each ToR switch in the network, the tool is executed on the Agent computer and it provides to the user the means for the scheduling of the network from the view point of the specific ToR switch. The user can control graphically and more importantly in real time the transmission of Nephela frames originating at the ToR switch (that is controlled by the Agent computer) to the other Nephela ToRs in the data center network [11].



Figure 9: Live Demo of the Management Tool

The benefit of designing, developing and effectively using the proposed management tool has been already proven during test procedures and demonstrations. An illustrious example is the application, which has been shown during a presentation of the control plane of the Nephela data center. The scenario for this demonstration has as follows: the control plane includes parts of the FPGA's implementations of the data plane, the Agent, and the SDN controller. Given that a functional data center Agent was not available, we presented the control plane by dividing it into two experiments. The first experiment demonstrates the SDN controller and the second the FPGA's operation controlled by the management tool. The management tool has successfully imitated

the functions of the Agent; the majority of the people that interacted with the management tool understood the concepts behind the architecture of the Nephele network and the function of the Agent in the Nephele data center. The demonstration as shown in Figure 9 consists of the SDN controller software presentation, the FPGA that represents the ToR switch, and the Desktop PC that executes the management tool, which is connected to the FPGA board via PCI Express. The user can interact with the management tool and give his/her own commands to the demonstration system.

3.3. Back-End of the Agent's Management Tool

In the Nephele data center the ToR switch design includes multiple FPGAs; all the FPGAs that belong to a single ToR implementation use PCI Express to communicate with the host ToR Agent computer. The management tool divides the scheduling information to distinct parts, so that each part corresponds to the scheduling information concerning the corresponding FPGA; then it creates distinct threads to complete the entire operation. We use a single thread to communicate with a single FPGA and transfer the respective part of the ToR switch traffic schedule. Note here that, the communication is performed in parallel for all the FPGAs belonging to the same ToR switch.

In order to develop the PCI Express interface of the FPGAs we used the Xilinx IP Core for PCI Express and the RIFFA (Reusable Integration Framework for FPGA Accelerators) framework [12]. The framework consists of an API (Application Programming Interface), a driver/kernel module and an IP core for the FPGAs. All the above parts are open-source. It is designed to perform with the Xilinx IP core that handles the physical layer of the PCI Express interface. The API is designed to support multiple languages like C/C++, Java and Python. Moreover, it includes the necessary function/methods that the management tool needs to invoke, in order to communicate with the FPGA. The entire API is designed to be executed by threads and the design of the management tool takes full advantage of this capability.

The communication that is directed from the Agent PC to a FPGA operates according to the following steps. In the first, the application initiates the transaction by calling the `fpga_send` method. Then, the thread invokes the operation of the kernel driver, which writes to the FPGA configuration registers the necessary information to begin the transaction. The FPGA uses DMA (Direct Memory Access) to read the scatter gather elements [12] that the driver instructed. At the time that the transaction will be completed the driver will read the final count of the data passed, the amount of the data is then returned to the management tool as the return value of the `fpga_send` method.

In the design of the tool special attention was paid to the operation of the RIFFA API, because the RIFFA's driver requires all the data in contiguous memory locations (in an array). Note here that, the Java's Array object can't be used in this case. An attractive solution to this problem is the employment of the

ByteBuffer Class of Java, which is a class that is created to handle a stream of raw data. The operations on the buffer can be carried out byte by byte, but casting is also supported for the user to be able to write a whole Java data type, like an integer.

Finally, the endian of the data has been tackled as follows. The JVM (Java Virtual Machine) stores class files in big endian byte order, where the high byte comes first. Multibyte data items are always stored in big-endian order. Given that the Xilinx FPGAs operate in little-endian byte order, the change of the endianness could be arranged either during the construction of the ByteBuffer or at the receiving buffer in the FPGA. The latter choice has been proven more efficient and gave us the advantage of the ByteBuffer casting, which would not be useful in the case of changing the order of the byte inside the ByteBuffer in the Java application.

4. Conclusion

The current paper presented a management tool for the Agent of the Nephele data center. The advantage of creating and using the proposed management tool is that the data center designers and engineers can create their own schedule as the tool's GUI users and then transfer that schedule to each data plane ToR switch. The user can control graphically in real time the transmission of Nephele frames originating at the ToR switch to the other Nephele ToRs in the data center network. Moreover, the management tool can be of even further use if it will be extended to create the scheduling tables of a PoD switch in the Nephele network.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work has been funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 645212 (NEPHELE).

References

- [1] A. Kyriakos, T. Tsavalos and D. Reisis, "GUI for the communication agent of the "Nephele" data center," 2017 South Eastern European Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Kastoria, 2017, pp. 1-5. doi: 10.23919/SEEDA-CECNSM.2017.8088237
- [2] P. Bakopoulos, K. Christodouloupolos, G. Landi, M. Aziz, E. Zahavi, D. Gallico, R. Pitwon, K. Tokas, I. Patronas, M. Capitani, C. Spatharakis, K. Yiannopoulos, K. Wang, K. Kontodimas, I. Lazarou, P. Wieder, D. Reisis, E. Varvarigos, M. Biancani, H. Avramopoulos, "NEPHELE: an end-to-end scalable and dynamically reconfigurable optical architecture for application-aware SDN cloud datacenters", *IEEE Communications Magazine*, 2018
- [3] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. 2008. "OpenFlow: enabling innovation in campus networks." *SIGCOMM Comput. Commun. Rev.* 38, 2 (March 2008), 69-74.
- [4] Yi-Bing Lin, Joe Geigel, "A graphical user interface design for network simulation," *Journal of Systems and Software*, Volume 36, Issue 2, Pages 181-190, February 1997.
- [5] M. Turon. 2005. "MOTE-VIEW: a sensor network monitoring and management tool." In *Proceedings of the 2nd IEEE workshop on Embedded*

Networked Sensors (EmNets '05). IEEE Computer Society, Washington, DC, USA, 11-17.

- [6] S. Corazza, S. Reale, "Network management system graphical interface", published in Eighth International Conference on Software Engineering for Telecommunication Systems and Services, 1992.
- [7] G. Landi, I. Patronas, K. Kontodimas, M. Aziz, K. Christodoulopoulos, A. Kyriakos, M. Capitani, A. Hamedani, D. Reisis, E. Varvarigos, P. Bakopoulos, H. Avramopoulos, "SDN control framework with dynamic resource assignment for slotted optical datacenter networks," 2017 Optical Fiber Communication Conference, Los Angeles, California, USA, March 2017.
- [8] Ioannis Patronas, Angelos Kyriakos, Dionysios Reisis, "Switching functions of a data center Top-of-Rack (ToR)," 23rd IEEE International Conference on Electronics Circuits and Systems, Monte Carlo Monaco, Dec. 2016.
- [9] K. Christodoulopoulos, K. Kontodimas, K. Yiannopoulos, E. Varvarigos, "Bandwidth Allocation in the NEPHELE Hybrid Optical Interconnect", 2016 18th International Conference on Transparent Optical Networks (ICTON), July 2016.
- [10] Johan Vos, Weiqi Gao, Stephen Chin, Dean Iverson, James Weaver Pro, JavaFX 8: A Definitive Guide to Building Desktop, Mobile, and Embedded Java Clients [1 ed.] 2014 p.206-208.
- [11] Chen J-W, Zhang J., "Comparing text-based and graphic user interfaces for novice and expert users," *AMIA Annual Symposium Proceedings*. 2007;2007:125-129.
- [12] Matthew Jacobsen, Dustin Richmond, Matthew Hogains, and Ryan Kastner. 2015 RIFFA 2.1: A Reusable Integration Framework for FPGA Accelerators. *ACM Trans. Reconfigurable Technol. Syst.* 8, 4, Article 22 (September 2015), 23 pages. Available: <http://dx.doi.org/10.1145/2815631>

Actual Use and Continuous Use of Retail Mobile App: A Model Comparison Perspective

Sunday Adewale Olaleye^{1,*}, Ismaila Temitayo Sanusi², Bisola Adepoju³

¹*Oulu Business School, Department of MMI, University of Oulu, 90014, Finland*

²*Philosophical Faculty, University of Eastern Finland, 80100 Joensuu, Finland*

³*Texas Instrument, Dallas, Texas, 75243, United States*

ARTICLE INFO

Article history:

Received: 09 July, 2018

Accepted: 26 September, 2018

Online: 14 November, 2018

Keywords :

Mobile app

Retail

Use

Continuous Use

Gratification

ABSTRACT

A mobile retail app is a growing innovation in a retailing setting and there is an argument on the prominent status of a mobile application in contrast with a mobile website and web applications. The study used quantitative data to run multiple regression analysis with keen attention to linear regression assumption and compare four models for mobile retail app use and continuous use based on mobile retail app technology, trust, and gratification. Theoretically, the study integrates the unified theory of acceptance and use of technology (UTAUT), trust and gratification and expand the technology acceptance model with the trust and gratification elements. To have a better understanding of the hypothesized theory and clearer perception of the model that have explanatory power, the study employs SPSS linear regression and general linear regression to look at the relationship of mobile app technology, trust and gratification predictors and the outcome variable. The study emphasized the importance of trust, privacy assurance, learning and relaxation features in a mobile retail app as an antecedent of its use and continuous use. This is a novel contribution to the literature on technology acceptance and retailing. This study also shed more light on the importance of age as a moderator of gender and marital status regarding mobile retail app use and continuous use. Further, it also explicates the managerial implication of mobile app and makes a necessary future recommendation.

1. Introduction

There is an argument on the prominent status of a mobile application in contrast with a mobile website and web applications most especially in a retailing context. The mobile app is a challenger of the web app, mobile website and since its emergence in 2008, there are more than 700,000 different apps that operate on different operating systems platforms such as Android, Windows, and iOS [1]. The mobile app is "a type of software that allows the user to perform a specific task that can be installed and run on a range of portable digital devices such as smartphones and tablets". In [2], the author defines mobile app as "an IT software artefact that is specifically developed for mobile operating systems installed on handheld devices, such as smartphones or tablet computers". The two definitions reveal that mobile app cannot be used in isolation but through a device. A mobile app can be distinguished as software for smart mobile devices, it could

be premium or freemium and downloadable through a centralized online market with an opportunity to rate and review based on the users' experience [3]. Despite the challenge of poor performance of some mobile retailing app, as at 2015, the global app revenue has soared to 8.3 billion dollars, and by the year 2020 mobile apps are predicted to reach 189 billion dollars in revenues [4]. A mobile app is dynamic and flexible for modification to meet the need of an individual retail store. A mobile app is attracting more customers daily and positively impacting the business of app developers, mobile device manufacturers and internet service providers [5-6]. Due to mobile app different potential and benefit, it is regarded as one of the fast-growing technology markets globally. A mobile app is an interesting research domain for the researchers and [6], integrates the Theory of Planned Behavior, the Technology Acceptance Model, and the Uses and Gratification Theory to examine the American consumers' mobile apps attitudes, intent, and use. Despite the result of this study, [2] discovered a gap in mobile app evaluation regarding theory and methodological clarity and recommended their instrument as an

*Corresponding Author: Sunday Adewale Olaleye, *University of Oulu, 90014, Finland*, +358466424139 sunday.olaleye@oulu.fi

adequate measurement of a mobile app. Reference [1] explore app consumption and exploratory analysis of the uses and gratifications of mobile apps while [2] focus on mobile app usability, conceptualize and developed mobile app instrument for general use. Reference [7] investigates the intention of a mobile app to disclose their information dwelling on privacy calculus theory. In [5], the author suggested an expansive study of a mobile app to researchers. This mobile app study is country focused. Finland is a famous country in mobile banking and other emerging technologies. S Group, Finland's renown retailer won NACS Insight European Technology Implementation Award in 2016 for an added fueling feature for its S-mobile app. The S-mobile app exhibits three features of retail, banking and fuel services [8]. Kesko is another significant store chain in Finland with a 33.1% share of the 16.7 billion euros retail market in 2014 with 900,000 customers in attendance of chain's 900 grocery stores every day. K-ruoka mobile app is being driven by its mission to make the daily lives of its grocery store customers easier with intelligent shopping tools [9]. A mobile app is futuristic but has a fundamental problem of short lifespan as the users delete it from the smart devices because of poor functionality and performance [10-11]. This study intends to compare four models for mobile retail app use and four models for mobile retail app continuous use based on mobile retail app technology, trust and gratification. The results of this research help to fill the gap that premature of mobile app use has created. The significant objective of this study is to examine how the combination of mobile retail app technology, trust and gratification model can extend the mobile retail app use and continuous use based on the optimised mobile app technology, increased trust confidence and embedded gratification elements and the sub-objective is to examine how to increase the mobile app user's autonomy and efficiency and to enrich the mobile app user's experience. To have a better understanding of the hypothesised theory and clearer perception of the model that have explanatory power, the study employs SPSS linear regression and general linear regression to look at the relationship of mobile app technology, trust and gratification predictors and the outcome variable. Further, it also explicates the managerial implication of mobile app and makes a necessary recommendation. The study is divided into five parts. The first section gives an overview of mobile app. The second section synthesizes applicable literature while section three explores the appropriate methodology for the study. Section four displays the data analysis and the result while the last section shows the theoretical and managerial implications with future study alertness.

2. Literature Review and Hypotheses

Based on UTAUT, Trust and Gratification model the researchers adopted two constructs include: Performance expectancy, and effort expectancy, social influence, facilitating condition, trust, security, privacy, cognitive, affection and tension free. These constructs have been proved in the literature as salient predictors for accepting technology. In this section, we define each of the constructs and their relationship.

2.1. Performance Expectancy

The level of belief by an individual that using the system will help to attain gains in job performance is known as Performance

expectancy. In literature, performance expectancy plays a significant role in the intention to adopt information technology as shown in study [12-14]. This model is a combination of previous ones, five factors from last model helped information of performance expectancy variable consisting of perceived usefulness (technology acceptance models), external motivation (motivational model), job fit (PC utilization model), relative advantages (innovation diffusion theory) and outcome expectations (social cognition theory) [15-17]. The perception that using a mobile app will be more useful and improve performance will determine its use and continuous use.

2.2. Effort Expectancy

Effort expectancy is defined as “the degree of ease associated with the use of the system” [16]. This construct is a combination of three constructs from the existing models as stated by researchers. Such as perceived ease of use from the study of [18] [15] [19] and ease of use from [20]. The perception of comfort in using the mobile app will determine its purpose and continuous use.

2.3. Social Influence

Reference [21] defined Social influence as the degree of impact on the interaction among people in the social network. It was further described by [22] as the perceived pressure gained to perform a specific behavior. Service experiences from technology use can be shared by people to form a collective basis for conversations within a social network. Social influence perspective has been found to be significant in the adoption of innovative product and services [23-24]. From a social perspective, we consider that social ties predict mobile retail app use and usefulness in the current study.

2.4. Facilitating Conditions

Facilitating conditions refers to the extent to which an individual perceives that the technical and organizational infrastructure required to use the proposed system are available [16]. According to [15-17], the definition covers constructs of perceived behavioral control (planned behavior theory and decomposed planned behavior theory), facilitating conditions (PC utilization model) and adaptability (innovation diffusion theory). Promoting Conditions is significantly related to technology use [25]. Technology-wise, there is a deep connection between PE, EE, SI and FC as an influencing factor of mobile retail app actual use and continuous use and the study hypothesized that:

H1: Performance expectancy, social influence and facilitating condition were predictors of mobile retail app use and usefulness.

H2: Performance expectancy and effort expectancy are predictors of mobile retail app continuous use.

2.5. Trust

According to [26], trust promotes transaction success because it can reduce social uncertainties that would otherwise be too complex, if not impossible, to figure out on a rational basis.

Online purchase requires consumer trust since consumers have to provide their personal information during the transaction process. [27] and [28], found trust to be a key predictor of both initial online purchase and repeat purchase. Trust has also been found to be significant in decision making for online transactions [29-30]. To obtain the necessary assurance, [31] opined that customers must depend more heavily on trust in the online vendor. [32] states that consumers rely on their trust in the vendor or the Internet to mitigate the effects of their uncertainty toward their relationships (as buyer and seller) in the online environment. In this study, it is proposed that users trust in the mobile app use will motivate its continuous use.

2.6. Security

Security threats have really been a point of concern in the online environment, threats such as fraudulent access or attack on consumer's mobile devices and online accounts. According to [33], in an internet context, security refers to the perceptions about safety regarding the means of payment and the mechanism for storing and transmission of information. [34] explains security as the consumer belief that their data will not be abused or their stored data cannot be modified by third parties without permission as data can only be seen by authorized individuals and certain actions can only be undertaken if proper authentication has taken place. [35] opined that an individual's perceived need for security should influence the perception of the usefulness of the device which is confirmed in the study of [36] that users show concern about unsecured websites. In this study, security concerns measure respondents' belief that security concerns of the mobile app will affect its use.

2.7. Privacy

The need for confidentiality of individuals becomes a significant challenge considering an automatic exchange of different personal information. Privacy concerns, especially through technology platform, must be addressed to maintain user control. Lessig defines privacy as the combination of "empowerment to control", "utility to protect", "dignity to establish an equilibrium", and "regulating agent to balance power" [37]. [38] opined that privacy is considered to exist when users of technology can control their personal information. Reference [39] found that users' perception of privacy assurance affects the use of a mobile platform. Thus, continuous use could probably be predicted by privacy confidence. Privacy assurance and a secured mobile retail app are antecedents of trust. The retail mobile app users will go an extra mile when they are assured that the mobile app they are using secured with privacy confidence. Regarding this proposition, the study hypothesized that:

H3: Trust and security perception are predictors of mobile retail app use and usefulness.

H4: Privacy confidence is a predictor of mobile retail app continuous use.

2.8. Cognitive

According to [40], cognitive factors related to mobile technology are important to examine because they influence the users' feelings about a technology. The cognitive phase represents

the conscious decisions regarding the behavioral purpose of serving the users' needs [41]. The cognitive comprises of some factors such as the perceived usefulness of the mobile app, confirmation of expectations, and contextual factors/characteristics of the system (i.e., perceived mobility, personalization, and responsiveness). These factors have been found to positively influence the use and continuous use of a system [42-45].

2.9. Affective

The behaviors that capture the personal feelings users have about an object that affects their behavior is known as affective [46]. According to [41], affect can exhibit positive or negative feelings about an object and provide an evaluation of the product and is considered an essential part of users' attitude. Zhang found that affect has a strong impact on decision-making behavior and consumer shopping behavior, which suggests that affect helps explain significant variance in one's cognition and behavior (Zhang, 2013). Affective factors have been found to be significant in IT use and continuance intention [45].

2.10. Tension Free

The current study examines the gratification among other constructs as an antecedent of mobile app continuous use in the retailing setting. The use and gratification theory posit that the use of media and technology is determined by individual users needs or motivations [47-48]. Uses and gratification research has highlighted consumers' hedonistic motives for using new communication technology, the need for entertainment, pleasure or enjoyment (Shin, 2007; Huang, 2008) [49-50]. Reference [51] sees gratification as the extent that the customers' needs are satisfied, while they assert that the stronger the degree of gratification, the higher the intention to use mobile apps. It was further stressed that some mobile app has gratification features that can calm tension while findings show that tension-free features on mobile apps positively affects the use of retailing mobile apps [51]. Mobile retail app with gratification elements will create an avenue for users learning, pleasure and tension calmness. It is therefore hypothesized that:

H5: Cognitive, affective, and tension free as an element of gratification were predictors of mobile retail app use and usefulness.

H6: Cognitive, affective, and tension free as an element of gratification were predictors of mobile retail app continuous use.

2.11. Actual Use

This construct was used in theories such as unified theory of acceptance and use of technology (UTAUT) by [18] which aims to explain user intentions to use an information system and subsequent usage behavior. Also, the Technology Acceptance Model (TAM) by [15], which is an information systems theory that models how users come to accept and use technology. Actual usage is used in the study as the actual use of the retail mobile app. Because of some antecedent behaviors, e.g. trust, privacy confidence in the platform and facilitating conditions.

2.12. Continuous Use

After the actual usage of a new technology or system, constant use is the next. This is the subsequent usage behavior of

a platform. At the level of continuous usage of a system, the antecedent of usage of that system comes to play, whereby experience and knowledge of the previous usage suggest constant use.

It is more profitable for the mobile retail app users to use the facilities embedded into the app and to enjoy it. This initial gratification, efficiency perception and assurance of security will motivate them to continue to use the retail app. Based on this view, the study hypothesized that:

H7: Combined elements of technology and gratification are predictors of mobile retail app use and usefulness.

H8: Combined elements of technology, trust and gratification were predictors of mobile retail app continuous use.

3. Research Design and Methodology

3.1. Sample and Data Collection

Purposeful sampling methodology was employed to gather the views and opinion of respondents online. This sampling method was used because only users of a mobile retail app in recent times or in the time past are the respondent target. Users that have used a mobile retail app to view products videos view current products new arrivals, shop and purchase products, received sales and coupon alerts, view products reviews and view products description, specifications and details. The total respondents that attended to the online questionnaire are 235. The data was subjected to a reliability test of Cronbach Alpha to ascertain the reliability of the instrument used and the results reach and above

the threshold of 0.7 with a minimum of 0.87 and a maximum of 0.96.

3.2. Measurement

In this study, items were adopted to measure the twelve latent variables from works of literature. Items Measuring Performance expectancy, Effort expectancy, Social influence, Facilitating conditions and Behavioral intention adopted from [16] user acceptance of information technology scale. Items Measuring Use behavior adopted from [2] mobile application usability scale). Items Measuring continuous use adopted from [52] expectation disconfirmation and technology adoption scale). Items Measuring cognitive, affective and tension free adopted from [53] Use and gratifications of mobile SNSs scale. Items Measuring tension free adopted from [54] the uses and gratifications of using Facebook music listening applications scale). Items Measuring affective adopted from [55] uses and gratifications theory and e-consumer behaviors scale. Items Measuring trust adopted from [56] addressing the personalization-privacy paradox scale. Items Measuring privacy and security adopted from [57] consumer trust, perceived security and privacy policy scale. The first section of the four parts of the instrument used to elicit information features the demography detail of the mobile app users followed by mobile app users experience. The next section extracts information on mobile app usage and the fourth section elicit a question from respondents on seven Likert Scale type question from strongly disagree (1) to strongly agree (7) based on the theories and works of literature reviewed.

Table 1: Regression Analysis result for retail mobile app use

Models (AU)	β	SE	St.β	T	R	R ²	Adj. R ²	AIC	SE	F
RMA Technology										
Constant	-.46	.240		-1.91	.824	.679	.674	559.8	.786	162.6
PE	.74	.046	.684	16.03						
SI	.20	.050	.166	3.98						
FC	.13	.040	.125	3.12						
RMA Trust										
Constant	1.64	.256		6.41	.540	.292	.286	743.5	1.165	47.75
TR	.366	.099	.354	3.68						
SE	.204	.093	.210	2.19						
RMA Gratification										
Constant	.492	.200		2.459	.778	.605	.600	608.1	.871	118.1
CO	.458	.062	.433	7.44						
AF	.275	.086	.263	3.19						
TF	.157	.075	.163	2.11						
Combined Models										
Constant	-.53	.222		-2.37	.848	.719	.714	530.9	.737	146.8
PE	.580	.051	.539	11.27						
FC	.107	.038	.108	2.82						
CO	.316	.051	.299	6.18						
Continent	.127	.044	.105	2.90						

***<0.0001. **0.001 <p ≤0.01, *0.01 <p ≤ 0.05

MRA: Retail Mobile App β: Beta, SE: Standard Error, SE: Standard error,

St.β: Standardized beta, SE: Standard error, T: T-test

R²: Coefficient of determination, F: F-test.

Dependent variable: AU – App use, Predictors: PE: Performance Expectancy, SI: Social Influence, FC: Facilitating Condition, TR: Trust, SE: Security, CO: Cognitive, AF: Affective

Table 2: Regression Analysis result for retail mobile app continuous use

Models (CU)	B	SE	St.β	T	R	R ²	Adj. R ²	AIC	SE	F
RMA Technology										
Constant	.163	.26		.63	.78	.61	.60	618.7	.893	178.9
PE	.795	.05	.72	14.7						
EE	.121	.06	.10	2.1						
RMA Trust										
Constant	1.912	.24		8.1	.55	.31	.30	750.4	1.18	102.4
PR	.545	.05	.55	10.1						
RMA Gratification										
Constant	.677	.21		3.2	.76	.58	.57	637.2	.927	105.5
CO	.361	.07	.33	5.5						
AF	.346	.09	.32	3.8						
TF	.187	.08	.19	2.4						
Combined Models										
Constant	-.072	.21		-3.5	.82	.68	.67	537.0	.816	119.3
PE	.497	.07	.45	7.6						
PR	.124	.05	.13	2.7						
CO	.180	.06	.17	2.9						
AF	.221	.07	.21	3.4						

***≤0.0001. **0.001 <p ≤0.01, *0.01 <p ≤ 0.05

RMA: Retail Mobile App, β: Beta, SE: Standard Error, SE: Standard error, St.β: Standardized beta, SE: Standard error, T: T-test,

R²: Coefficient of determination, F: F-test.

Dependent variable: CU – Continuous use, Predictors: PE: Performance Expectancy, SI: Social Influence, FC: Facilitating Condition, TR: Trust, SE: Security, CO: Cognitive, AF: Affective

4. Results

4.1. Multiple Regression Model

The study used quantitative data to run multiple regression analysis with keen attention to linear regression assumption. The study used SPSS 24 version to ensure that the prediction errors are independent over cases, follow a normal distribution, lack heteroscedasticity and have a constant variance (homoscedasticity) and the relationship among the variables are linear. In order not to violate the assumptions, the study undertakes the following steps. We examined the factorability of the 48 items and used several criteria. Initially, 45 items correlated with other items (.215 - .841), indicating reasonable factorability. Secondly, the Kaiser-Meyer-Olkin measure of sampling adequacy was .94, above the rule of thumb of .6, and Bartlett's test of sphericity was significant ($\chi^2(990) = 11678.29, p < .001$). Finally, the commonalities were all above .5, confirming that each item shared some common variance with other items. Based on the overall indicators, factor analysis was conducted with 45 items.

The study variables did not contain any system missing values and the frequency distributions look plausible. The descriptive statistics test for mean was (2.91 – 5.39) and standard deviation (1.29 – 1.64). The plotted scatterplots of the predicted values (x-axis) with the outcome variables (y-axis) did not show any clear curvilinearity. Cronbach Alpha ranges between (.972 - .974) and the variance inflation factors (VIFs) are lower than 10 (Myers, 1990) [58], the tolerances greater than 0.2 (Menard, 1995) [59] and the condition index less than 30. In all, the chart plotter, and the data analysis did not show any violation of the independence, homoscedasticity and linearity assumptions.

The author run linear regression analysis for four models for technology, trust and gratification predictors as against mobile retail app use and continuous use. All the predictors of mobile retail app (Model 1) were significant, performance expectancy → mobile app use ($\beta=0.68, t=16.03, P \text{ Value} = <.001$), social influence → mobile app use ($\beta=0.17, t=3.98, P \text{ Value} = <.05$), facilitating conditions → mobile app use ($\beta=0.13, t=3.12, P \text{ Value} = <.05$). Mobile retail app trust (Model 2) trust → mobile app use, ($\beta=0.35, t=3.68, P \text{ Value} = <.05$), security → mobile app use, ($\beta=0.21, t=2.19, P \text{ Value} = <.05$). Mobile retail app gratification (Model 3) cognitive → mobile app use, ($\beta=0.43, t=7.44, P \text{ Value} = <.05$), affective → mobile app use, ($\beta=0.26, t=3.19, P \text{ Value} = <.05$), tension free → mobile app use, ($\beta=0.16, t=2.11, P \text{ Value} = <.05$). Combined models (Model 4) performance expectancy → mobile app use, ($\beta=0.54, t=11.27, P \text{ Value} = <.001$), facilitating conditions → mobile app use, ($\beta=0.11, t=2.82, P \text{ Value} = <.05$), cognitive → mobile app use, ($\beta=0.30, t=6.18, P \text{ Value} = <.05$), continents → mobile app use, ($\beta=0.11, t=2.90, P \text{ Value} = <.05$).

For continuous use all the predictors of mobile retail app (Model 1) were significant, performance expectancy → mobile app continuous use ($\beta=0.72, t=14.67, P \text{ Value} = <.001$), effort expectancy → mobile app continuous use ($\beta=0.10, t=2.11, P \text{ Value} = <.05$). Mobile retail app trust (Model 2) privacy → mobile app continuous use, ($\beta=0.55, t=10.12, P \text{ Value} = <.05$). Mobile retail app gratification (Model 3) cognitive → mobile app continuous use, ($\beta=0.33, t=5.51, P \text{ Value} = <.05$), affective → mobile app continuous use, ($\beta=0.32, t=3.77, P \text{ Value} = <.05$), tension free → mobile app continuous use, ($\beta=0.19, t=2.35, P \text{ Value} = <.05$). Combined models (Model 4) performance expectancy → mobile app continuous use, ($\beta=0.45, t=7.64, P$

Value = <.001), privacy → mobile app continuous use, ($\beta=0.13$, $t=2.71$, P Value = <.05), cognitive → mobile app continuous use, ($\beta=0.17$, $t=2.91$, P Value = <.05), affective → mobile app continuous use, ($\beta=0.21$, $t=3.40$, P Value = <.05).

4.2. Model Comparison

The study model is multistage indicating the need for general linear regression analysis. We conducted general linear analysis for model comparison, and we extracted the value of Akaike Information Criterion (AIC) and compared it with the coefficient of determination. The first model of mobile retail app (mobile retail app technology) revealed AIC = 559.76, $R^2 = 67.9\%$, model 2 (mobile retail app trust) AIC = 743.52, $R^2 = 29.2\%$, model 3 (mobile retail gratification) AIC = 608.12, $R^2 = 60.5\%$, and model 4 (combined models) AIC = 530.93, $R^2 = 71.9\%$. According to the rule of thumb of linear model assessment, the lower the AIC, the better and the higher the coefficient of determination the better. In this study, the combined model of technology, gratification and continent had the lowest AIC and the highest R^2 .

Similarly, to mobile retail app use, the continuous use first model of mobile retail app (mobile retail app technology) revealed AIC = 618.66, $R^2 = 60.7\%$, model 2 (mobile retail app trust) AIC = 750.42, $R^2 = 30.5\%$, model 3 (mobile retail gratification) AIC = 637.20, $R^2 = 57.8\%$, and model 4 (combined models) AIC = 537.02, $R^2 = 67.5\%$. The combined model of technology, trust and gratification had the lowest AIC and the highest R^2 .

4.3. Moderation Effects

A two-way analysis of variance was conducted on the influence of two independent variables (gender, marital status) on the age of mobile retail app users. Marital status included four levels (single, married, cohabitation, divorcement) and gender consisted of two levels (male, female). All effects were statistically significant at the .05 significance level except for the gender factor. The main effect for marital status yielded an F ratio of $F(3, 227) = 57.7$, $p < .001$, indicating a significant different between male ($M = 2.35$, $SD = 1.32$), female ($M = 2.05$, $SD = 1.17$). The main effect for gender yielded an F ratio of $F(1, 227) = 0.004$, $p > .005$, indicating that the main effect for gender was not significant. The interaction effect was significant, $F(3, 227) = 3.62$, $p < .014$. In this case, the effect for gender interacts with marital status. That is, age affects females differently than males.

5. Discussion, Theoretical and Practical Implications

The goal of this study was to examine the impact of combined model of a mobile retail app on mobile retail app use and continuous use based on the optimised mobile app technology, increased trust confidence and embedded gratification elements and how the integrated model can enhance mobile app user's experience. Using SPSS, linear regression and general linear regression statistical technique were used.

Eight hypotheses were tested and supported. Out of the eight hypotheses, four focused on mobile retail app use and usefulness while the other four dwells on mobile retail app continuous use. Performance expectancy being the technology acceptance

variable was the highest predictor of mobile app use in the first model. In addition to performance expectancy, social influence and facilitating conditions were found significant as predictors of mobile app use. There is an influence of the third parties such as mentors, relatives, and the retailers on the retailing customers to use a mobile retail app but the social influence variable was not significant as a predictor of mobile retail app continuous use. In model two, trust and security are the predictors of mobile app use while three elements, cognitive, affective and tension free of gratification predicted mobile retail app and cognitive was the highest predictor. To get more insight from the study, the study combined the three models of technology, trust and gratification and performance expectancy, facilitating conditions, cognitive and continent path coefficient with the mobile retail app were significant. Respondents that lived in Finland from Europe, Asia and Africa participated in the study, and the result shows the neutral response of the Europeans to the mobile retail app while Asia and Africa's response supported the use of mobile retail app. The Europeans users are indifferent while the Asia and Africa users in Finland are enthusiastic. Performance expectancy still maintained the highest predictor of a mobile retail app. Extant studies emphasised the role of performance expectancy in technology use [60-61].

Unlike mobile retail app use model one, only two variables are the predictors of mobile retail app continuous use and effort expectancy that was not found significant in mobile retail app use was substantial in a constant model. This can be explained that mobile retail app users have been using for a long period of time and its use is not cumbersome to the users any more. Only privacy confidence was significant in model two and it is an indication that privacy assurance will prolong the mobile retail app continuous use. Like the mobile retail app model three all the gratification elements were significant and coincidentally, cognitive was the highest predictor of mobile retail app continuous use. This signifies the importance of learning as the retailers add more features to their mobile app, there will be a need for the mobile retail app to learn how to use the new features either through a video or text instructions. The model four showcase performance expectancy, privacy, cognitive and affective as a predictor of mobile retail app continuous use but continent was not significant. In all the model, the mobile retail use and continuous combined model of technology, trust and gratification were found more robust than others. In addition to model comparison, the interaction effects of gender and marital status reveals that different age brackets influence gender differently based on four levels of marital status.

A mobile retail app is a growing innovation in a retailing setting, and theoretically, the study integrates the unified theory of acceptance and use of technology (UTAUT), trust and gratification and expand technology acceptance model with the trust and gratification elements. The study emphasized the importance of trust, privacy assurance, learning and relaxation features in a mobile retail app as an antecedent of its use and continuous use. This is a novel contribution to the literature on technology acceptance and retailing. This study also shed more light on the importance of age as a moderator of gender and marital status regarding mobile retail app use and continuous use.

Practically, the retailing managers and the technology professionals need to put age, gender, marital status and people from different continents into consideration when they are strategizing for the mobile retail app use and continuous use. The add-on technology, hedonic and trust features should factor in the demographic profile of the mobile retail app. Second, the retailer should reposition their mobile app as a multitasking app of an information database for online e-shopping, gamification for rewards and video curation for socialization.

5.1. Limitations and Future Research

The mobile app market is vast, and this study only focuses on retailing segment which may not represent the state-of-the-art of mobile app market. Due to this limitation, the future researcher in the mobile app research stream should conduct a mobile app comparative study with a focus on countries and different business sectors. The prospective studies also should use structural equation modelling (SEM) to compare different mobile app users intrinsic and extrinsic motivation.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] R.N. Gerlich, K. Drumheller, J. Babb, D. Armond, "App consumption: An exploratory analysis of the Uses & Gratifications of mobile apps" *Academy of Marketing Studies Journal*, 19(1), 69, 2015.
- [2] H. Hoehle, V. Venkatesh, "Mobile Application Usability: Conceptualization and Instrument Development." *MIS Quarterly*, 39(2), 2015.
- [3] C.Z. Liu, Y.A. Au, H.S. Choi, "Effects of freemium strategy in the mobile app market: an empirical study of Google play" *Journal of Management Information Systems*, 31(3), 326-354, 2014.
- [4] A. Coleman, "Behind Sumoing's App Store Success: Finland's Mobile Tech Legacy" 2016. Available at: <https://www.forbes.com/sites/alisoncoleman/2016/04/24/behind-sumoings-app-store-success-finlands-mobile-tech-legacy/#385b8077668c>. Cited on 21.12.2017.
- [5] R. Garg, R. Telang, "Inferring app demand from publicly available data". *MIS Quarterly*, 37(4), 1253-1264, 2013.
- [6] H.C. Yang, "Bon Appétit for apps: young American consumers' acceptance of mobile applications" *Journal of Computer Information Systems*, 53(3), 85-96, 2013.
- [7] T. Wang, T.D. Duong, C.C. Chen, "Intention to disclose personal information via mobile applications: A privacy calculus perspective" *International Journal of Information Management*, 36(4), 531-542, 2016.
- [8] D. Munford, "S Group innovates with mobile payment technology at the petrol pump" 2017. Available at: <https://www.linkedin.com/pulse/group-innovates-mobile-payment-technology-petrol-pump-dan-munford/>. Cited on 21.12.2017.
- [9] M. Kakko, "Mobile app makes customers' daily lives easier – K-ruoka" 2017. Available at: <https://futurice.com/cases/k-ruoka-mobile>. Cited on: 21.12.2017.
- [10] eMarketer, "Why Consumers Download, and Delete, a Retailer's Mobile App, Promos and rewards drive downloads" 2016. Available at: <https://www.emarketer.com/Article/Why-Consumers-Download-Delete-Retailers-Mobile-App/1014212>. Cited on: 21.12.2017.
- [11] Insights, "Retail Apps: Why Consumers Download, Use, and Delete" 2017. Available at: <http://newsroom.synchronyfinancial.com/document-library/retail-app-engagement>. Cited on: 21.12.2017.
- [12] I. Benbasat, H. Barki, "Quo vadis TAM?" *Journal of the Association for Information Systems*, 8(4), 212-218, 2007.
- [13] L. Carter, L.C. Schaupp, M.E. McBride, "The U.S. E-file initiative: An investigation of the antecedents to adoption from the individual" *E-service Journal*, 7(3), 2-19, 2011.
- [14] M.N. Alraja, "User acceptance of information technology: A field study of an e-mail system adoption from the individual students' perspective" *Mediterranean Journal of Social Sciences*, 6(6), 19-25, 2005.
- [15] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models". *Management Science*, 35(8), 982-1002, 1989.
- [16] V. Venkatesh, M.G. Morris, G.B. Davis, F.D. Davis, "User Acceptance of Information Technology: Toward A Unified View" *MIS Quarterly*, 27(3), 425 – 478, 2003.
- [17] V. Venkatesh, F.D. Davis, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies" *Management of Science*, 46(2): 186-204, 2000.
- [18] F.D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". *MIS Quarterly*, 13(3), 319-339, 1989.
- [19] R.L. Thompson, C.A. Higgins, J.M. Howell, "Personal Computing: Toward a Conceptual Model of Utilization" *MIS Quarterly*, 15(1), 124-143, 1991.
- [20] G.C. Moore, I. Benbasat, I. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation" *Information Systems Research*, 2(3), 192-222, 1991.
- [21] R.E. Rice, C. Aydin, "Attitudes toward new organizational technology: network proximity as a mechanism for social information processing. *Adm. Sci. Q.* 36 (2), 219–244, 1991.
- [22] V. Venkatesh, S.A. Brown, "A longitudinal investigation of personal computers in homes: adoption determinants and emerging challenges". *MIS Quarterly*, 25 (1), 71–102, 2001.
- [23] O. Turel, A. Serenko, N. Bontis, "User acceptance of wireless short messaging services: deconstructing perceived value" *Inf. Manag.* 44 (1), 63–73, 2010.
- [24] V. Venkatesh, J.Y.L. Thong, X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS Quarterly* 36 (1), 157–178, 2012.
- [25] G.E. Heilman, G.A. Johnson, G.O. Seshie, B. Greene, "Perceived effect of facilitating conditions on college student computer use" *Journal of Academy of Business and Economics*, 9(2), 148-151, 2009.
- [26] N. Luhmann, *Trust and Power*, London: John Wiley & Sons, 1979.
- [27] C. Flavian, M. Guinaliu, R. Gurrea, "The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty," *Information & Management*, 43, 1-14, 2006.
- [28] I. Qureshi, Y. Fang, E. Ramesy, P. McCole, P. Ibboston, D. Compeau, "Understanding Online Customer Repurchasing Intention and the Mediating Role of Trust: An Empirical Investigation in Two Developed Countries" *European Journal of Information Systems*, 18(3),205-222, 2009.
- [29] P.A. Pavlou, "Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model" *Int. J. Electron. Commer.* 7(3), 101–134, 2003.
- [30] B.J. Corbitt, T. Thanasankit, H. Yi, "Trust and e-commerce: a study of consumer perceptions". *Electron. Commer. Res. Appl.* 2(3),203–215, 2003.
- [31] C.L. Corritore, B. Kracher, S. Wiedenbeck, "On-Line Trust: Concepts, Evolving Themes, a Model," *International Journal of Human-Computer Studies* 58(6), 737-758, 2003.
- [32] S. Ha, L. Stoel, "Promoting customer-retailer relationship building: Influence of customer trustworthiness of customer loyalty programme marketing" *Journal of customer behaviour* 7(3), 215–229, 2008.
- [33] A. Kolsaker, C. Payne, "Engendering trust in e-commerce: a study of gender-based concerns", *Marketing Intelligence and Planning*, 20(4), 206-14, 2002.
- [34] M. Mandić, "Privacy and Security in E-Commerce". *TRŽIŠTE*, 21(2), 247 – 260, 2009.
- [35] T. James, T. Pirim, K. Boswell, B. Reithel, R. Barkhi, "An Extension of the Technology Acceptance Model to Determine the Intention to Use Biometric Devices" In Ed. Steve Clarke *End User Computing Challenges and Technologies: Emerging Tools and Applications*, IGI Global, 57-78, 2008.
- [36] M.H. Shah, R. Okeke, R. Ahmed, "Issues of Privacy and Trust in E-Commerce: Exploring Customers' Perspective" *Journal of Basic Applied Science Research*, 3(3), 571-577, 2013.
- [37] L. Lessig, *Code and other laws of cyberspace*. Basic books, 1999.
- [38] H. McCloskey, "Privacy and the right to privacy" *Philosophy*, 55(211), 17-38, 1980.
- [39] S. Pearson, "Taking account of privacy when designing cloud computing services" In *Software Engineering Challenges of Cloud Computing. CLOUD'09. ICSE Workshop*, 44-52, 2009, IEEE.
- [40] M.B. Holbrook, "Beyond attitude structure: Toward the informational determinants of attitude" *Journal of Marketing Research*, 545-556, 1978.
- [41] L. Cobos, "Determinants of continuance intention and word of mouth for hotel branded mobile app users. *Electronic Theses and Dissertations*. 5719, 2017. <http://stars.library.ucf.edu/etd/5719>
- [42] A. Bhattacharjee, G. Premkumar, "Understanding changes in belief and attitude toward information technology usage: a theoretical model and longitudinal test" *MIS Quarterly*, 229-254, 2004.

- [43] A. Bhattacharjee, C.P. Lin, "A unified model of IT continuance: three complementary perspectives and crossover effects" *European Journal of Information Systems*, 24(4), 364- 373, 2015.
- [44] Z. Zhong, J. Luo, M. Zhang, "Understanding antecedents of continuance intention in mobile travel booking service" *International Journal of Business and Management*, 10(9), 156, 2015.
- [45] A. Nabavi, M.T. Taghavi-Fard, P. Hanafizadeh, M.R. Taghva, "Information Technology Continuance Intention: A Systematic Literature Review" *International Journal of E-Business Research (IJEER)*, 12(1), 58-95, 2016.
- [46] I. O. Pappas, P.E. Kourouthanassis, M.N. Giannakos, V. Chrissikopoulos, "Explaining online shopping behavior with fsQCA: The role of cognitive and affective perceptions. *Journal of Business Research*, 69(2), 794-803, 2016.
- [47] J.G. Blumler, E. Katz, "The uses of mass communications: current perspectives on gratification research". Sage publication, Beverly Hills, CA, 1974.
- [48] C.A. Lin, "Looking back: The contribution of Blumler and Katz's uses of mass communication" *Journal of Broadcasting and Electronic media*, 40(4), 574-681, 1996.
- [49] D. Shin, "User acceptance of mobile internet: Implications for convergence technologies" *Interacting with Computers*, 19(4), 472-483, 2007.
- [50] E. Huang, "Uses and Gratification in e-consumers" *Internet Research*, 18(4), 405-426, 2008.
- [51] S.A. Olaleye, J. Salo, I.T. Sanusi, A. Okunoye, "Retailing Mobile App Usefulness: Customer Perception of Performance, Trust and Tension Free. *International Journal of E-Services and Mobile Applications (IJESMA)*. 10(4), 1-17, 2018.
- [52] V. Venkatesh, S. Goyal, "Expectation disconfirmation and technology adoption: polynomial modeling and response surface analysis" *MIS quarterly*, 281-303, 2010.
- [53] Y.W. Ha, J. Kim, C.F. Libaque-Saenz, Y. Chang, M.C. Park, "Use and gratifications of mobile SNSs: Facebook and KakaoTalk in Korea". *Telematics and Informatics*, 32(3), 425-438, 2015.
- [54] A. E. Krause, A. C. North, B. Heritage, "The uses and gratifications of using Facebook music listening applications". *Computers in Human Behavior*, 39, 71-77, 2014.
- [55] Y. Lu, W. Beck. Do you see what I see? Infants' reasoning about others' incomplete perceptions. *Developmental Science*, 13(1), 134-142, 2010.
- [56] J. Sutanto, E., Palme, C.H. Tan, C.W. Phang, "Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users". *MIS Quarterly*, 37(4), p. 1141-1164, 2013.
- [57] C. Flavián, M. Guinalú, "Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site". *Industrial Management & Data Systems*, 106(5), 601-620, 2006.
- [58] R.H. Myers, "Classical and modern regression with applications, 2nd edition. PWS Kent, Boston, MA, 1990.
- [59] S. Menard, *Applied logistic regression analysis. Quantitative applications in the social sciences*, No. 106. Thousand Oaks, CA & London: Sage, 1995.
- [60] T. Zhou, "Understanding mobile Internet continuance usage from the perspectives of UTAUT and flow. *Information Development*, 27(3), 207-218, 2011.
- [61] T. Escobar-Rodríguez, E. Carvajal-Trujillo, "Online purchasing tickets for low cost carriers: An application of the unified theory of acceptance and use of technology (UTAUT) model" *Tourism Management*, 43, 70-88, 2014.

cv4sensorhub – A Multi-Domain Framework for Semi-Automatic Image Processing

Kristóf Csorba*, Ádám Budai

Budapest University of Technology and Economics, Department of Automation and Applied Informatics, 1117 Budapest, Magyar Tudósok krt. 2/Q, Hungary

ARTICLE INFO

Article history:

Received: 08 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

Software framework

Image processing

Microscopy images

Grain boundaries

Cell tracking

ABSTRACT

Although there are many research domains with very good software support and workflow automation, there are even more which do not have it: software development is too expensive to create domain specific applications for every research topic. This leaves many domain experts to work for example with general purpose image processing and statistics tools. Many research ideas get even omitted as requiring unfeasible much manual work. This paper presents a multi-disciplinary image processing software framework called cv4s (cv4sensorhub). Its aim is to create an environment where reusable components make development of domain specific image processing software solutions easier, and thus, more feasible. The paper presents the basic architecture of the framework and two example applications: GrainAutLine which is for analysis of microscopy images of marble, sandstone and schist thin sections, and ChemoTracker which is designed for the motion analysis of white blood cells. As many image processing operations are relatively domain independent, the possible application areas are not limited to petrographic and medical images processing.

1. Introduction

Image processing is an area of computer science with a very wide range of possible application areas. Beside the classic directions there are lots of interdisciplinary application areas where the software supported workflow automation is much less common and the domain specific software development is too expensive to be feasible. In order to apply state-of-the-art image processing in these areas, a software framework with the following properties is needed:

1. The framework has to provide reusable software components, so that domain specific software solutions can be created much faster and thus cheaper, than by starting from scratch. It has to provide the necessary elements for the rapid integration of state-of-the-art image processing methods, and an environment for convenient R&D experimentations.
2. It has to provide a user interface to allow professionals with less IT related skills to test and use the resulting techniques. As we cannot expect these professionals to get used to a scientific computing environment and its user interface, we need to provide a state-of-the-art user interface similar to common applications.

3. The framework needs to emphasize the support for semi-automatic image processing which involves significant user interaction into the image processing operations. This is essential in many application domains where either very high accuracy is required, or the users simply do not necessarily trust a fully automatic solution. This means that many operations have to be designed so that the user has full control over the results: all automatic suggestions can be overridden by the user before finalizing it.
4. Both the underlying techniques and the user interface has to be highly customizable, so that they can be accommodated to the application domain. This includes for example hiding irrelevant functions, or to adapt visualization and color schemes to other well-known expert systems.

There are interdisciplinary areas with enough resources for large-scale software development projects like many medical image processing tasks and geo-information (GIS) systems. But there are lot more which do not have. The goal of our cv4s project (<http://bmeaut.github.io/cv4sensorhub/>) is (1) to create a framework meeting the requirements mentioned above, and (2) to create domain specific applications for interdisciplinary areas where the software-based process automation could have very significant impact.

* Corresponding Author: Kristóf Csorba, Budapest University of Technology and Economics, Hungary Email: kristof@aut.bme.hu
www.astesj.com
<https://dx.doi.org/10.25046/aj030620>

In this paper we present the core architecture and features of the framework, and two domain specific applications built over it. The first application is GrainAutLine aiming for the semi-automatic segmentation and analysis of marble, sandstone, and schist thin section images. The second is ChemoTracker aiming for the semi-automatic tracking of white blood cells in microscopy image sequences.

Both applications emphasize the high accuracy requirements mentioned before: the results of the automatic operations are only suggestions which can always be checked, and if necessary, modified by the user. During the research and development of the application, it shows a smooth transition from the initially almost fully manual operation to a very high level of automation needing very few user interaction. But the user still has the option to fall back to the manual operation if necessary, for example due to bad imaging conditions.

The research questions this paper focuses to answer are the following:

- (1) Is the proposed data model suitable for creating software support for multiple domains? If it is, it allows creating a set of reusable software components which are easy (and thus fast) to build upon, making the process automation feasible.
- (2) How far can the involvement of software engineers improve the workflow in terms of speed, result quality and reliability. This question is also important from a research planning and management point of view, as many research topics located far from software engineering could significantly benefit from it.

In the remaining parts of the paper, after an overview of related work, first we present the core functionality and architecture of the framework. Then we show the way GrainAutLine and ChemoTracker was created by customizing and extending it, and finally, conclusions are summarized.

2. Related work, levels of software support

Previous work related to this project can be observed from two directions: the theoretical research results related to the tasks the GrainAutLine and ChemoTracker applications aim for, and the already existing software solutions available for these problems.

From the theoretical side, both of these applications perform image segmentation which is a very rich and well researched topic [1]. GrainAutLine extends this with shape recognition [2] and ChemoTracker applies motion tracking [3].

Starting from the application side, the options are much more limited: if a researcher who's domain is far from information technology needs image processing, the most common option is to use a general purpose image manipulation program like PhotoShop and Gimp. As these software are not specialized in research domains like material or medical technology, their applicability is limited. It often needs time consuming manual steps and corrections in the workflow, even if some important steps are automated.

The next level is achieved if one can use a more domain specific software package, like the one of some microscope manufacturers who provide software solutions with their devices.

In this case, many steps can be automated, except the ones too specific to the current research. For example if someone needs to count the number of neighbors for every grain, and that is not supported by the software, the solution gets problematic.

The next step is achieved if the researcher can use the scripting features of highly customizable (or programmable) tools like ImageJ (<https://imagej.net/>). In these cases, significant part of the process can be customized and the required amount of programming skills can be learnt relatively fast. It should be noted that ImageJ's manual tracking plugin was used before ChemoTracker has been developed.

The level of software support we are investigating in our research is the result of the integration of two teams: the one of the target domain and a software engineering team. Although learning to speak the language of the other and cooperate smoothly takes time, the two sides will learn what is easy, hard, important or useless for the other side. If it works, the domain experts can tell what they need, and the software engineers make the software support it as much as possible. End-to-end workflow support means for example from downloading the images from the microscope to generating the reports and diagrams ready for publication. This goal is what the eScience Center in the Netherlands is also aiming for (<https://www.esciencecenter.nl/>).

3. Core functionality

The cv4s framework is a .NET based system using the OpenCV [4] computer vision library, a stand-alone component of the SensorHUB [5] data collection and processing system. The architecture of the system is visualized in Fig. 1. The core data model is modified by the operations and visualized by common user interface (UI) functions. Derived data is a cache of information derived from the core data model and frequently used by operations. The communication and persistence component is responsible for supporting several input and output, various reporting formats. The topmost layer is the domain specific application built over the reusable components.

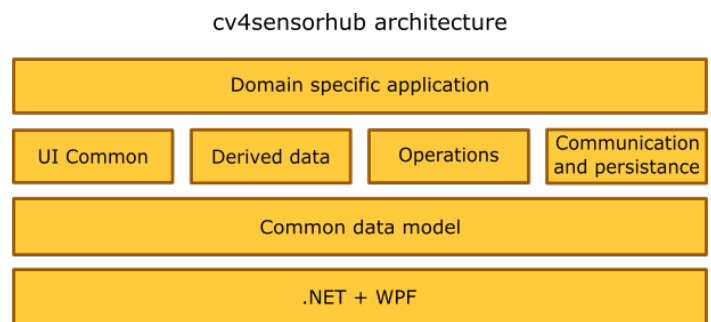


Figure 1: Layers of the cv4sensorhub framework.

3.1. Data model

The data model is organized around a set of entities with the following subtypes: polygons, polylines, single points, and raster images. Raster images are usually used as the original input, and the result usually consists of a combination of polygons and further information attached to them as tags. The tagging system allows adding key-value pairs to any entity. A simple tag like "selected"

may have irrelevant value, but tags like “ID”, “groupID”, or “parentID” utilize the value of the key-value pair as well. As the tags are heavily used by the operations, the system provides a tag based indexing to accelerate the search for entities with given tags or tag-value pairs.

Based on our experiences, this data model is sufficient to describe all necessary information we had to manage so far. Beside the input as raster images and the segmentation results as polygons, the tagging system allows for the representation of the following:

- Simple selection of entities either by the user or by an operation via adding the “selected” tag.
- Hierarchies using unique “ID” tag values for every entity and “parentID” tags to store the relationships. This is a way for example to match the occurrences of the same cell in a sequence of microscopy images.
- Entity groups by using “groupID” tags, like grains with same type, or cells with similar motion speed.

3.2. Derived data

There are several information which can be derived from an entity and may be often needed by the operations. Derived data are data collections associated with entities which are cached and kept up-to-date whenever the corresponding entity changes. For example a derived data set contains the up-to-date bounding box for every entity. The most important features are the following:

- Bounding boxes
- Bounding box distances: Manhattan distance between the bounding boxes of every entity pair. This can be used to quickly retrieve all entities in the proximity of a given one.
- Distance maps: the distance transform results for the image of the corresponding entity. This allows very fast retrieval of the Euclidean distance between a given point and the nearest point of the entity.
- Exact distances: Euclidean distance between the nearest points of every two entities. This can be used to retrieve all neighbors of a polygon within a given distance limit.
- Neighbors: the list of direct neighbors for all entities. This can be used for example to retrieve the number of grain neighbors.

3.3. Operations

A significant set of operations is related to directly manipulating the entities from the user interface, while the remaining ones offer more complex operations. The most important operations for editing the entities:

- Polygon and polyline drawer and modifier operations allowing direct modifications. They are designed for easy and fast usage: a single mouse stroke across polygons allows merging or slicing them. If the automatic image processing can guess most boundaries correctly, there is no need to manually mark them, the user only needs to quickly spot the mistakes and give the software a hint to fix that.
- Mover operation allows the movement of a whole entity, for example an inaccurate detection of a blood cell.

- A generic tool to change the tags of a given entity. In a customized application, this can be used for example as a grain selection tool by manipulating a “selected” tag.
- Z-Index modifier: as described later, the visualization uses the “ZIndex” tag to decide which polygons are above the others. This tool allows moving a polygon to the front or to the back.

These were the most common user interface tools. Some more complex ones are the following:

- Edge detector: an image segmentation tool using the Canny[6] edge detector to create new polygons.
- Polygon simplifier: simplifies the boundary of the selected polygons by removing noise.
- Color detection based on the histogram of some manually marked areas. This can be used for example to detect painted porosities in sandstones or concrete sections.
- Smart region growing to extend polygons to cover whole grains, even in the presence of textures.
- Thresholding and adaptive thresholding functions.
- Superpixel segmentation to locate small areas with homogeneous color.
- Stitching and other automatic alignment tools to handle series of images and automatically fix their displacements or other alignment errors.

Beside these, the load and save persistence functions are also implemented as operations. For convenient integration with other software solutions, vector image (SVG) export is also available and Shapefile (SHP) export for GIS systems is under development. The latter is implemented because many geosciences use GIS systems and if the image analysis can export its results into such Shapefiles, those can be imported into GIS systems opening up their wide range of shape and topology analysis tools.

The above mentioned operations are provided by the framework as general purpose operations. Customizations for given application domains can define their own tags and additional operations as needed.

3.4. Common user interface functions

Beside the common data model, the cv4s framework provides user interface components which can be used as the central element of customized user interfaces.

The most important user interface element is the viewer which takes care of the visualization of the whole data model. It is meant to be the central editor area of the applications. It allows the operations to capture mouse and keyboard events, and features a customizable visualization of the entities based on their assigned tags, in terms of the appearance engine.

The appearance engine maintains a list of appearance commands which can modify the appearance of an entity like fill color, stroke color and thickness, opacity etc. For example an appearance command can instruct the viewer to set the fill color of every entity having the “selected” tag green. From that point on, as soon as an entity gets this tag as a result of any operation, it

immediately turns green. (Appearance commands have priorities to resolve contradicting instructions.) Using the appearance commands, customizations can modify the appearance of entities according to any tags or tag-value pairs, like changing the fill color based on the number of neighbors, marking concave grains, or distinguishing fast moving cells.

Beside the viewer, the framework provides two additional user interface components: one to select and configure the operations to be applied, and one to configure the appearance commands.

If the proposed data model is suitable for a specific domain, then with these components, the developer of a new domain specific application does not have to care about several user interface functions like visualization and manual corrections. They can focus on the special requirements of the domain which makes the development much easier and faster.

3.5. Customization

The primary objective of the framework is reusability and flexibility. The main steps required when creating a new, domain specific application are the following:

- The new applications user interface is built around the viewer control. The operation selection and configuration control is often replaced by a menu bar or ribbon, and the appearance command editor is rarely required as these settings are hardwired in a domain specific setup designed for the end users.
- Used tag names and the corresponding appearance commands are created.
- Domain specific operations not provided by the framework are created. This is the biggest part of the development.
- As the output format of the applications is very different, the framework provides a basic support for statistics and data exporting, but most of these functions should be implemented in the domain specific customizations, usually also in the form of operations.

The plug-in architecture of the operations allows the developers to enhance and experiment with the operations in a very flexible way. Operations created for a given domain but having potential in other domains as well are often merged into the framework as reusable components, so that later projects are even faster to implement.

4. Domain specific customizations and extensions

After introducing the cv4s framework, in this section we present two domain specific applications as customizations and extensions of the functions mentioned before. These applications are now ready to be used by end users which allows us to provide preliminary results obtained using them.

4.1. GrainAutLine

Marble provenancing is an important task in archaeometry: identifying the mine a marble sample originates from is used both in archeology and during the genuinity check of historic artifacts as well [7]. One of the most commonly used features of marbles is the distribution of the grain diameters, or at least its maximum

called the Maximal Grain Size (MGS). This is a typical image segmentation task which is especially hard with marble thin section images: a geomorphological process called crystal twinning creates grains with many straight lines inside them as shown in Fig. 2. This phenomena prohibits the convenient use of standard segmentation methods based on edges. In order to achieve a high quality result, we decided to create an application for semi-automatic marble thin section segmentation.

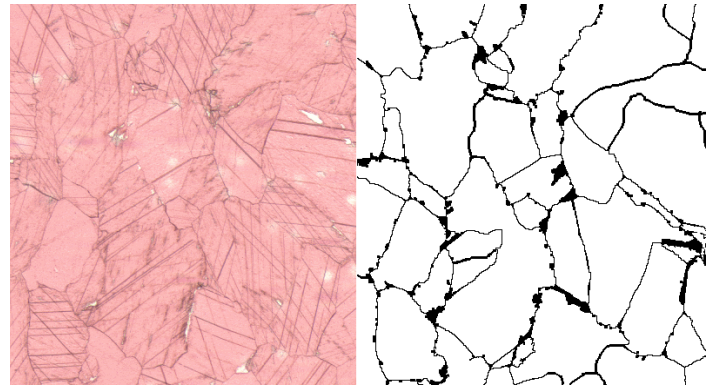


Figure 2: Marble thin section with twin crystals, sample taken from the MissMarble data set [8].

The workflow usually starts with a general purpose image segmentation to retrieve boundary suggestions. Among these, the software tries to identify the twin crystal lines and presents a suggestion for merging several pieces of grains. At this point, the user can review all the suggestions, add further ones or slice some grains into pieces. When finished, the merge and slice operations are performed and the final reports with the grain size histograms can be generated.

The intermediate manual review has two purposes: on one hand it is required to fix the mistakes the automatic methods may have left. On the other hand, this allows the system to gain trust among the users: they still have full control and there is no need to hope that the automatic process will detect everything correctly. With the human help, the results can be free of mistakes.

Even if the still experimental twin crystal detection methods are not sufficiently reliable, the cv4s frameworks merge and slice operations allow the user to work very fast after the initial segmentation.

During the development of GrainAutLine, an additional task came across us regarding sandstones: we need to count how many neighbors a quartz grain has, and create a histogram from these neighbor counts. This is required during the material selection of restoration works as a key indicator of porosity. If the grains have less than 3.45 neighbors in average, the domain experts consider it unsuitable for replacing missing parts of buildings for example. Fig. 3. shows the original image, the retrieved neighborhood graph, and the resulting histogram.

To achieve this, the following operations were added to the GrainAutLine application:

- We use seeded watershed segmentation which needs a marker in every grain. We added an operation to mark the center of large clear areas, but the user gets the possibility to add further ones or modify existing suggestions as needed.

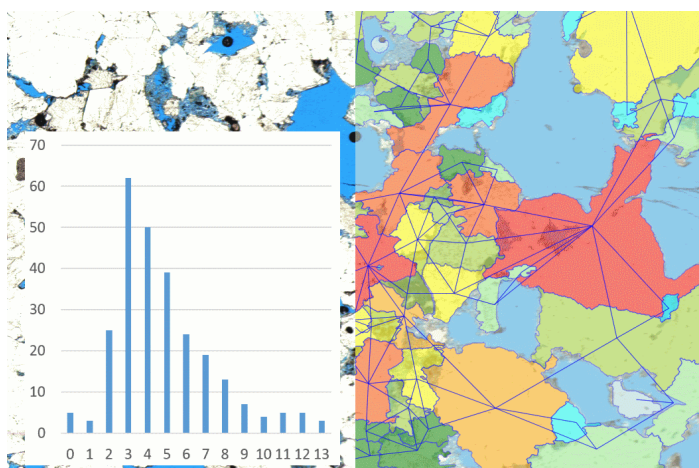


Figure 3: Sandstone image with painted porosities (left background), identified grains with neighborhood graph and neighbor count histogram.

- The watershed segmentation operation is a wrapper of the OpenCV implementation.
- The color marker operation is used to remove the porosity areas. It takes selected polygons as input, learns their color and covers all areas with these colors with additional polygons. These porosity polygons can be subtracted from all other polygons using the subtraction operation of the framework.
- The “neighbors” derived data was already in the framework, we only added an operation showing the graph as an image overlay as seen in Fig. 3. It is only a set of polylines which have the “decorator” tag marking them not to be included in further operations.
- The neighborhood counter operation stores the neighbor number for every polygon in a dedicated tag. This value will also control the color of the corresponding polygons.
- An exporting operation reads the neighbor count tag from every polygon and exports the histogram with a chart into an Excel table for convenience.

The GrainAutLine application is not yet finished, further operations are being developed to reduce the amount of manual work, but it is already suitable for production use. Evaluating the GrainAutLine system in terms of our research goals is not easy. A quantitative analysis would need the monitoring of several users while they use the system. Our experiences until now show that the work can be accelerated even by a factor of 10, but this depends highly on the complexity of the images to process. But even if the automatic tools are less effective, guiding a software with simple mouse strokes is much more convenient and less tiresome than to carefully draw the boundary lines one-by-one.

4.2. ChemoTracker

The white blood cell tracking application called ChemoTracker is used for immunology related medical research, in collaboration with the Semmelweis University. Since the accumulation of white blood cells is critical for the development of inflammation, it is important to understand the molecular mechanism of white blood cell migration. This is tested by time-lapse microscopic recording of migration of neutrophils (one type of white blood cells) towards an inflammatory compound [9,10], followed by automated

analysis of the movement path of individual cells in microscopy image sequences. This requires the segmentation of a series of images distinguishing between white blood cells and other visual anomalies, and the tracking of the individual cells.

The customization and extension of the framework required the following steps:

- The user interface needs to show the available operations as a convenient ribbon.
- A specialized image segmentation operation was added which can recognize the white blood cells properly. It tries to distinguish them even if they are stuck together.
- As the robotic microscope is observing several samples periodically, its slight positioning errors have to be compensated, so that steady objects in consecutive images do not move.
- The tracking operation takes the polygons of all the images and assigns them to cells, one polygon for every cell from every image which is stored in the data model as a dedicated “CellID” tag. Further operations are available for the ergonomic manual modification of these assignments. If the user sets a polygon to be the successor of another one in the previous frame, the “manual” tag is added to prevent further tracking to override this. The user can achieve the correct tracking by iteratively running the tracking and applying corrections to the results.
- The visualized motion trails of the cells are polylines which get their opacity from a tag value. This tag value is updated whenever the current image index changes.

It should be noted that the main challenges of the tracking are caused by two factors: (1) when the microscope is taking pictures of many samples after each other, the time gap between images is long and the faster cells may move far away from their previous location, sometimes with sharp turns. And (2) the shape of the white blood cells changes significantly between images, so that shape and appearance based identification is near to useless. Fig. 4. shows the user interface of ChemoTracker.

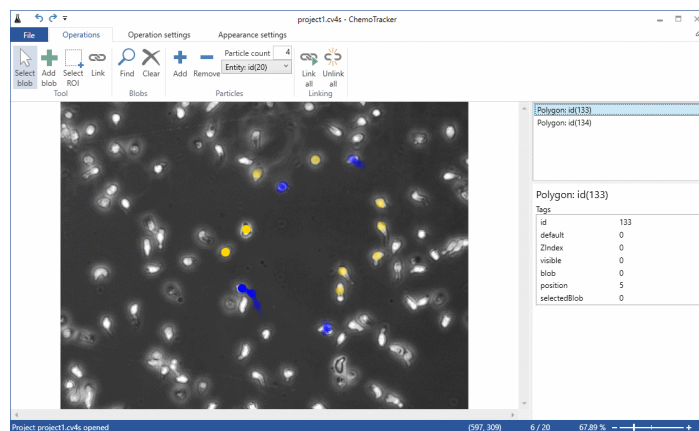


Figure 4: User interface for cell motion tracking in ChemoTracker

The visualization of the cell motions required a semi-transparent trail. As that one could be added to the data model as a polyline, every required function of ChemoTracker could fit into the common data model. During evaluation tests, using the

ChemoTracker proved to be better than having to click on all cells on all images all day. Although the best way for visualization is still an open question, we estimate a 5-10 factor acceleration. Even if the system is unable to track the fast (and relative unpredictably moving) cells, the manual tracking is much more convenient and less tiresome after the software has marked the cells which it is unsure about.

5. Conclusions

This paper presented the cv4s framework as an environment for creating software solutions to support several research domains utilizing image processing. Beside the core capabilities of the system, two software solutions have been presented which were built over the framework: GrainAutLine and ChemoTracker. Demonstrating the high domain independence of the project, we are investigating collaborations with experts in archaeometry, petrography, construction material analysis, and road fracture evaluation. The basic data representation and plug-in like architecture of the framework proves to be an efficient foundation for these kinds of tasks.

The research questions raised in the introduction were challenging the applicability of the general data model for very different domains which we have proven to be true: both applications could use it without extensions. The second question asked the advantage a combination of domain specific researchers and IT engineers can achieve. We believe that it is significant and collaborations of this kind can open up research directions which have been flagged “unfeasible” until now.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to express their special thanks to Anna Váry, Márton Sebők, and Zsolt Dudás for their software development efforts, Anikó Bere, Zsuzsanna Pató, and Attila Mócsai for the ChemoTracker cooperation, Zsófia Koma, Balázs Székely, and Judit Zöldföldi for the GrainAutLine cooperation.

This work was performed in the frame of FIEK_16-1-2016-0007 project, implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the FIEK_16 funding scheme.

6. References

- [1] Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, 4th Edition, Pearson, 2017.
- [2] Mark Nixon, “Feature Extraction and Image Processing for Computer Vision”, 3rd Edition, Academic Press; 3 edition, 2012.
- [3] S. Ojha and S. Sakhare, "Image processing techniques for object tracking in video surveillance- A survey" in International Conference on Pervasive Computing (ICPC), Pune, 2015, pp. 1-6.
- [4] G. Bradski: The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.
- [5] L. Lengyel, P. Ekler, T. Ujj, T. Balogh, H. Charaf, “SensorHUB – An IoT Driver Framework for Supporting Sensor Networks and Data Analysis” in International Journal of Distributed Sensor Networks, Vol. 2015, Article ID 454379, 12 pages, 2015.
- [6] J. F. Canny. A computational approach to edge detection. IEEE Trans. Pattern Analysis and Machine Intelligence, pages 679-698, 1986.

- [7] J. Zöldföldi, M. Satir, “Provenance of the White Marble Building Stones in the Monuments of Ancient Troia” in G. A. Wagner, E. Pernicka, E. H. P. Uerpmann (eds.) Troia and the Troad. Springer, Berlin, 2003.
- [8] J. Zöldföldi, P. Hegedűs, B. Székely, “MissMarble, an Interdisciplinary Data Base of Marble for Archaeometric, Art History and Restoration Use” in I. Turbanti-Memmi (ed.), Proceedings of the 37th International Symposium on Archaeometry, Springer-Verlag, Berlin, Heidelberg, 355-361., 2011.
- [9] K. Futosi, T. Németh, R. Pick, T. Vántus, B. Walzog, A. Mócsai, “Dasatinib inhibits proinflammatory functions of mature human neutrophils”, *Blood*. 2012 May 24;119(21):4981-91. doi: 10.1182/blood-2011-07-369041. Epub 2012 Mar 12., PMID: 22411867
- [10] T. Németh, K. Futosi, C. Hably, MR Brouns, SM Jakob, M. Kovács, Z. Kertész, B. Walzog, J. Settleman, A. Mócsai, “Neutrophil functions and autoimmune arthritis in the absence of p190RhoGAP: generation and analysis of a novel null mutation in mice.”, *J Immunol*. 2010 Sep 1;185(5):3064-75. doi: 10.4049/jimmunol.0904163. Epub 2010 Jul 30., PMID: 20675588.

Medium Height Dual Buildings with Masonry and Concrete Walls in High Seismic Areas

Sorina Constantinescu*

Technical University of Construction Bucharest, Department of Civil Engineering, ZIP Code 011711, Romania

ARTICLE INFO

Article history:

Received: 28 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

Plastic mechanism

Masonry stiffness

Concrete walls ductility

ABSTRACT

This is a comparative study on the behavior of a dual medium height building with different walls solutions, in a high seismic area (Bucharest, Romania). The main feature of those buildings is the placement of load bearing walls on the perimeter of the building. This is done to limit the lateral displacement when the structure is subjected to seismic loading. The walls cannot be placed inside because of architecture demands. The structure has frames on the inside. The walls may be made of confined masonry, as the medium height buildings are allowed to have, but they may also be made of reinforced concrete. The wall area on direction Y is smaller than on X. There may be high efforts on direction Y. This study will show witch walls solution ensures the best behavior for the building. It is also interesting to see the way frames, masonry and concrete walls work together. The study contains both elastic and plastic state analysis results. This study results may be used for any dual medium height building with perimeter walls, and smaller walls area on one direction.

1. Introduction

This study highlights the behavior of confined masonry walls in a medium height dual building. The walls will be used to limit the building drifts. This is an office building, 3 stories high, which will be built in Bucharest Romania. This is considered a high seismic area, the seismic acceleration is 0.30g (g is the gravity acceleration). According to the architecture demands, the building's partitioning needs to be flexible. This is why walls will be placed on the perimeter of the building and not inside it. Walls assure lower drift values. Both masonry and reinforced concrete walls may be used in this example because this is a medium height building. However, masonry is susceptible to cracking. Dynamic loads may cause irregular deformations [1]. Load-bearing masonry walls show a complex behavior due to the load eccentricity. Walls slenderness reduces the bearing capacity [2]. Studies on masonry walls have shown that shear force capacity is reduced with the increase of bending moment [3]. Masonry walls reinforcements help the masonry to work together with the confining elements [4] and an increased number of confining columns improves the walls strength, the energy dissipation capacity, the ductility and the cracking pattern [5]. According to the model experiments, the confined masonry and concentrated reinforced masonry structures have been used for low and medium-rise buildings in seismic areas [6]. Seismic actions cause

vibrations in walls that create lateral loads variable in time. The masonry walls lateral load bearing capacity depends much on its slenderness and the axial load value [7]. Masonry walls may give in to shear force by diagonal shear and sliding shear, so reinforcement is recommended for high seismic areas. There is also another failure called corner crushing, that is not considered in the masonry design [8]. For dual buildings, seismic shear failure is expected to occur in walls. The frames should show a large deformation capacity in the plastic stage [9]. Pushover analysis may give information both about the building's failure pattern, by the plastic hinges development but also about the building's safety when the plastic stage is reached. The pushover diagrams are created in terms of base force and top of building displacement [10]. Medium-rise reinforced concrete walls show a good seismic behavior for different earthquake patterns. Buildings with slender walls may show important ductility. Plastic hinges mostly develop at the beams ends [11]. It is interesting to see the interaction between frames in the building's center and masonry or concrete walls both in the elastic and plastic state. It is not a usual solution to use both masonry and reinforced concrete walls for the same structure. The codes in force used to design the building are [12–19].

2. Structural Solutions

There are 3 structural solutions that may be used for this building. The first solution, seen in Figures 1 and 2, use just confined

*Corresponding Author: Sorina Constantinescu, Bucharest, 0742265890, sorina.constantinescu@yahoo.com

masonry walls, on the building's perimeter, on both directions. For the second solution, seen in Figures 3 and 4, the 4 walls on direction Y are made of reinforced concrete. In the third solution, in Figures 5 and 6, on direction Y, the walls are small, made of reinforced concrete. In Figures 1 to 6 the beams are blue, columns and reinforced concrete walls are green, confined masonry walls are red and slabs are light gray. The software used for analysis is ETABS 2016.

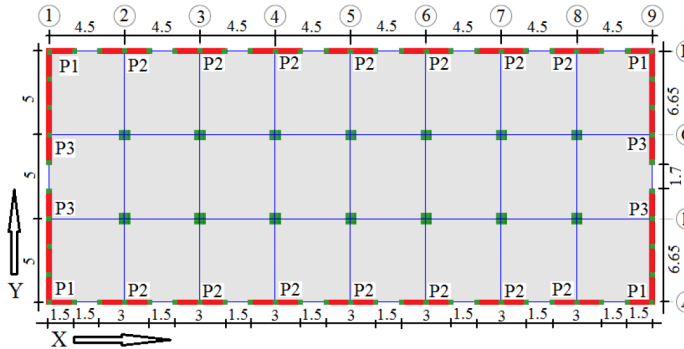


Figure 1: Story plan for solution 1

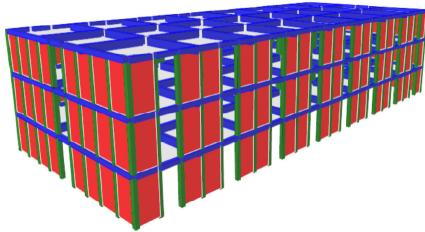


Figure 2: 3D building image for solution 1

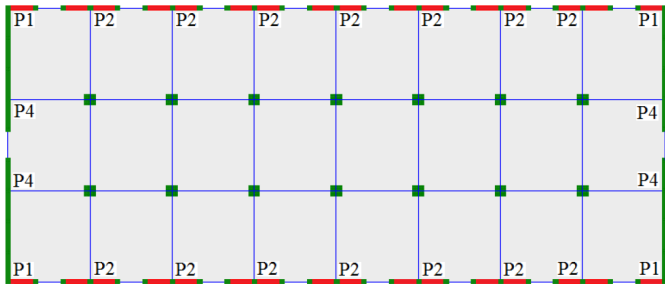


Figure 3: Story plan for solution 2

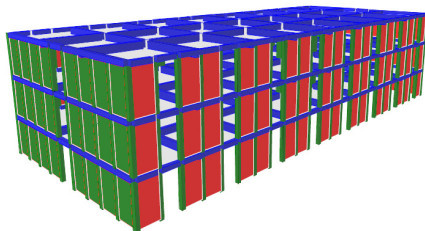


Figure 4: 3D building image for solution 2

3. Structural Elements Design

The structure design is done using the seismic loads combination that contains 1.0·permanent loads+0.4·variable loads+1.0·seismic loads.

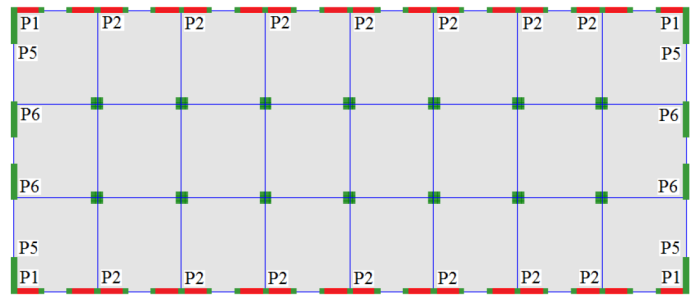


Figure 5: Story plan for solution 3

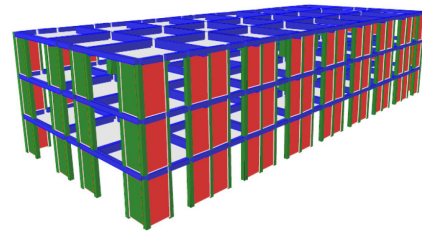


Figure 6: 3D building image for solution 3

The building is composed of a ground floor and 2 stories above it. Walls P1, P2, P3, P4, P5 and P6 horizontal sections are shown in detail in Figure 7.

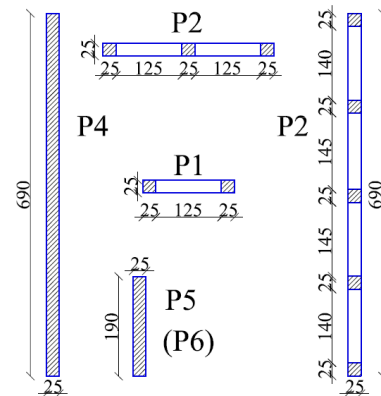


Figure 7: Walls details

Materials used here are concrete C20/25 [16], with elasticity modulus $E_C=30000\text{N/mm}^2$ and full bricks $240 \cdot 115 \cdot 63$ (mm) with standard strength $f_b = 10\text{N/mm}^2$, mortar M7.5 and elasticity modulus $E_M=4050\text{N/mm}^2$ [12]. Reinforcement bars are S355 with elasticity modulus $E_S=210000\text{N/mm}^2$ [16]. The seismic coefficient c_s introduces the seismic load. The base force F_b is calculated using [17]. $\gamma_{l,e} = 1.2$ is the building's importance-exposure coefficient, $\beta_0 = 2.5$ is the maximum value of the elastic spectrum and q is the structure's behavior factor, $q=2.25 \cdot \alpha_w/\alpha_1=2.25 \cdot 1.35$ [17], α_w/α_1 = the base shear force value for the failing mechanism/the base shear force value for the first plastic hinge, m = building's mass [17]. $\eta = 0.88$ is the reduction factor, $\lambda = 0.85$ for 3 stories buildings, $a_g = 0.30g$ [17], G = building's weight.

$$F_b = \gamma_{l,e} \cdot \beta_0 \cdot a_g/q \cdot m \cdot \eta \cdot \lambda = c_s \cdot G = 0.22 \cdot G \text{ [kN]} \quad (1)$$

3.1. Masonry Walls Design Theory

The masonry walls stresses analyzed are: σ_x , σ_z , and τ_{xz} . They are compared to the masonry design strengths [12], that are the following: horizontal compression f_{dh} , vertical compression f_d , and shear strength for horizontal direction $f_{vd,i}$. Those design strengths are determined from their characteristic values: f_{kh} , f_k , and $f_{vk,0}$, using the characteristic masonry strengths insurance factor $\gamma_M=1.9$ and the unitary vertical stress (σ_d) [12]. The concrete compression design strength f_{cd} , is determined using the characteristic strength (f_{ck}) and $\gamma_M=1.5$, for concrete [16].

$$f_{dh} = f_{kh}/\gamma_M = 1.91/1.9 = 1 \text{ N/mm}^2 \quad (2)$$

$$f_d = f_k/\gamma_M = 4.05/1.9 = 2.13 \text{ N/mm}^2 \quad (3)$$

$$f_{vd,i} = f_{vk,0}/\gamma_M + 0.4 \cdot \sigma_d = 0.3/1.9 + 0.4 \cdot 0.1 = 0.2 \text{ N/mm}^2 \quad (4)$$

$$f_{cd} = f_{ck}/\gamma_M = 20/1.5 = 13.3 \text{ N/mm}^2 \quad (5)$$

Wall percentages are calculated on both X and Y directions. The formula used is (6), A_w is the walls area on the direction the percentage is calculated and A_t is the total area of the floor [12].

$$p\% = A_w/A_t \cdot 100 \quad (6)$$

M_{Rd} (wall's bearing bending moment) calculated as in [12]. C_A is the compressed area of the wall, $M_{Rd(M)}$ is the bearing bending moment from the masonry area.

$$M_{Rd} = M_{Rd(M)} + M_{Rd(As)} \text{ [kNm]} \quad (7)$$

$$C_A = N_{Ed}/(0.85 \cdot f_d) \text{ [mm}^2\text{]} \quad (8)$$

$$M_{Rd(M)} = N_{Ed} \cdot y_c \text{ [kNm]} \quad (9)$$

y_c is the distance between the wall's weight center and the compressed masonry area weight center [12]. The bearing bending moment from the slender columns reinforcement at the wall edges is $M_{Rd(As)}$ [12].

$$M_{Rd(As)} = I_s \cdot A_s \cdot f_{yd} \text{ [kNm]} \quad (10)$$

$f_{yd} = f_{yk}/\gamma_M = 310 \text{ N/mm}^2$ is the design strength of the reinforcement bars. It is determined from the characteristic strength f_{yk} and the safety coefficient $\gamma_M=1.15$ for steel [16]. I_s is the distance between the slender columns at the wall margins centers. A_s is the horizontal reinforcement area of the slender columns.

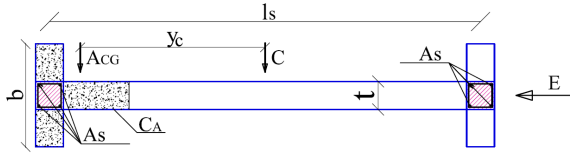


Figure 8: Confined masonry wall section

A_C is the wall's compressed area. A_{CG} is the compressed area gravity center. C is the wall section gravity center. E is the earthquake action. In Figure 9, $b = t \cdot f_{cd}/f_d$ [12]. V_{Rd} is the masonry wall bearing shear force and V_{Ed} is the horizontal shear force from the seismic loads combination.

$$V_{Rd} = V_{Rd1}^* + V_{Rd2} + V_{Rd3} \text{ [kN]} \quad (11)$$

$$V_{Rd1}^* = 0.4 \cdot (N_{Ed} + 0.8 \cdot V_{Ed} \cdot h_{pan}/l_{pan}) \text{ [kN]} \quad (12)$$

$$V_{Ed} \leq I_{pan} \cdot t \cdot f_{vd,0} \quad (13)$$

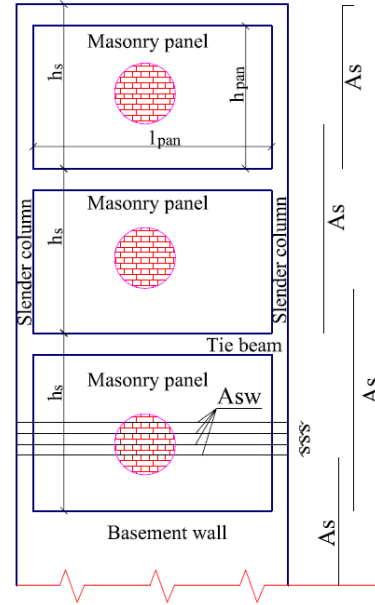


Figure 9: Confined masonry wall elevation

$f_{vd,0} = 0.16 \text{ N/mm}^2$ and $f_{vk,0} = 0.30 \text{ N/mm}^2$ are the design and characteristic initial shear strengths for no axial stress [12], h_{pan} and l_{pan} are the height and length of the masonry area panel. V_{Rd2} is the bearing horizontal shear force from the reinforcement in the slender column at walls compressed edge [12]. A_s is the reinforcement area of the slender column at the walls compressed edge. λ_c is the reinforcement participation factor. Here, $\lambda_c = 0.25$, for longitudinal reinforcement $\Phi 16$.

$$V_{Rd2} = \lambda_c \cdot A_s \cdot f_{yd} \text{ [kN]} \quad (14)$$

$$V_{Rd3} = 0.8 \cdot I_w \cdot A_{sw} \cdot f_{yd}/s \text{ [kN]} \quad (15)$$

V_{Rd3} is the bearing shear force taken by the horizontal reinforcement area in the bricks joints A_{sw} . s is the vertical distance between two horizontal reinforced joints. Here, the masonry walls have tie beams at each story level, as well as at each landing. The reinforcement in those tie beams can create V_{Rd3} , even if there are no horizontal reinforcement bars in the masonry wall. The load combination used to design the structure is $1.0 \cdot \text{permanent loads} + 0.4 \cdot \text{variable loads} + 1.0 \cdot \text{seismic loads}$.

3.2. Reinforced Concrete Walls Design Theory

$M_{Ed,0}$ and $M_{Ed,s}$ are the ground floor and upper floors design bending moments and M_{Rd} is the wall's bearing bending moment. M'_{Ed} is the bending moment from the moment diagram given by the seismic loads combination for any story [19].

$$M_{Ed,0} = M'_{Ed} \quad (16)$$

$$M_{Ed,s} = k_M \cdot \Omega \cdot M'_{Ed} \leq \Omega \cdot M'_{Ed} \quad (17)$$

$$\Omega = M_{Rd,0}/M'_{Ed,0} \quad (18)$$

$$V_{Rd,ww}=0.18 \cdot b_{w0} \cdot l_w \cdot f_{cd} \quad (19)$$

$$V_{Rd,h}=\Sigma A_{s,h} \cdot f_{yd} \quad (20)$$

$$V_{Rd,h}=\Sigma A_{s,h} \cdot f_{yd} + V_{Rd,c} \quad (21)$$

$$V_{Rd,c}=0.5 \cdot \sigma_{cp} \cdot b_{w0} \cdot l_w \quad (22)$$

$$\sigma_{cp}=N_{Ed}/(l_w \cdot b_{w0}) \quad (23)$$

$k_M=1.15$ for DCM (medium ductility buildings). V_{Ed} and V_{Rd} are the design and bearing shear forces. There are 3 V_{Rd} values that must surpass V_{Ed} [19]. Wall web V_{Rd} ($V_{Rd,ww}$), horizontal reinforcement V_{Rd} ($V_{Rd,h}$) and casting joint V_{Rd} ($V_{Rd,s}$). Equation (20) is used for the ground floor and (21) is for the upper floors. $A_{s,h}$ is the sum of horizontal reinforcement bars intersected by a 45° angle crack, $V_{Rd,c}$ is the shear force taken by the concrete wall area and σ_{cp} is the medium compression stress on the wall web.

$$V_{Rd,s}=\mu_f \cdot [\Sigma(A_{s,v}+A_s) \cdot f_{yd}+N_{Ed}] \quad (24)$$

$A_{s,v}$ is the sum of vertical reinforcement bars and $\mu_f=0.7$ is the friction coefficient [19]. Equations (16) to (24) are taken from [19]. Figure 10 is drawn according to [19].

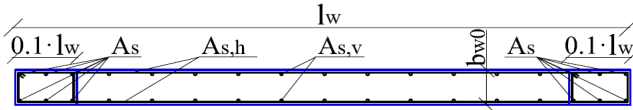


Figure 10: Reinforced concrete wall section

3.3. Reinforced Concrete Beams and Columns Design Theory

Bending reinforcement of beams is designed according to M_{Ed} according to [13– 18].

$$M_{Ed}=b \cdot \lambda x \cdot f_{cd} \cdot (d-\lambda x/2)=A_s \cdot f_{yd} \cdot z \quad [kNm] \quad (25)$$

$$m=M_{Ed}/(b \cdot d^2 \cdot f_{cd}) \quad (26)$$

$$z=d-\lambda x/2=d-d \cdot (1-(1-2m)^{0.5})/2 \quad [mm] \quad (27)$$

$$A_{s,min} = \min\{0.26 \cdot f_{ctm}/f_{yk} \cdot b \cdot d; 0.0013 \cdot b \cdot d\} \quad (28)$$

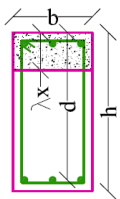


Figure 11: Reinforced concrete beam section

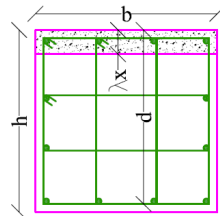


Figure 12: Reinforced concrete column section

A_s is the minimum reinforcement area for beams. λx is the beam section compressed area height [16]. $f_{ctm}=2.6N/mm^2$ is the medium value of the concrete tensile strength. Columns bending moment is calculated using according to [18]. $\gamma_{Rd}=1.2$ is the steel stiffening factor for DCM (medium ductility buildings) [18], ΣM_{Rc} and ΣM_{Rb} are the sums of bearing bending moments

of columns and beams near a frame joint. The longitudinal reinforcement percent for columns minimum value is $p_{min}=1\%$ and the maximum is $p_{max}=4\%$ [18]. If $\lambda x < 2 \cdot a_s$, A_s will be determined from (32), and from (33), if $\lambda x \geq 2 \cdot a_s$. N_{Ed} is the axial force in the calculated columns [16]. Here $a_s=45mm$.

$$\Sigma M_{Rc} \geq \gamma_{Rd} \cdot \Sigma M_{Rb} \quad [kNm] \quad (29)$$

$$p=A_s/(b \cdot d) \cdot 100 \quad (30)$$

$$x=N_{Ed}/(b \cdot \lambda \cdot f_{cd}) \quad [mm] \quad (31)$$

$$A_s=[M_{Ed}-N_{Ed}(d-a_s)/2]/[f_{yd} \cdot (d-a_s)] \quad [mm^2] \quad (32)$$

$$A_s=[M_{Ed}+N_{Ed}(d-a_s)/2-b \cdot \lambda x \cdot f_{cd}(d-\lambda x/2)]/[f_{yd}(d-a_s)] \quad (33)$$

4. Results for the Elastic Stage Analysis

For all 3 solutions, element dimensions and longitudinal reinforcement in beams, tie beams, columns and slender columns are seen in Table 1. A_s is the longitudinal reinforcement area. The bars are seen as black discs and the diameter (Φ) of bars (in mm) is written for each element.

Table 1: Element dimensions and longitudinal reinforcement in beams, tie beams, columns and slender columns

Beam 25x50 As → 3Φ16 up and down	Tie beam 25x30 As → 4Φ16	Column 60x60 As → 12Φ20	Slender column 25x25 As → 4Φ16

4.1. Results for Walls for Solution 1

Efforts in confined masonry walls are seen in Table 2. P1, P2 and P3 are the wall labels from Figure 1. For walls P3 it is necessary to place horizontal reinforcement bars. In Table 2, $A_{sw}: 2\Phi(8)10/15$ means 2 bars of (8mm)10mm diameter placed at every 15cm in the masonry wall and S1, S2 and S3 mean story 1, 2 and 3.

Table 2: Confined masonry walls efforts

	P1			P2			P3		
	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]
S3	207	19	529	178	165	1406	271	795	3409
S2	482	124	734	355	541	1689	542	2547	4309
S1	730	312	912	532	1063	1918	812	4640	5192
	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw} □	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw}	V_{Ed} [kN]	V_{Rd} [kN]	A_{sw}
S3	17	179	0	102	167	0	486	749	2Φ8/ 150
S2	78	289	0	201	238	0	859	945	2Φ10/ 150
S1	97	388	0	230	309	0	904	1053	2Φ10/ 150

4.2. Results for Walls for Solutions 2 and 3

For solution 2 and 3, the elastic stage design results for the reinforced concrete walls are seen in Table 3. Walls P4, P5 and P6 are mainly subjected to bending moments. The minimum horizontal reinforcement required by the code is enough to bear the shear forces. P5 and P6 have the same dimensions, so they are both calculated with the same sectional efforts.

Table 3: Elastic stage design results for the reinforced concrete walls

	P4				P5 and P6			
	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]	A_s	N_{Ed} [kN]	M_{Ed} [kNm]	M_{Rd} [kNm]	A_s
S3	335	1183	2815	6Φ12	138	242	536	8Φ25
S2	667	3587	4771	6Φ12	278	1131	1154	8Φ25
S1	984	6676	7297	6Φ25	418	2021	2136	8Φ25
	l_w	t	$A_{s,v}$	$A_{s,h}$	l_w	t	$A_{s,v}$	$A_{s,h}$
	6.9m	0.25m	2Φ14/ 350	2Φ12/ 250	1.9m	0.25m	2Φ14/ 350	2Φ12/ 250
	V_{Ed} [kN]	$V_{Rd,ww}$ [kN]	$V_{Rd,h}$ [kN]	$V_{Ed,s}$ [kN]	V_{Ed} [kN]	$V_{Rd,ww}$ [kN]	$V_{Rd,h}$ [kN]	$V_{Ed,s}$ [kN]
S3	501	4139	863	2054	151	1139	628	1178
S2	933	4139	1728	2745	340	1139	627	2096
S1	1123	4139	1541	5118	438	1139	490	2194

5. Nonlinear Analysis Results

5.1. Nonlinear Analysis Results for Solution 1

The two pushover cases used for the building’s nonlinear analysis for the first solution are PX and PY. The plastic hinges development for both nonlinear cases are seen in Figures 13 and 14. The color code is the following: B (green) means the plastic hinge is formed, C (light blue) means the plastic hinge reaches the limit and the element gives out, D (pink) means the load was redistributed and E (red) means collapse. Those colors seen in Figures 13 and 14 show the pushover analysis last steps. For case PX, at step 70, the plastic hinges reach collapse stage at the bottoms of all columns in the center and slender columns in walls on both directions. The same stage is reached by the hinges in the short beams connecting walls on direction X at the top stories. There are hinges developed to stages D in the long beams close to the building edges on X.

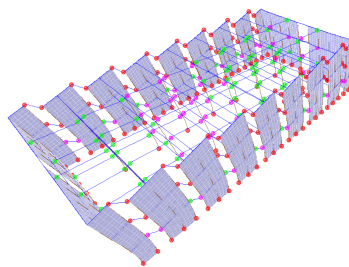


Figure 13: Plastic hinges: PX step 70

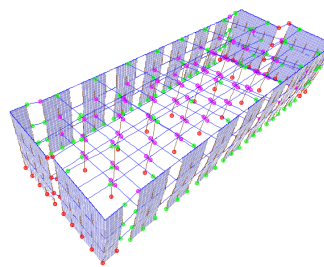


Figure 14: Plastic hinges: PY step 72

Plastic hinges in stage B are seen in all the other beams and columns. On direction Y, at step 72, plastic hinges reach stage E at the bottoms of slender columns in walls on Y and in the beams that connect these walls.

Also hinges in stage E appear at the bottoms of walls close to those on Y and at the bottoms of columns in the center. Hinges in stage D are seen in beams on direction Y. It is important to notice that for case PY, the walls on X are bended as they are slender, so no plastic hinges develop to stage E at their bottoms. For case PX, on the other hand, walls on direction Y do develop plastic hinges, as they are stiff. In both cases, plastic hinges reach collapse in the walls on the same direction as the nonlinear case. The same stage is reached in the beams that connect these walls, and not in the long beams. Efforts are always higher in the beams connecting walls than in frame ones, because walls are stiffer than columns. Stresses in each confined masonry wall are shown in Figures 15 – 20.

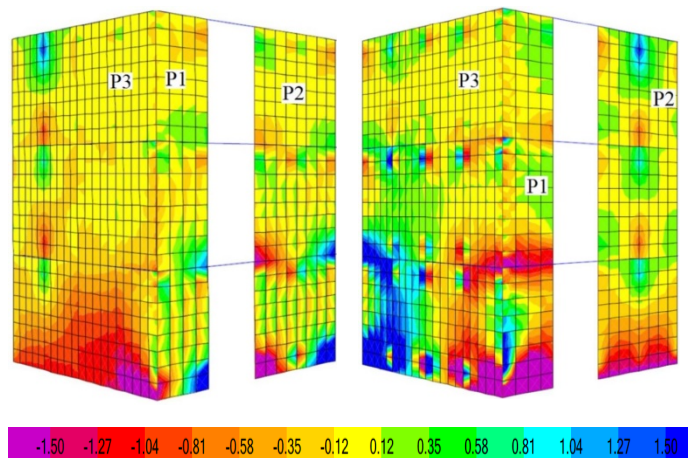


Figure 15: PX step 2 $\sigma_x=1.5N/mm^2$ Figure 16 PY step 2 $\sigma_x=1.5N/mm^2$

Stresses σ_x show increased values near the walls base and near the tie beams. The stresses show opposite signs from one side of the walls to the other. This means one side of each wall is stretched and the other is squashed. This is due to the horizontal direction of the stress. Beams and tie beams bring high stress values to the walls. They are stiffer than masonry walls and they transmit both vertical and horizontal efforts to the walls. The highest stress values are seen in P3, those walls are subjected to higher seismic loads because they are stiffer and fewer.

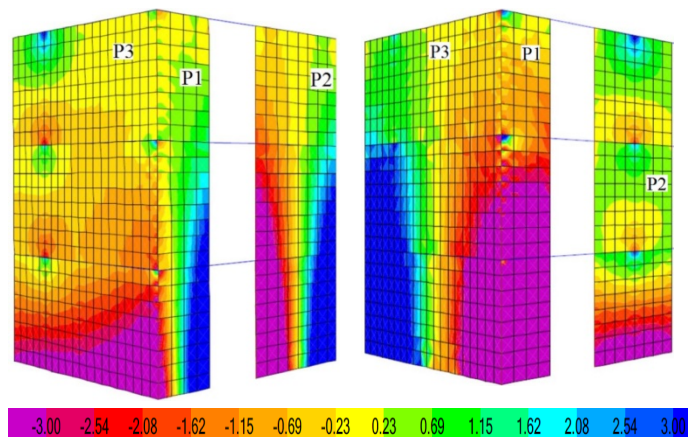


Figure 17: PX step 2 $\sigma_z=3N/mm^2$ Figure 18: PY step 2 $\sigma_z=3N/mm^2$

Stress σ_z reaches important values at the walls base. At the bottom edges of walls on the same direction as the nonlinear case there are opposite sign stresses. This is because the walls are subjected to axial stress but also horizontal loads from the nonlinear load cases.

Stress values reach comparable values at the same walls height. In the walls perpendicular to the load case direction there are high stress values at the bottoms. Slender walls on X are completely cracked at the bottom. Stresses τ_{xz} reach the highest values in the masonry walls panels at the walls lowest stories, as the walls are designed as fixed at the bottoms. Near the tie beams or slender columns, the stresses are taken less by the masonry and more by the confining elements. The highest stress values are reached in walls on direction Y.

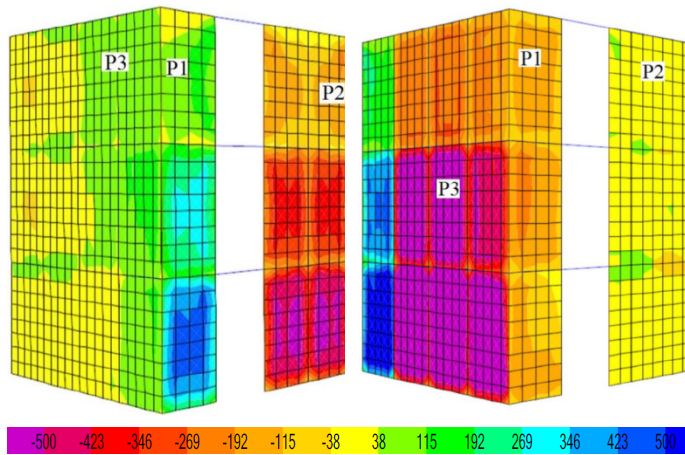


Figure 19: PX step 2 $\tau_{xz}=0.5N/mm^2$ Figure 20: PY step 2 $\tau_{xz}=0.5N/mm^2$

All stresses surpass the strength values from step 2 of the analysis. The masonry is cracked before the full plastic mechanism is formed. It is clear that walls on direction Y have higher loads to bear compared to those on X.

5.2. Nonlinear Analysis Results for Solutions 2 and 3

For the second solution, the plastic hinges stages and distributions when the plastic mechanism is reached are seen in Figures 21 and 22. PX C1 and PY C1 are the static nonlinear load cases for the structure in which walls labeled P3 are made of reinforced concrete and renamed P4, for the second solution. Cases PX C2 and PY C2 are used for the third structure solution, where wall P4 is replaced by 2 small reinforced concrete walls P5 and P6, connected by beams.

For the second solution, the plastic mechanism is formed in very much the same way as for the first one.

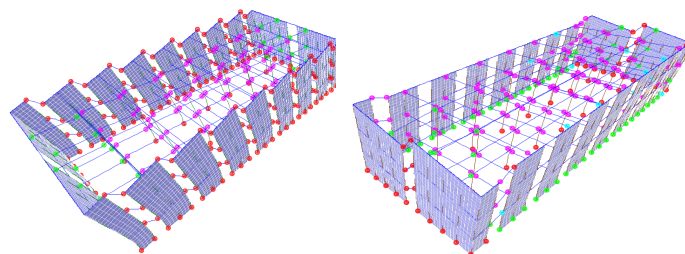


Figure 21: Plastic hinges: PX C1 step 30

Figure 22: Plastic hinges: PY C1 step 22

There are fewer steps for the nonlinear analysis for both directions. Reinforced concrete walls labeled P4, are subjected to higher stresses, especially for case PY C1. This is because they are the main load bearing elements on direction Y.

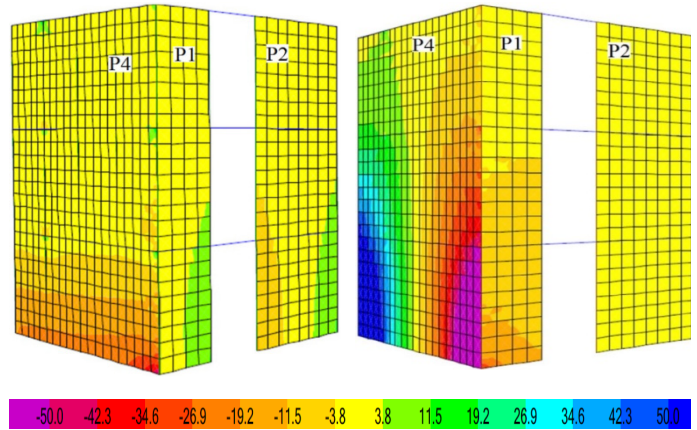


Figure 23: PY C1 step 2 $\sigma_z=50N/mm^2$

Figure 24: PY C1 step 2 $\sigma_z=50N/mm^2$

On this direction, at step 2 of the analysis, the reinforced concrete design compression strength f_{cd} is surpassed up to the third story. At base, the concrete walls are completely cracked. The damaged area was not as extended if confined masonry walls P3 were used. Although the masonry strengths are much lower than that of concrete, the stresses are much smaller and more evenly distributed among the masonry walls. This behavior is not the same for direction X, because stresses on this direction are taken mostly by the confined masonry walls on X. Concrete walls are corner cracked early in the analysis.

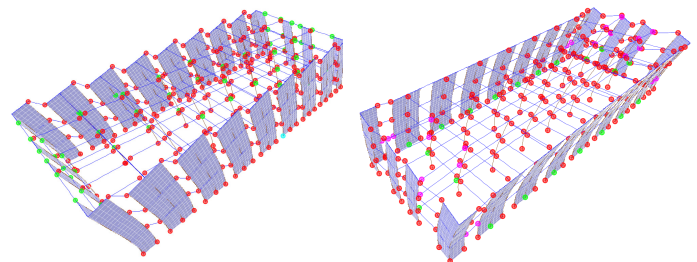


Figure 25: Plastic hinges: PX C2 step 29

Figure 26: Plastic hinges: PY C2 step 93

The third solution uses 2 reinforced concrete walls instead of wall P4, as seen in Figure 5. This will assure a less stiff behavior on direction Y, as the 2 concrete walls are slender but they can bear the loads they are subjected to. Of course, the new concrete walls are designed according to the sectional efforts from the load combination. This solution is also studied in the plastic state, by using the nonlinear cases PX C2 and PY C2. Hinges development at the final stage is seen in Figures 25 and 26. Most hinges are in stage E when the maximum displacement considered (100 cm) is reached. These hinges are equally distributed, at the beams ends and at the walls bottoms on each direction. For both cases PX C2 and PY C2 concrete walls reach greater stresses than the masonry ones. The differences in values are seen particularly for direction Y, because concrete walls are on that direction. It is seen that for this third solution, the use of more concrete walls makes the

corner cracking area much smaller than for the second solution. This is expected, because walls are less stiff and attract smaller effort values.

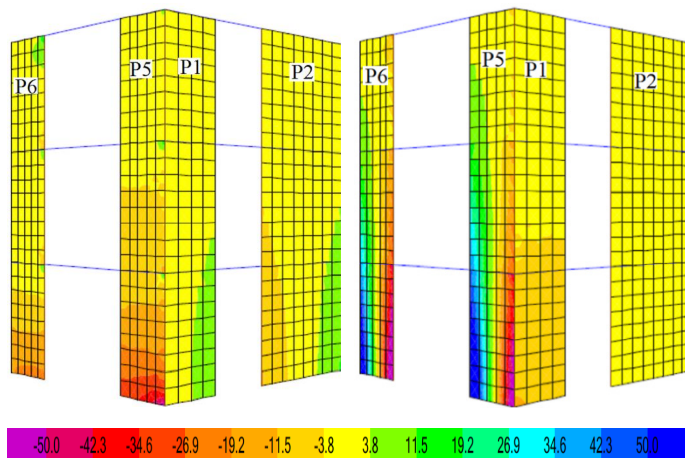


Figure 27: PX C2 step 2
 $\sigma_x=50\text{N/mm}^2$

Figure 28: PY C2 step 2
 $\sigma_x=50\text{N/mm}^2$

5.3. Pushover Diagrams

The pushover diagrams for the 6 pushover cases are shown in Figure 30. The analysis is performed until the displacement 100cm. This is enough for the structures to reach the plastic stages for all pushover cases.

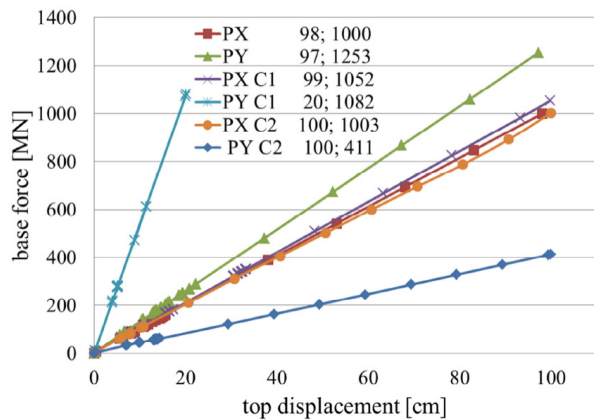


Figure 29: Pushover diagrams for all 3 structural solutions

In figure 29, the displacement values are 98, 97, 99, 20 and 100 cm. They are followed by the base force values. In the first case, when only confined masonry walls are used, there are a lot of small steps for both PX and PY diagrams at the beginning of the analysis. This means a lot of nonlinear hinges are formed in the structure, or hinges already formed advance to the next stage. However, the building's rigidity is maintained the same, the pushover diagrams are straight lines. Towards the end of the analysis, the steps are widely spaced. This means that more base force and displacement increase is necessary to advance the plastic hinges to the next stages. The structure is stiffer for case PY, because there are 4 long walls (P3) on direction Y. These walls are stiffer than the 18 short walls on X. The walls stiffness is created not by the sum of walls areas but mostly by each walls length on that direction. The building rigidities are 10.2MN/m for PX and 12.9MN/m PY. They are calculated as base

force/displacement. For solution 2, when reinforced concrete walls (P4) are used on direction Y, the impact of these walls is less seen on direction X in pushover curve PX C1. This diagram shows the building is capable of large displacement, the same as in the previous case. The analysis steps are spaced differently. There are some groups of steps very close to each other, and there are some steps spaced apart. The structure's rigidity is maintained, as the pushover curve is a straight line. It is seen that for PX C1 the building is slightly stiffer than for PX. Rigidity for PX C1 is 10.6NM/m. The reinforced concrete walls do have an effect, although they are placed on the other direction. For case PY C1 the building behaves completely different than in all other cases. The plastic mechanism is reached for a much lower displacement. The stiffness is very high: 54,1MN/m. The long reinforced concrete walls have a huge impact on the building's behavior.

For solution 3, when 2 small reinforced concrete walls (P5 and P6) are used instead of P4, PX C2 shows the same behavior pattern as PX. For PY C2, the building shows a very low rigidity (4.11MN/m), compared to PY. This means that P5 and P6, although they are reinforced concrete, they are less stiff than confined masonry wall P3.

6. Conclusions

The building behaves well both in the elastic and plastic state if all walls are made of confined masonry and also if 2 reinforced concrete walls are used instead of each long confined masonry wall on direction Y. For the first solution, masonry walls on direction Y have greater efforts as they are stiffer and the walls area on Y is half the one on X. They need horizontal reinforcement to bear shear forces, while no such reinforcement is needed for walls on X. The failure mechanism allows great lateral displacements for both directions. For the second solution, when there are 4 long reinforced concrete walls on Y, the building behaves stiffer. Walls P4 are subjected to high efforts and important damage at smaller displacements in case PY C1. Both first and third solutions are good, but the first solution is recommended. This is because all walls are made of confined masonry and they behave similarly.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] M. Kaluza, "Analysis of in plane deformation of walls made using AAC blocks strengthen by GFRP mesh" International Conference on Analytical Models and New Concepts and Masonry Structures (AMCM) 2017 Elsevier Procedia Engineering Vol 14. p 393-400, 2017. doi:10/1016/j.proeng.2017.06.229
- [2] C. Cornado, J.R. Rosell, J. Leiva, C. Diaz, "Experimental study of brick masonry walls subjected to eccentric and axial load" International RILEM Conference on Materials, Systems and Structures in Civil Engineering Conference segment on Historical Masonry Technical University of Denmark, Lyngby, Denmark p 33- 40, 2016. www.rilem.net/publications/proceedings-500218
- [3] J. J. Perez-Gavilan, L. E. Flores, A. A. Mazano, "New Shear Strength Design Formula for Confined Masonry Walls: Proposal to the Mexican Code" tenth U.S. National Conference on Earthquake Engineering, Frontiers of Earthquake Engineering (10 NCEE) Anchorage Alaska, 2014. <https://www.eeri.org/products-page/national-conference-on-earthquake-engineering/10th-u-s-national-conference-on-earthquake-engineering-frontiers-of-earthquake-engineering-proceedings-thumb-drive/>

- [4] K. Yoshimura, K. Kikuchi, M. Kuroki, H. Nokana, K. Tae Kim, R. Wangdi, A. Oshikata, "Experimental study for developing higher seismic performance of brick masonry walls" 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada; 2004 paper No. 1597, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [5] A. Marinilli, E. Castilla, "Experimental evaluation of confined masonry walls with several confining columns" 13th World Conference on Earthquake Engineering Vancouver 2004, B.C., Canada; paper No. 2129, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [6] D. H. Liu, M. Z. Wang, "Masonry structures with beams and columns" 12 World Conference on Earthquake Engineering (WCEE) Auckland, New Zealand paper 2720, 2000. <http://www.worldcat.org/title/12wcee-2000-12th-world-conference-on-earthquake-engineering-auckland-new-zealand-sunday-30-january-friday-4-february-2000>
- [7] A. T. Vermeltoort, D.R.W. Martens, "Preliminary Tests on the Lateral Load-bearing Capacity of Slender Masonry Walls 13th Canadian Masonry Symposium Halifax, Canada, 2017. <https://www.canadamasonrydesigncentre.com/symposiums/13th-canadian-masonry-symposium/13th-cms/>
- [8] M. Dhanasekar, "Shear in Reinforced and unreinforced masonry: response, design and construction. The 12th East Asia-Pacific Conference on Structural Engineering and Construction. Elsevier Procedia Engineering 14 2069-2076, 2011 doi:10.1016/j.proeng.2011.07.260
- [9] K. Leng, C. Chintanapakdee, T. Hayashikawa "Seismic Shear Forces in Shear Walls of a medium Rise Building By Response Spectrum Analysis" Engineering Journal Volume 18 Issue 4, 2014. <http://dx.doi.org/10.4186/ej.2014.18.4.73>
- [10] P.Naik, S. Annigeri, Performance evaluation of 9 story RC building located in North Goa. 11th International Symposium on plasticity and Impact Mechanics, Implast 2016 Elsevier Procedia Engineering 173 (2017) 1841 -1846 doi:10.1016/j.proeng.2016.12.231
- [11] H. Akiyama, M. Teshigawara, H. Kuramoto, F. Kumazawa, Y. Inoue, K. Watanabe "Development and structural design guideline for medium/high rise RC wall-frame structures with flat beams" 13th World Conference on Earthquake Engineering Vancouver, B.C., Canada; 2004, paper No. 2354, 2004. https://www.iitk.ac.in/nicee/wcee/thirteenth_conf_Canada/
- [12] CEN EN 1996-1-1-2006 Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures, 2006.
- [13] CEN EN 1991-1-1-2004 Eurocode 1: Actions on structures - Part 1-1: General actions- Densities, self-weight, imposed loads for buildings, 2004.
- [14] CEN EN 1990-2004 Eurocode 0: Basics of structural design, 2004
- [15] CEN EN 1991-1-3-2005 Eurocode 1: Actions on structures - Part 1-3: General actions- Snow loads, 2005
- [16] CEN EN 1992-1-1-2004 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, 2004.
- [17] CEN EN 1998-1-2004 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, 2004.
- [18] P100-1/2013 Seismic design code – Part 1- General rules for buildings, 2013
- [19] CR 2-1-1.1/2013 Reinforced concrete walls buildings design code, 2013

Linear Evaluation on Weak Story Medium Rise Structures Placed in High Seismic Areas

Sorina Constantinescu*

Technical University of Construction Bucharest, Department of Civil Engineering, ZIP Code 011711, Romania

ARTICLE INFO

Article history:

Received: 28 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

Soft story

Partitioning Walls

Plastic Mechanism

ABSTRACT

The paper shows the influence of nonbearing masonry walls for medium rise framed buildings placed at higher stories but not at the ground floor. The building studied here is a hotel that will be built in Bucharest, Romania. This is a high seismic area according to the seismic codes in force. The structure is composed of a basement, a ground floor and 3 stories above it. It is necessary to use a frames structure to have a free partitioning at the ground story. Masonry nonbearing walls placed only at the upper stories may generate a weak story behavior for the structure. This becomes particularly difficult when the building is subjected to seismic loads. The ground floor columns are subjected to high bending moments and shear forces. This study will show the behavior of such buildings in the elastic state, but also the failure mechanism. The influence the nonbearing walls have on the structure's behavior, their ability to bear the efforts they are subjected to and whether or not this solution is usable are very important aspects to be highlighted in the study. The results may be used for framed buildings with small bays and masonry partitioning walls placed at the higher stories and not at the ground floor.

1. Introduction

The “weak story” is phenomenon that can cause serious earthquake damage. These story configurations appear from architectural requirements such as: open first floor, free architectural plan, free façade and roof terrace gardens [1]. Many medium and high rise framed buildings contain masonry infill walls. Such buildings are usually used as dwellings or office buildings with moderate bay size. The frames are usually designed without the masonry walls being taken into account [2]. There are advantages in using masonry walls as partitions for high structures. They may increase the lateral stiffness [3]. Masonry walls are thought to fail through diagonal shear. There is also another failure called corner crushing, that is not considered in the masonry design [4]. Masonry panels may generate different failure mechanisms to reinforced concrete framed structures, so they should be taken into account in the structure design [5]. The interaction between frames and masonry infill walls may generate a very different seismic response from what it was originally assumed. According to tests on model framed buildings, shear failures appeared at the slab-column connection when no masonry walls were present. If masonry infill walls were used, they prevented slab collapse and

increased the structure stiffness and strength [6]. According to laboratory tests, the reinforced concrete frames lateral strength increases and the displacement ductility decreases as a result of using infill masonry walls stiffly connected to the structure. If the masonry walls have a less stiff connection to the frames, then these effects are diminished [7]. The most common masonry units are either burned clay bricks or concrete blocks. They can be solid or can contain hollows. The concrete blocks are more brittle than the burned clay bricks. Different mortars cause masonry properties to vary [8]. The masonry prism strength decreases as the bricks height/thickness ratio increases. Masonry prisms subjected to compression strength tests show diagonal cracks close to the corners and vertical cracks in the center [9]. Masonry infill walls with rigid connections to the structure have caused several undesired effects during earthquakes, such as short-column effect, soft story effect, torsion and out of plane collapse. One of the requirements in ductile frames design is strong columns and weak beams, so most structures will have enough ductility to survive an earthquake. Elements will yield and deform, but they will bear the loads. After seismic events, it was seen that if the weak beams and strong columns rule isn't followed, plastic hinges appear at the column ends. It is also important that cast in place slabs increase the beams stiffness [10]. The presence of masonry infill walls significantly changes the dynamic response behavior of the building model compared to the bare frame model. The existence

* Corresponding Author: Sorina Constantinescu, Bucharest, 0742265890, sorina.constantinescu@yahoo.com

of a soft story shows dramatic variation in the dynamic behavior of the infill walls model, compared to the one without such a story. The story displacement, overturning moment and stiffness change from one model to the other [11]. It is adequate to analyze a structure's behavior in the plastic stage too, to predict its possible failure mechanism and the maximum base force it can bear. The moment-curvature diagram may show the ductility or stiffness of the building in study [12]. Moreover, many other researchers have also evaluated high rise composite concrete-steel structures without considering the frame wall [13-14]. The paper presents the behavior of a framed building composed of a basement, a ground floor and two floors above it. The structure will be designed using the codes in force [15-21]. It will be built in Bucharest, Romania. This is considered a high seismic area, as $a_g=0.30g$ (g is the gravity acceleration) [21].

2. Building description

The structure is composed of a basement, a ground floor with 3.5m story height and 3 floors above it with 3m story height. This study will only highlight the behavior of the structure above the basement. According to the seismic codes in force, it is considered that the columns at the ground floor are fixed at the bottoms. The basements are designed as stiff boxes, stiffer than the structures. The structure is presented in Figures 1 and 2.

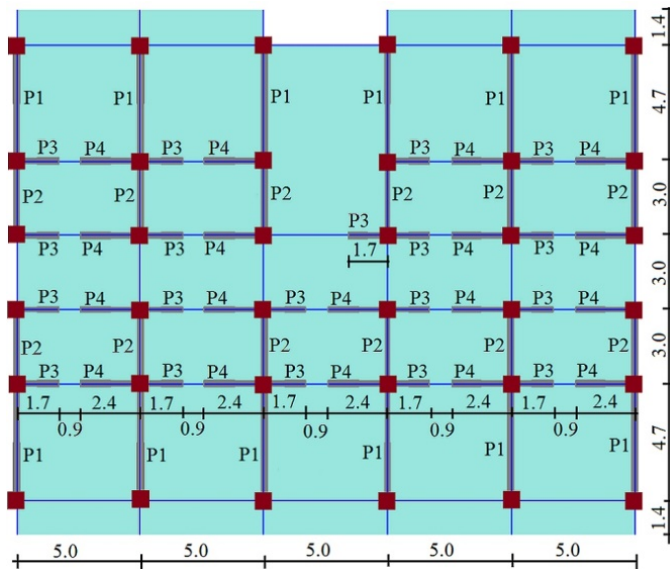


Figure 1: Story plan

The lengths seen in Figure 1 are written in meters. The particularities of this structure are the small bays and masonry walls placed at the upper stories. The ground story has no partitioning walls, so a weak story phenomenon may appear there. There are four types of nonbearing walls, according to their lengths. They will be named P1, P2, P3 and P4. The beams are blue, the slabs are green, walls are gray and columns are brown. In order to highlight the masonry walls influence to the building, 2 structural solutions are studied. One doesn't use the stiff partitioning masonry walls. It will be named solution 1. Another uses the masonry walls. It will be named solution 2. This will help to understand the nonbearing walls importance.

In the research conducted in the last decade, the finite element method has been widely considered. The finite element software used for analysis is ETABS 2016.

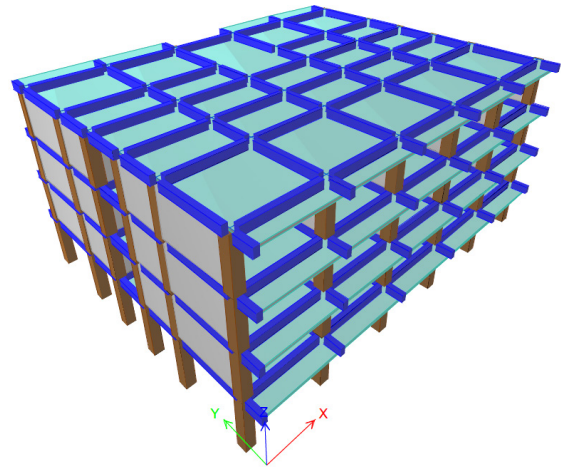


Figure 2: 3D building image

3. Theory Used in Paper

3.1. Materials Properties

To determine the behavior of frames they will have to be designed and reinforced to bear both the vertical and horizontal loads they are subjected to in the elastic state. Also, the plasticity of the materials is ignored. Moreover, the theoretical elastic and plastic behavior of steel and concrete are explained in [22-24]. Concrete used here is C20/25 [19], with elasticity modulus $E_C=30000N/mm^2$. Reinforcement bars are S355 with elasticity modulus $E_S=210000N/mm^2$ [19]. The nonbearing masonry walls are made of autoclaved concrete 600·250·240 (mm) with standard strength $f_b =5N/mm^2$, mortar M10 and elasticity modulus $E_M=3380N/mm^2$ [15]. The masonry design strengths [15] are horizontal compression (f_{dh}), vertical compression (f_d), and shear strength for horizontal direction ($f_{vd,0}$). Those design strengths are determined from their characteristic values: f_{kh} , f_k , and $f_{vk,0}$, using the masonry strength insurance factor γ_M [15]. The concrete compression design strength (f_{cd}) is determined using the characteristic strength (f_{ck}) and γ_M for concrete [19]. The steel reinforcement bars design strength is determined in an analog way.

$$f_{dh} = f_{kh}/\gamma_M = 1.91/1.9 = 0.1 N/mm^2 \quad (1)$$

$$f_d = f_k/\gamma_M = 3.38/1.9 = 1.78 N/mm^2 \quad (2)$$

$$f_{vd,0} = f_{vk,0}/\gamma_M = 0.25/1.9 = 0.13 N/mm^2 \quad (3)$$

$$f_{cd} = f_{ck}/\gamma_M = 20/1.5 = 13.3 N/mm^2 \quad (4)$$

$$f_{yd} = f_{yk}/\gamma_M = 355/1.15 = 309 N/mm^2 \quad (5)$$

3.2. Reinforced Concrete Frames Design

Bending reinforcement of beams is designed according to M_{Ed} (bending moment from the load combination: $1.0 \cdot$ permanent loads + $0.4 \cdot$ variable loads + $1.0 \cdot$ seismic loads) [15-21].

$$M_{Ed} = b \cdot \lambda x \cdot f_{cd} \cdot (d - \lambda x / 2) = A_s \cdot f_{yd} \cdot z \quad [\text{kNm}] \quad (6)$$

$$m = M_{Ed} / (b \cdot d^2 \cdot f_{cd}) \quad (7)$$

$$z = d - \lambda x / 2 = d \cdot (1 - (1 - 2m)^{0.5}) / 2 \quad [\text{mm}] \quad (8)$$

$$A_{s,\min} = \min\{0.26 \cdot f_{ctm} / f_{yk} \cdot b \cdot d; 0.0013 \cdot b \cdot d\} \quad (9)$$

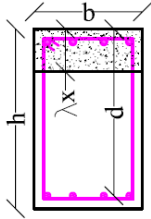


Figure 3: Concrete beam section

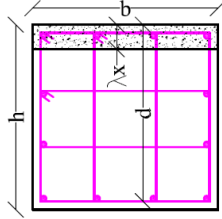


Figure 4: Concrete column section

A_s is the minimum reinforcement area for beams. λx is the beam section compressed area [19] length. $f_{ctm} = 2.6 \text{ N/mm}^2$ is the medium value of the concrete tensile strength. Bending moments of columns are calculated using (10), according to [21].

$$\Sigma M_{Rc} \geq 1.20 \cdot \gamma_{Rd} \cdot \Sigma M_{Rb} \quad [\text{kNm}] \quad (10)$$

$\gamma_{Rd} = 1.2$ is the steel stiffening factor for DCM (medium ductility buildings) [21], ΣM_{Rc} and ΣM_{Rb} are the sums of bearing bending moments of columns and beams along a line. Minimum longitudinal reinforcement percent value for columns is $p_{\min} = 1\%$ and the maximum is $p_{\max} = 4\%$ [21]. If $\lambda x < 2 \cdot a_s$, A_s will be determined from (13), and from (14), if $\lambda x \geq 2 \cdot a_s$. Here, $a_s = 45 \text{ mm}$. N_{Ed} is the axial force in the calculated columns [21].

$$p = A_s / (b \cdot d) \cdot 100 \quad (11)$$

$$x = N_{Ed} / (b \cdot \lambda \cdot f_{cd}) \quad [\text{mm}] \quad (12)$$

$$A_s = [M_{Ed} - N_{Ed}(d - a_s) / 2] / [f_{yd} \cdot (d - a_s)] \quad [\text{mm}^2] \quad (13)$$

$$A_s = [M_{Ed} + N_{Ed}(d - a_s) / 2 - b \cdot \lambda x \cdot f_{cd}(d - \lambda x / 2)] / [f_{yd}(d - a_s)] \quad (14)$$

3.3. Seismic Base Force

The base force F_b is calculated according to [21]. The factors in the formula are: $\beta_0 = 2.5$ is the maximum elastic spectrum value, q is the structure's behavior factor, $\gamma_{I,e} = 1.2$ is the building's importance-exposure coefficient, $q = 3.5 \cdot \alpha_u / \alpha_1 = 3.5 \cdot 1.35$ [21], α_u / α_1 is the base shear force value for the failing mechanism/the base shear force value for the first plastic hinge, m = building's mass. $\lambda = 0.85$ for 3 stories buildings, $a_g = 0.30g$ [21], G = building's weight and c_s is the seismic coefficient.

$$F_b = \gamma_{I,e} \cdot \beta_0 \cdot a_g / q \cdot m \cdot \lambda = c_s \cdot G = 0.17 \cdot G \quad [\text{kN}] \quad (15)$$

3.4. Masonry Walls Design Method

$$A_1 = N_{Ed} / (0.85 \cdot f_d) \quad [\text{mm}^2] \quad (16)$$

$$M_{Rd} = N_{Ed} \cdot y_z \quad [\text{kNm}] \quad (17)$$

A_1 is the wall's compressed area and M_{Rd} is the walls bearing bending moment. y_z is the distance between the wall's weight center (C) and the compressed autoclaved concrete area weight center (A_{1G}) [15]. S is the seismic action direction.

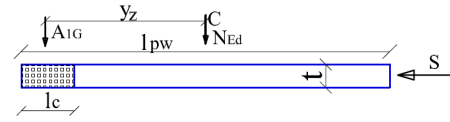


Figure 5: Autoclaved concrete wall section

$$V_{Rd} = 0.4 \cdot (N_{Ed} + 0.8 \cdot V_{Ed} \cdot h_{pan} / l_{pw}) \quad [\text{kN}] \quad (18)$$

$$V_{Ed} \leq l_{pw} \cdot t \cdot f_{vd,0} \quad (19)$$

V_{Rd} is the masonry wall bearing shear force and V_{Ed} is the horizontal shear force from the seismic loads combination. $f_{vd,0}$ is the design initial shear strength for no axial stress [15], h_{pan} and l_{pw} are the height and length of the masonry panel.

3.5. Seismic Force Perpendicular to the Masonry Walls

$$F_{NBW}(z) = \gamma_{I,e} \cdot a_g \cdot \beta_{NBW} \cdot k_z \cdot m_{NBW} / q_{NBW} = 0.7 \text{ kN/m}^2 \quad (20)$$

The force is considered uniformly distributed, perpendicular to the non-bearing walls [21]. The factors in the formula are explained here. $\beta_{NBW} = 1$ is the non-bearing walls amplification factor, k_z is a coefficient according to the non-bearing wall's level (the distance to the building's base), z is the non-bearing wall's level and H is the building height [21]. k_z is expressed by factors k_{z1} and k_{z2} that refer to the highest and lowest points of the nonbearing wall. Of course, the greatest value for k_z is calculated at the top building story. $q_{NBW} = 2.5$ is the behavior factor for non-bearing walls. $m_{NBW} = \gamma_{mas} \cdot t = 7 \cdot 0.25 = 1.75 \text{ kN/m}^2$ is the wall mass/ m^2 .

$$k_z = 1 + 2 \cdot z / H \quad (21)$$

$$k_z = (k_{z1} + k_{z2}) / 2 = (1 + 2 \cdot z_1 / H + 1 + 2 \cdot z_2 / H) / 2 \quad (22)$$

Seismic force value F_{NBW} is limited as (23) shows [21].

$$0.75 \cdot \gamma_{I,e} \cdot a_g \cdot m_{NBW} \leq F_{NBW} \leq 4 \cdot \gamma_{I,e} \cdot a_g \cdot m_{NBW} \quad (23)$$

4. Elastic Analysis Results

4.1. Natural Vibration Periods

The natural vibration periods for the first 3 natural vibration modes are 20% smaller if stiff partitioning walls are used.

Table 1 Natural vibration periods

Natural vibration periods	Solution without stiff walls (solution 1)	Solution with stiff walls (solution 2)
T1	0.510 s	0.398 s
T2	0.462 s	0.359 s
T3	0.443 s	0.356 s

4.2. Story Displacements in the Elastic Stage

For the first solution with no nonbearing walls, the displacements reach slightly greater values at story 1 and then, they visibly increase towards story 4. For the second solution, where the nonbearing walls are present at the upper stories, displacements

are smaller and the values gradually increase towards the building top.

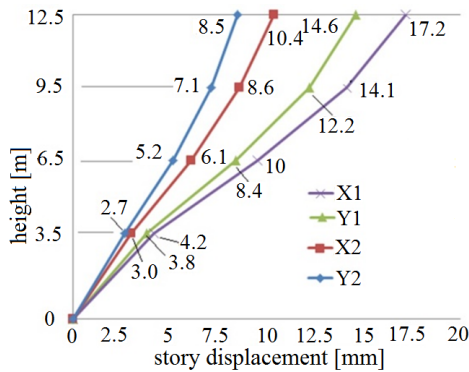


Figure 6: Story elastic displacements

4.3. Frames Efforts and Reinforcements

Efforts in beams and columns are shown in Figures 7–16 for the load combination with the seismic load on direction X (GX) and on Y (GY). The dark blue graphic lines show the efforts in frames for the first solution and the light blue graphic lines show effort values for the second one.

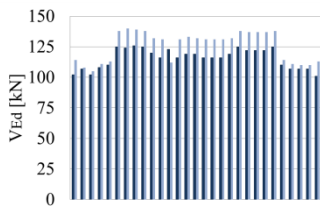


Figure 7: V_{Ed} in beams on direction X

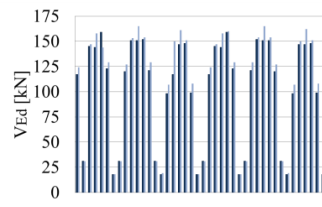


Figure 8: V_{Ed} in beams on direction Y

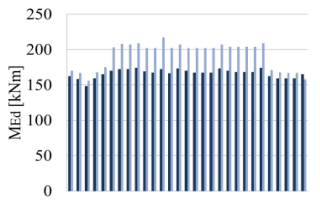


Figure 9: M_{Ed} in beams on direction X

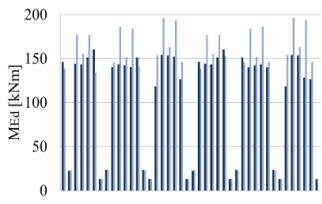


Figure 10: M_{Ed} in beams on direction Y

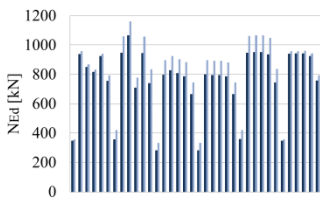


Figure 11: N_{Ed} in columns on X

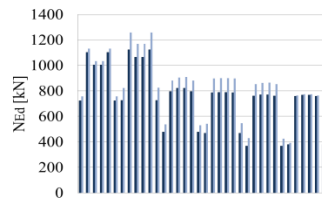


Figure 12: N_{Ed} in columns on Y

It is clear that the presence of nonbearing walls increases the frames efforts. There are 29 beams on direction X, 42 on Y and 36 columns. Each one of the line graphics is the effort value in a frame or a column. The beams with very low efforts on Y are the short beams at the edges. There are some beams on direction X with no masonry walls to bear. Their effort values do increase for the case when the walls are present. This is because of the influence from the other elements in the structure.

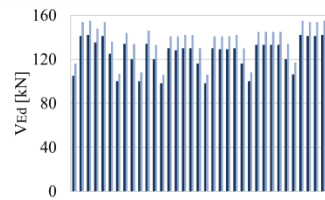


Figure 13: V_{Ed} in columns on X

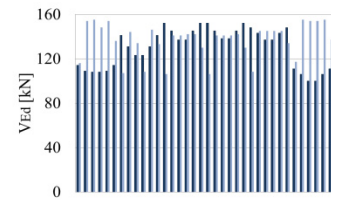


Figure 14: V_{Ed} in columns on Y

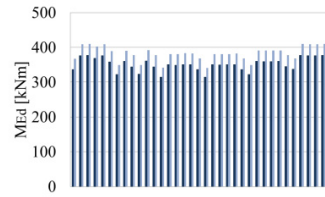


Figure 15: M_{Ed} in columns on X

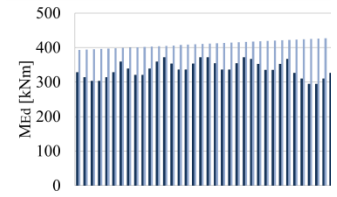


Figure 16: M_{Ed} in columns on Y

If no stiff partitioning walls are present the efforts in columns are about 90% of these that are seen if nonbearing autoclaved concrete walls are used.

Table 2: Frame elements reinforcement bars

Beam 30x50 As →4Φ22 up and 4Φ20 down (solution 1)	Beam 30x50 As →4Φ25 up and 4Φ22 down (solution 2)	Column 60x60 As →12Φ22 (both solutions)

4.4. Masonry Nonbearing Walls Results

Table 3 Design and bearing efforts in masonry walls

Wall (P)	N_{Ed} [kN]	M_{Ed} [kNm]	V_{Ed} [kN]	M_{Rd} [kNm]	V_{Rd} [kN]
P1	28.2	10.7	3.57	65	11.9
P2	18	6.9	2.3	26.4	7.8
P3	10.2	3.9	1.2	8.47	7.64
P4	14.4	5.5	1.8	17	6.36

5. Nonlinear Stage Results

5.1. Plastic Mechanisms

The nonlinear analysis is performed for solution 1 by using pushover cases PX and PY and for solution 2, by using cases PXW and PYW. This way, the masonry walls influence on the building is clearly seen on both directions X and Y. Figures 17 – 20 show the last stages of the analysis. The color code is the following: B (green) means the plastic hinge is formed, C (light blue) means the plastic hinge reaches the limit and the element gives out, D (pink) means the load was redistributed and E (red) means collapse.

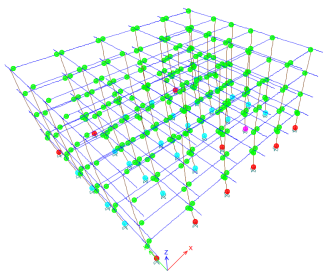


Figure 17: PX step 11 solution 1

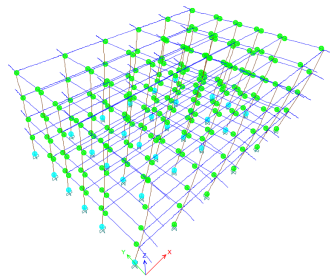


Figure 18: PY step 22 solution 1

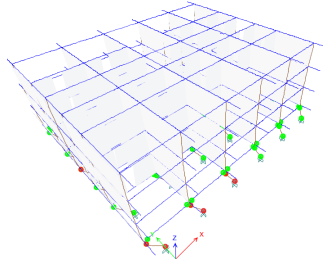


Figure 19: PXW step 7 solution 2

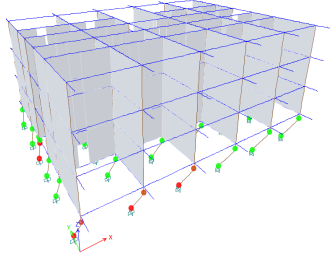


Figure 20: PYW step 25 solution 2

If stiff nonbearing walls are not used, the plastic mechanism on direction X is created when plastic hinges reach stage D at some columns bottoms and E at others. There are plastic hinges at all the beams ends on both directions. The last nonlinear analysis step on direction Y shows plastic hinges in stage C at the columns bottoms and in stage B at the beams ends on direction Y. If stiff masonry walls are used, the plastic mechanism is formed when plastic hinges reach stages D and E at the ends of the ground floor columns. There are plastic hinges at the ground story beams ends on direction X for case PXW and on Y for PYW. There are no plastic hinges seen in the floors above. It looks as if the ground floor columns are folded. The nonbearing walls do have a stiffening effect to the upper stories.

5.2. Plastic Stage Story Displacements

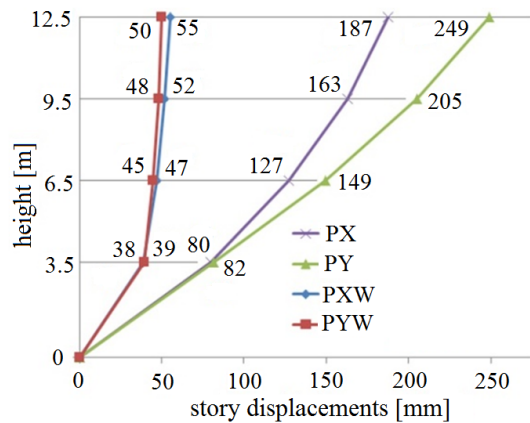


Figure 21: Story plastic displacements

For solution 1, where no masonry walls are present, the displacements reach high values. These values are greater for direction Y. The story displacements for the structure with masonry walls at the upper stories reach a certain value at story 1 and then show a slight increase towards the building top. The values are very similar for both X and Y directions. The values are 4 to 5 times smaller than for solution 1.

5.3. Pushover Diagrams

For cases PX and PY, where no partitioning walls are present, the maximum displacements are 5 times greater than for PXW and PYW. The maximum base forces reached for PY, PXW and PYW are similar. For case PY, the maximum force reached is smaller. Cases PX and PY diagrams show a slightly reduced rigidity at the beginning of the analysis, compared to PXW and PYW.

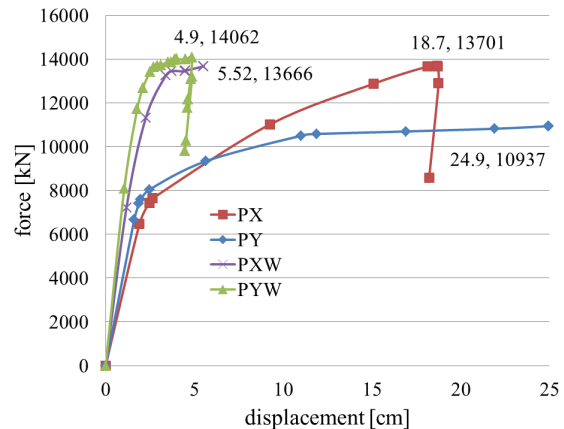


Figure 22: Pushover diagrams

When the base force approaches 8000kN, the rigidity visibly drops more for PY and less for PX. As the base force reaches 11000kN, PY rigidity drops nearly to 0 and is maintained like that until the analysis ends. For case PX, the base force nears 14000kN and then the base force drops while the displacement is maintained almost the same. This may be explained by the plastic hinges that reach collapse in the ground floor columns. Case PYW shows a greater rigidity than PXW from the beginning of the analysis to the maximum force. After this force is reached, the rigidity drops to nearly 0 for both cases. There is a slight stiffness increase for case PXW before collapse, but PYW shows a drop in force while the displacements remains the same.

5.4. Masonry Nonbearing Walls Results

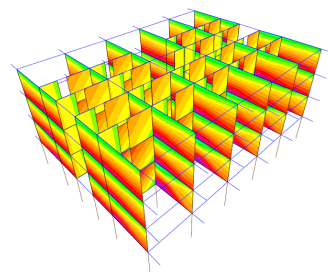


Figure 23: PXW step 2
 $\sigma_x = 0.1 \text{ N/mm}^2$

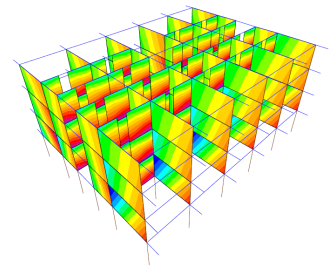


Figure 24: PYW step 8
 $\sigma_x = 0.1 \text{ N/mm}^2$

Stress σ_x surpasses strength f_{dh} from the nonlinear analysis second step. Both PXW and PYW cases generate corner crushing in the masonry panels on their directions. The highest stresses are in the walls at story 2. In the panels on the perpendicular direction, the stresses values increase from the bottom to top and also change the sign. This indicates a wall folding at about one third on story height measured from top to bottom. The stress reaches approximately the same values on both directions.

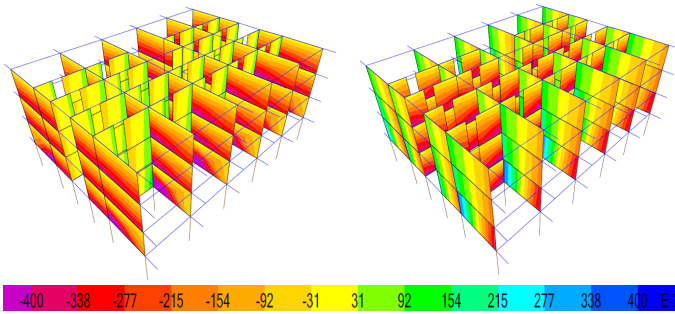


Figure 25: PXW step 8
 $\sigma_z=0.4\text{N/mm}^2$

Figure 26: PYW step 8
 $\sigma_z=0.4\text{N/mm}^2$

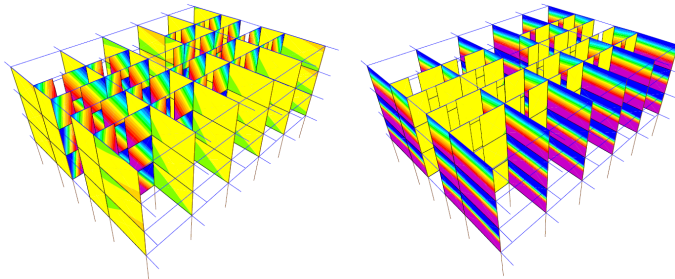


Figure 27: PXW step 1
 $\tau_{xz}=0.15\text{N/mm}^2$

Figure 28: PYW step 1
 $\tau_{xz}=0.15\text{N/mm}^2$

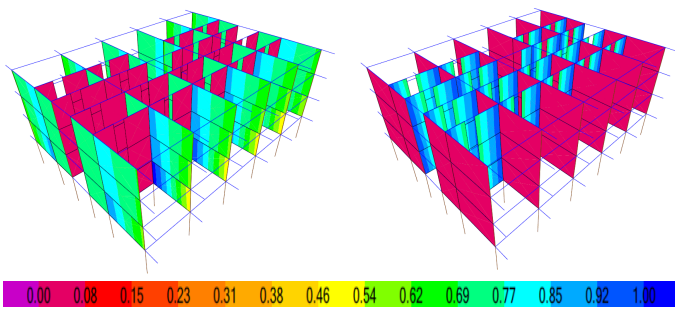


Figure 29: PXW step 1 F=1kN/m²

Figure 30: PYW step 1 F=1kN/m²

Stress σ_z remains below strength f_d throughout the nonlinear analysis. The stress pattern shows horizontal stripes indicating increased values from the wall top to bottom on the direction perpendicular to the load case. On the nonlinear case direction, the stress value stripes are vertical. This shows crushing on one side of the walls and stretching on the other.

Stress τ_{xz} surpasses strength $f_{v,d,0}$ from the first step of the analysis for both pushover cases. The stress values are very clearly seen in walls on the same direction as the pushover case. For PXW, the stresses are particularly increased at the walls corners. For PYW the stresses show horizontal stripes. The maximum values are reached both at the walls tops and bottoms, but the signs are opposite. There are stripes of zero stress at one third of the wall height from the top. This means the walls are „cut” at that zero stress line.

The horizontal seismic force perpendicular to the partitioning walls surpasses the theory design value from the first step of the analysis. The force values are practically 0 in walls parallel to the nonlinear case direction. In walls perpendicular to the loading case, the force values are seen as vertical stripes. The highest values are reached closer to the frame columns.

6. Conclusions

The autoclaved concrete walls are lightweight but they do have an influence on the structure. If the masonry walls are not present, the efforts values in frames are about 90% of these reached otherwise. This demands greater reinforcements in beams. The natural vibration periods are decreased by 20% if stiff partitioning walls are used. The walls can bear their efforts in the elastic stage. The plastic mechanism is greatly influenced by the nonbearing walls presence. The ground story columns are particularly affected, as plastic hinges form only there and develop until collapse. The pushover diagrams are greatly affected by the autoclaved concrete walls. They increase the structure rigidity and the maximum base force. Horizontal stresses surpass the masonry strengths at the first steps of the analysis. The horizontal seismic force reaches greater values than the theory design value from the first step of the analysis. The autoclaved concrete walls work well in the elastic stage. In the plastic stage they reduce the structure ductility and increase the maximum base force. The masonry walls reduce story displacements for both elastic and plastic stage.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] T. Guevara-Perez, „Soft Story and Weak Story in Earthquake Resistant Design: A multidisciplinary Approach” in 15 World Conferences on Earthquake Engineering (WCEE) Lisboa 2012 https://www.iitk.ac.in/nicee/wcee/article/WCEE2012_0183.pdf
- [2] C.V.R. Murty, S.K. Jain, “Beneficial influence of masonry infill walls on seismic performance of RC frame buildings” in 12th World conference on Earthquake Engineering (WCEE) Auckland, New Zealand, 2000. <http://www.iitk.ac.in/nicee/wcee/article/1790.pdf>
- [3] Madan A and Hashimi AK. “Performance Based Design of Masonry Infilled Reinforced Concrete Frames for Near-Field Earthquakes Using Energy Methods” WASET, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering Vol:8, No:6, p 689-695, 2014. <https://pdfs.semanticscholar.org/bbdd/eb245adce7d0b1604f474ad249e5a57c1397.pdf>
- [4] M. Dhanasekar, “Shear in Reinforced and unreinforced masonry: response, design and construction” in The 12th East Asia-Pacific Conference on Structural Engineering and Construction. Elsevier Procedia Engineering 14 2069-2076, 2011. doi:10.1016/j.proeng.2011.07.260
- [5] A. Furtado, H. Rodriguez, A. Arede, H. Varum, “Experimental evaluation of out-of-plane capacity of masonry infill walls” Elsevier Engineering Structures Vol. 111, p 48-63, 2016. <https://doi.org/10.1016/j.engstruct.2015.12.013>
- [6] S. Pujol, A. Benavent-Climent, M. E. Rodriguez, J.P. Smith-Pardo, “Masonry infill walls: an effective alternative for seismic strengthening of low-rise reinforced concrete building structures” in The 14th world conference on earthquake engineering. Beijing, China 2008 <http://www.jdwhite.com/sites/default/files/uploadedfiles/tp-Masonry Infill Walls.pdf>
- [7] H. Jiang, X. Liu, J. Mao, “Full scale experimental study on masonry infilled RC moment-resisting frames under cyclic loads”, Elsevier Engineering structures 91, p70-84, 2015. <http://dx.doi.org/10.1016/j.engstruct.2015.02.008>
- [8] M. Teguh, “Experimental evaluation of masonry infill walls of RC frame buildings subjected to cyclic loads” Elsevier Procedia Engineering 171 Sustainable Civil Engineering Structures and Construction Materials, p 191-200, 2016. doi:10.1016/j.proeng.2017.01.326
- [9] N. N. Thaickavil, J. Thomas, “Behavior and strength assessment of masonry prisms” Elsevier. Case Studies in Construction Materials 2018 Vol. 8, p 23-38. <https://doi.org/10.1016/j.cscm.2017.12.007>
- [10] B. Li, Z. Wang, K. Mosalam, H. Xie Wenchuan, “Earthquake Field Reconnaissance on Reinforced concrete on reinforced concrete framed buildings with and without masonry infill walls” in The 14th world conference on earthquake engineering. Beijing, China, 2008 <https://scholar.google.ro/>

- [11] A. Abd-Elhamed, S. Mahmoud, "Linear and Nonlinear Dynamic Analysis of Masonry Infill RC Framed Buildings" *C.E.J.* Vol. 3 Nr. 10, 2017. <http://dx.doi.org/10.28991/cej-030922>
- [12] P. Naik, S. Annigeri, "Performance evaluation of 9 story RC building located in North Goa" 11th International Symposium on plasticity and Impact Mechanics, Implast, Elsevier Procedia Engineering 173, 2017. p 1841 - 1846, 2016. doi:10.1016/j.proeng.2016.12.231
- [13] R. Rahnavard, F. Fathi Zadeh Fard, A. Hosseini, M Suleiman, "Nonlinear analysis on progressive collapse of tall steel composite buildings", Elsevier Case Studies in Construction Materials 8 ,2018 p 359-379. <https://doi.org/10.1016/j.cscm.2018.03.001>
- [14] R. Rahnavard, N. Siahpolo, "Function comparison between moment frame and moment frame with centrally braces in high-rise steel structure under the effect of progressive collapse", *Journal of structure and construction engineering*, Volume 4, Issue 4 - Serial Number 14, p 42-57, doi: 10.22065/jsce.2017.77865.1084
- [15] CEN EN 1996-1-1-2006 Eurocode 6: Design of masonry structures - Part 1-1: General rules for reinforced and unreinforced masonry structures, 2006.
- [16] CEN EN 1991-1-1-2004 Eurocode 1: Actions on structures - Part 1-1: General actions- Densities, self-weight, imposed loads for buildings, 2004.
- [17] CEN EN 1990-2004 Eurocode 0: Basics of structural design, 2004.
- [18] CEN EN 1991-1-3-2005 Eurocode 1: Actions on structures - Part 1-3: General actions- Snow loads, 2005.
- [19] CEN EN 1992-1-1-2004 Eurocode 2: Design of concrete structures - Part 1-1: General rules and rules for buildings, 2004.
- [20] CEN EN 1998-1-2004 Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings, 2004.
- [21] P100-1/2013 Seismic design code – Part 1- General rules for buildings, 2013.
- [22] R.Rahnavard, A. Hassanipour, A. Mounesi, "Numerical study on important parameters of composite steel-concrete shear walls", *Journal of Constructional Steel Research* 121 2016 p 441-456. doi:10.1016/j.jcsr.2016.03.017
- [23] R. Rahnavard, M. Naghavi, M. Abudi, M. Suleiman, "Investigating Modeling Approaches of Buckling-Restrained Braces under Cyclic Loads", Elsevier Case Studies in Construction Materials, Case Studies in Construction Materials 8, 2018 p 476-488. <https://doi.org/10.1016/j.cscm.2018.04.002>
- [24] R. Rahnavard, A. Hassanipour, M. Suleiman, A. Mokhtari, "Evaluation on eccentrically braced frame with single and double shear panels", Elsevier *Journal of Building Engineering* 10, 2017 p 13-25. <http://dx.doi.org/10.1016/j.jobbe.2017.01.006>

A Wi-Fi based Architecture of a Smart Home Controlled by Smartphone and Wall Display IoT Device

Tareq Khan*

School of Engineering Technology, Eastern Michigan University, Michigan, Ypsilanti, 48197, United States

ARTICLE INFO

Article history:

Received: 31 July, 2018

Accepted: 10 August, 2018

Online: 14 November, 2018

Keywords:

Android Things

Raspberry Pi

Wi-Fi

ABSTRACT

In this age of smart devices, many people are carrying a smartphone with them all the time. When they are at home, most of them are connected with the home Wi-Fi network. In this paper, a Wi-Fi network based architecture is proposed to control home appliances using a smartphone and also with a touchscreen-based wall display panel. The proposed system enables the user to control appliances from anywhere in the home without the pain of walking towards the switch panel on the wall. In this project, the mechanical switch based panel on the wall is replaced by the state-of-the-art touch-based liquid crystal display. Along with buttons, the display also shows current weather and time widgets. The smartphone app and a prototype of the display panel using Raspberry Pi with Android Things operating system is developed and tested.

1. Introduction

In a recent survey, respondents said that Wi-Fi Internet is more important to them than television, alcohol, sex, and everything else besides food. In this age of smart devices, many people – both young and adult - are carrying a smartphone with them all the time. When they are at home, most of them are connected with the home Wi-Fi network. Homes today have an average of 8 devices connected to the Wi-Fi and at the end of 2019, households worldwide will have more than 10 billion devices capable of connecting to their home network router [1]. Globally, total public Wi-Fi hotspots will grow sevenfold from 2015 to 2020, from 64.2 million in 2015 to 432.5 million by 2020. Wi-Fi hotspots will be key for the development of the Internet of Things (IoT) applications and services [2].

A smart home is equipped with electronic devices, such as sensors, microprocessors, relays, and appliances - that are tied to a network and controlled via interfaces such as phones, webs, or embedded computers [3], [4]. In this paper, a Wi-Fi network based smart home architecture is proposed to control home appliances using a smartphone and also with a touchscreen-based wall display IoT device. The proposed system enables the user to control appliances from anywhere in the home using a smartphone app - without the pain of physically walking towards the switch panel on the wall. In the proposed system, the smartphone acts as a remote control for the home appliances. With the smartphone in hand –

the appliances can be operated while lying on the bed, sitting on the couch, or even during eating – without the need of line-of-sight of traditional infrared based remotes. In this project, the mechanical switch based panel on the wall is also replaced by the state-of-the-art touch screen based liquid crystal display (LCD) unit which is connected to Wi-Fi. The display shows buttons and also shows the current weather and time widgets. The smartphone app and a prototype of the wall display unit using Raspberry Pi with Android Things operating system is developed and tested.

2. Related Work

Several recent related works are found in the literature about the design and architecture of a smart home. In [5], dual tone multi-frequency (DTMF) is used to control the home appliances remotely via a GSM network. The idea is to employ a DTMF decoder through which home appliances can be controlled by dialing a predefined number, via smartphone, for a specific load. In [6], the proposed hardware architecture is composed of remote user application terminals, the terminal control unit and various terminals distributed in different rooms. The communication between the terminal control unit and the field terminals is implemented using ZigBee technology and GSM/GPRS module is used for information transmission of sensor data and remote commands. Remote users can browse the household temperature and humidity data using a mobile app or the web page. Here, the proposed work can only do monitoring, but unable to control loads. In [7], a smart home environment monitoring system is designed with Raspberry Pi based smart home controller, EnOcean/Wi-Fi

*Corresponding Author: Tareq Khan, Eastern Michigan University, 118 Sill, Ypsilanti, MI 48197, United States. Email: tareq.khan@emich.edu.

wireless sensor control network, and smart home PC client. In [8], a solution is developed to control home appliances like light, fan, door cartons, energy consumption, and level of the Gas cylinder using various sensors and microcontrollers such as Node MCU ESP8266 and Arduino UNO. It provides information about the energy consumed by the house owner regularly in the form of a message. Also, it checks, the level of gas in the gas cylinder - if it reaches lesser than the threshold, it automatically places the order for refill gas cylinder and sends a reference number as a message to the house owner. However, load control using a smartphone app and using wall display panel is not discussed in [7] and [8]. In [9], the design of a Smart Power Strip is discussed where the loads can be turned on/off using a smartphone app that communicates wirelessly with the microcontroller using Bluetooth. In [18], a smart home architecture is proposed using ZigBee, Wi-Fi, and Ethernet. Here a ZigBee to Wi-Fi data conversion service is required, where the proposed work in this paper only uses Wi-Fi and do not require the network protocol conversion overhead. The work in [19] uses Arduino Ethernet, where this work uses Raspberry Pi with the state-of-the-art Android Things operating system for controlling the loads.

Compared with the other works, the proposed work uses the home Wi-Fi network. As most of the smart devices such as smartphones, tablets etc. are always connected with home Wi-Fi, no additional connection such as Bluetooth, ZigBee etc. are required – which can save battery life of the phone. Using the proposed system, electrical loads can be turned on/off using both smartphone app and a Wi-Fi connected LCD with a touchscreen device that can be attached on the wall – replacing the old-fashioned mechanical switch based panel. The wall display also shows colorful weather forecast and time widgets. To use the proposed system, the password of the home Wi-Fi network must be known to the user – thus the system is fairly secured from hacking.

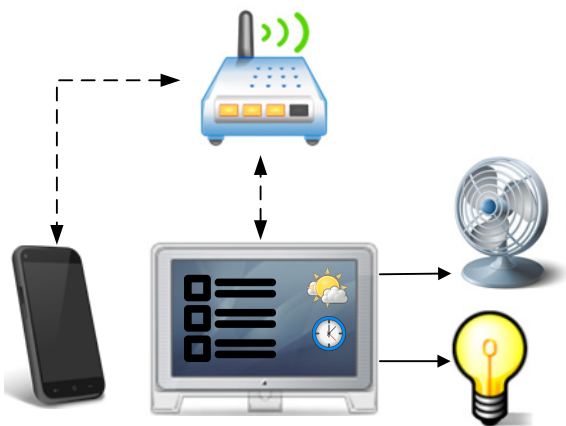


Figure 1: The architecture of the proposed system: Smartphone (a) connected to the wireless router (b) using Wi-Fi. The Raspberry Pi based IoT display panel (c) connected with the router (b), showing buttons, weather and time widgets. Loads (d) and (e) are connected with the Raspberry Pi ports of the display panel using solid state relays.

3. System Architecture

The overall architecture of the proposed system is shown in Figure 1. In Figure 1., the smartphone (a) and the wall display panel (c) gets connected to the wireless router (b) using Wi-Fi. www.astesj.com

Loads such as light (d) and fan (e) are connected with the wall display panel using solid state relays (SSR). The loads can be turned on/off by using the buttons of the display panel. The display panel also shows weather and time widgets. Loads can also be controlled using the Smartphone app wirelessly. A brief description of the Wi-Fi connected wall display panel and the smartphone app are given below.

3.1. Wall Display Panel

In the proposed system, the mechanical switch based panel on the wall is also replaced by the state-of-the-art touch screen based LCD unit which is connected with Wi-Fi. The display also shows current weather and time widgets. The hardware and software parts of the device are described below.

Hardware: The hardware architecture of the display panel is shown in Figure 2. A Raspberry Pi (RPI) v3 [10] single board computer is used as the main processing unit. It containing a 1.2 GHz 64-bit quad-core ARMv8 microprocessor, 1 GB of RAM, micro SD card slot supporting up to 32 GB, 40 general purpose input output (GPIO) pins with two pulse width modulation (PWM) module, onboard Wi-Fi module, and other built-in hardware peripherals such as universal asynchronous receiver/transmitter (UART), I2C, serial peripheral interface (SPI), universal serial bus (USB), audio output via 4-pole 3.5mm connector, LCD interface (DSI) etc. A 7-inch capacitive touch LCD [11] is interfaced with the RPi using the DSI and the I2C port. The LCD has 24-bit color depth and the screen resolution of 800 × 480 pixels.

Two solid-state relays (SSR) [12] are used, as shown in Fig. 2., to control the AC loads. The SSRs are controlled by two GPIO pins - GPIO18 and GPIO13 - of the RPi. The contact of the SSR can carry a maximum of 40 A current. Finally, the power supplies for the RPi board and the touch LCD are supplied using 110V AC to 5.1V DC adapters [13].

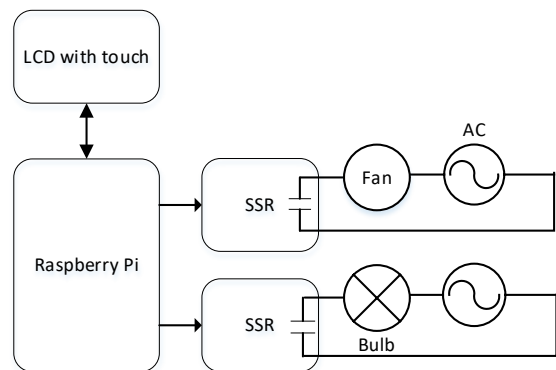


Figure 2: Block diagram of the wall display panel hardware.

App: The Android Things [14] embedded operating system is installed on a 16 GB secure digital (SD) card of the RPi board. Android Things is recently developed by Google, aimed to be used with low-power and memory constrained IoT devices. The platform can run any program targeting Android and it also comes with libraries for accessing hardware peripherals such as UART, I2C, and GPIO of the RPi. The device gets connected to the home Wi-Fi and can access the Internet.

The app of the display panel shows the *on* and *off* buttons for each load. The current on-off status is also displayed using labels.

The display panel also shows the current time and weather widget on the screen. Hypertext Markup Language (HTML) code was embedded in web views of the app for fetching time [15] and weather [16] information from the Internet.

The wall display panel is configured as a *server* and the smartphone is configured as a *client* for socket-based data communication in the local Wi-Fi network. The display panel's app shows its local Internet Protocol (IP) address, so the user can input that IP in the smartphone to establish a socket connection. The dynamic host configuration protocol (DHCP) server of the router dynamically assigns IP to its connected devices, so the IP address of the display panel may change on a new connection depending on how many devices are currently connected with the router. If IP of the display panel changes, then the user needs to re-enter the IP in the smartphone app to establish a socket connection each time– which is tiresome. To solve this problem, the IP of the display panel is fixed to 192.168.1.20 by configuring the router [17], so that the smartphone does not have to reenter the display panels IP.

Whenever On/Off button is pressed for a particular load in the display panel, a status file – named *status.dat* - is updated with the new value. This file contains the *on-off* status of all the AC loads. Then depending upon the content of the *status.dat* file, GPIO pins are either set or reset. After that, the labels showing the status of the loads are updated based on the *status.dat* file. If the display panel is already connected to the smartphone, then the load status is synchronized with the smartphone. To do that, a command data frame, as shown in Figure 3, is sent to the smartphone using socket-based data communication.

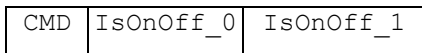


Figure 3: Command data frame.

In Figure 3, CMD is a byte containing the nature of the command or the response. When the display panel sends its status to the smartphone, the CMD field contains the constant `CMD_SERVER_RESPONSE`. The `IsOnOff_<n>` contains the Boolean on-off status; here `<n>` represents the load ID. After the phone receives the command frame, it updates its status with the new values and the status is synchronized with the smartphone.

3.2. Smartphone App

The smartphone app is developed for the Android platform. The first screen of the app shows the *on* and *off* buttons for each load. The current on-off status is also displayed using labels. The app contains a *settings* menu, where the user can input the IP of the wall display panel. The first screen of the app contains a *connect* button – when pressed by the user – a socket connection is established with the display panel through the router using Wi-Fi.

After the smartphone gets connected with the display panel using the socket connection for the first time, the phone sends a request to the display panel to get the current status of the loads. To do that, the smartphone sends a command data frame where the CMD field contains the constant `CMD_GET_STATUS` and the remaining fields contain *don't care* values. After receiving the

`CMD_GET_STATUS` command, the display panel app reads the contents of the *status.dat* file and makes a command frame where CMD field is set to `CMD_SERVER_RESPONSE` and the remaining fields contain the status of the loads. The command is then sent and after the phone receives the data frame, it updates its status with the new values and the status is synchronized.

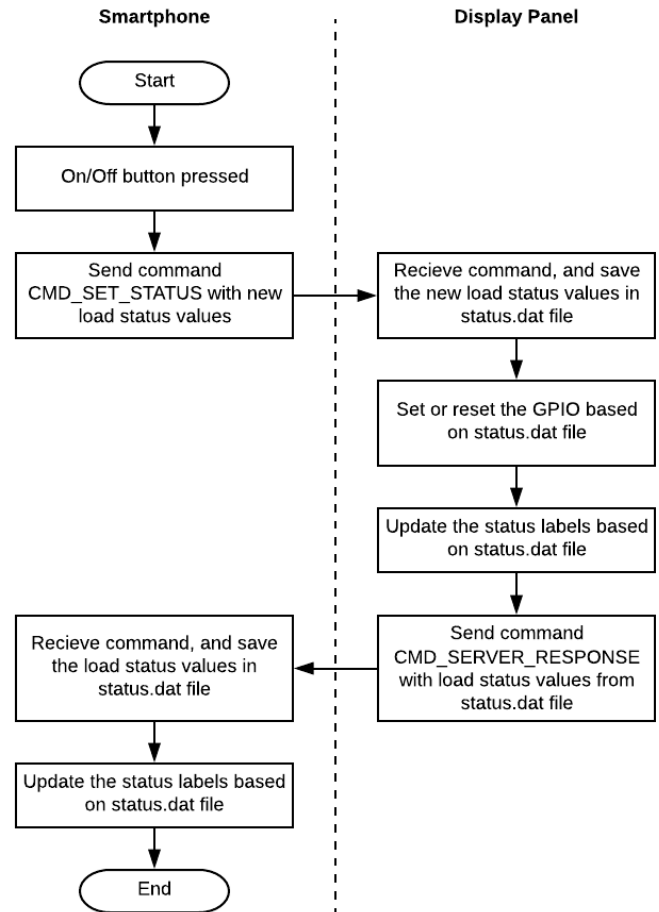


Figure 4: Flowchart when on/off button is pressed in the smartphone.

Whenever On/Off button is pressed for a particular load in the smartphone, the app makes a command frame - where CMD field contains the constant `CMD_SET_STATUS` and remaining fields contain the new values for the loads. The command is then sent to the display panel using socket connection. When `CMD_SET_STATUS` command is received, the display panel app updates the *status.dat* file with the new values. Then depending upon the *status.dat* file – the display panel updates the GPIO pins, updates the status labels and then sends a response data frame to the phone where the CMD field contains the constant `CMD_SERVER_RESPONSE` and remaining fields contains the new status of the loads. After the phone receives the data frame, it updates its status with the new values, stores them in the *status.dat* file, updates the status labels and status is synchronized. The flowchart of the operation is shown in Figure 4.

4. Result

A prototype of the proposed Wi-Fi-based smart home control system as discussed in Section 3 is developed and tested

successfully. The photograph of the experimental setup is shown in Figure 5. Loads can be controlled using the display panel and also with the smartphone with a 100% success rate. The status of the smartphone app automatically gets updated when a load is controlled by the display panel and vice-versa.

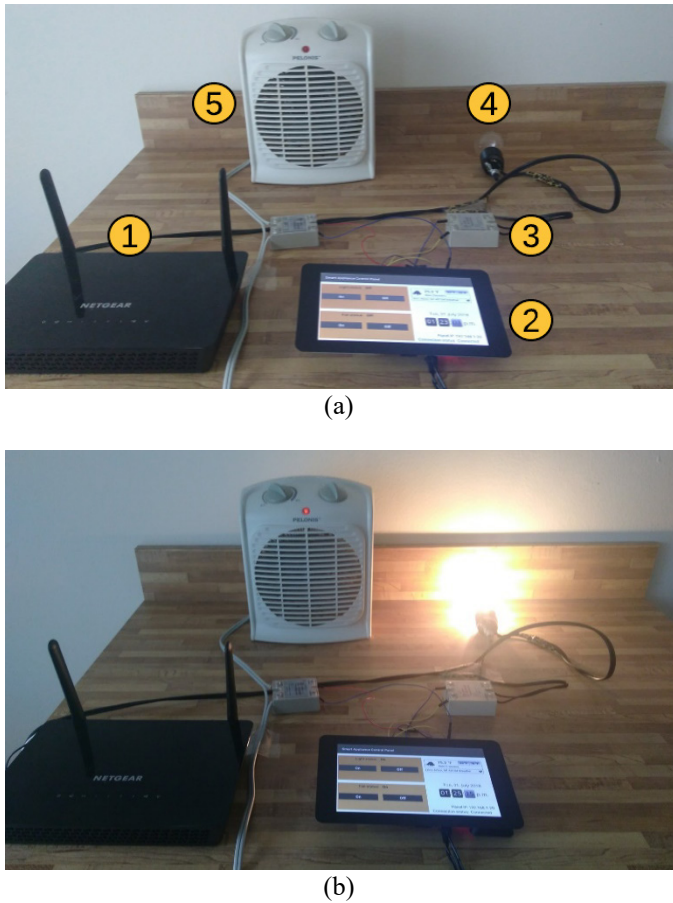


Figure 5: The experimental setup: (a) wireless router (1), wall display IoT device (2), SSR (3), Bulb at off state(4), Fan at off state (5); (b) Bulb and fan at on state.

The wall display panel IoT device, as shown in Figure 6, is developed as discussed in Section 3.1. This display device replaces the mechanical switches and loads – such as light and fan - can be controlled by pressing buttons on the touchscreen. It also shows current time and weather gadgets. It is possible to add more than two loads with the other GPIO pins of the RPi.

The smartphone app is developed according to the discussion in Section 3.2. A screenshot of the first screen of the smartphone app is shown in Figure 7. The app shows the *on* and *off* buttons, and the current on-off status for each load. The app contains a *connect* button and also shows the connection status with the display panel.

5. Conclusion

In this paper, a Wi-Fi network based architecture to control home appliances using a smartphone and also with a touchscreen-based wall display panel IoT device is proposed. Using the smartphone, the user can control appliances from anywhere in the home. A prototype of the proposed system is developed and tested.

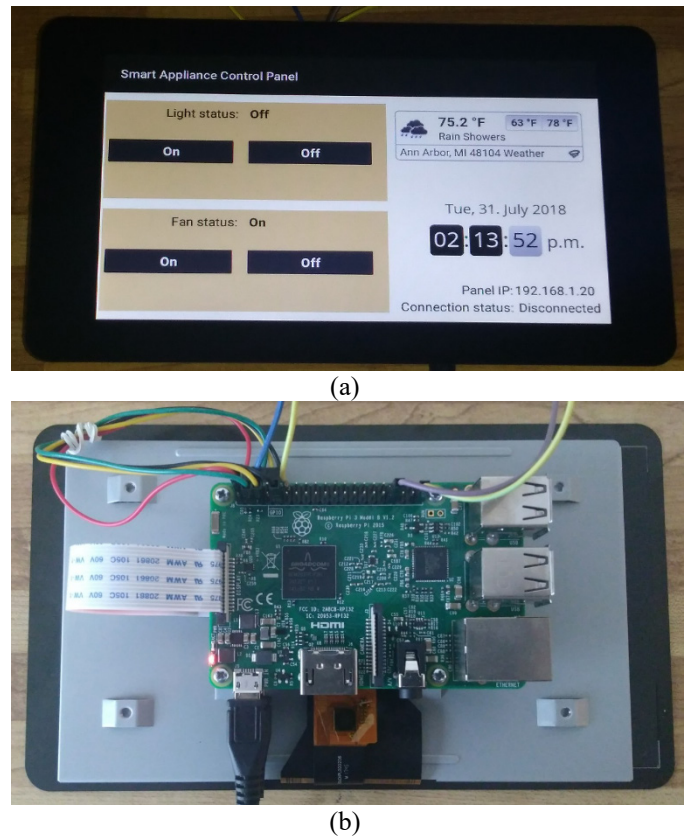


Figure 6. The wall display IoT device: (a) Top view – showing on-off buttons for each load, load and connection status, current time, and weather gadgets. (b) Bottom view –Raspberry Pi board interfaced with LCD and SSR.

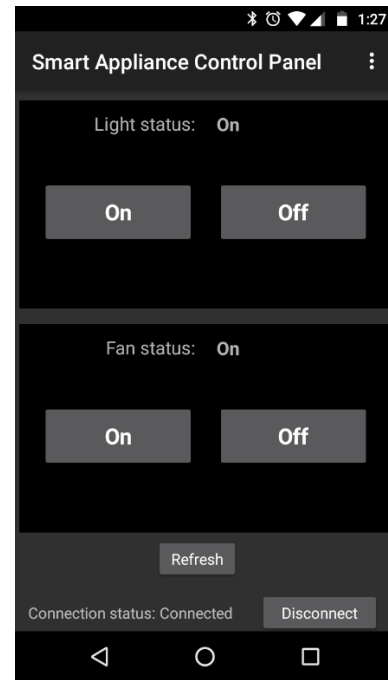


Figure 7. Screenshot of the smartphone app

Future work includes controlling the brightness or speed of the load using PWM, a cloud-based web interface for controlling and monitoring the loads and interfacing temperature and humidity sensors with the display panel.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] The Home Network: Our Neglected Workhorse. [Online], Available: http://www.linksys.com/resources/img/features/ea8500/The_Home_Internet_Network_IDC_Brief.PDF, 2015
- [2] Facts and stats about Wi-Fi. [Online], Available: <http://worldwifeday.com/about-us/facts/>, 2018
- [3] C. Paul, A. Ganesh and C. Sunitha, "An overview of IoT based smart homes," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 43-46.
- [4] A. Zanello, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," in *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22-32, Feb. 2014.
- [5] R. A. Johar, E. Fakieh, R. Allagani and S. M. Qaisar, "A smart home appliances control system based on digital electronics and GSM network," 2018 15th Learning and Technology Conference (L&T), Jeddah, 2018, pp. 52-58.
- [6] Z. Xiaodong and Z. Jie, "Design and implementation of smart home control system based on STM32," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 2018, pp. 3023-3027.
- [7] X. Wen and Y. Wang, "Design of smart home environment monitoring system based on raspberry Pi," 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 2018, pp. 4259-4263.
- [8] H. Singh, V. Pallagani, V. Khandelwal and U. Venkanna, "IoT based smart home automation system using sensor node," 2018 4th International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, 2018, pp. 1-5.
- [9] K. Laubhan, K. Eggenberger, T. Khan and K. Yelamarthi, "Design of a smartphone operated powerstrip," 2017 IEEE International Conference on Electro Information Technology (EIT), Lincoln, NE, 2017, pp. 317-320.
- [10] Raspberry Pi, [Online]. Available: <https://www.raspberrypi.org>, 2018.
- [11] Raspberry Pi LCD - 7" Touchscreen, [Online]. Available: <https://www.sparkfun.com/products/13733>, 2018.
- [12] Solid State Relay, [Online]. Available: <https://www.sparkfun.com/products/13015>, 2018.
- [13] DC Power Supply, [Online]. Available: <https://www.sparkfun.com/products/13831>, 2018.
- [14] Android Things, [Online]. Available: <https://developer.android.com/things/get-started/index.html>, 2018.
- [15] Clock widget, [Online]. Available: <https://www.zeitverschiebung.net/en/clock-widget>, 2018.
- [16] Weather widget, [Online]. Available: <https://www.willyweather.com/widget/create.html>, 2018.
- [17] How do I reserve an IP address on my NETGEAR router? [Online]. Available: <https://kb.netgear.com/25722/How-do-I-reserve-an-IP-address-on-my-NETGEAR-router>, 2018.
- [18] C. Y. Chang, C. H. Kuo, J. C. Chen, and T. C. Wang, "Design and Implementation of an IoT Access Point for Smart Home," *Appl. Sci.* vol. 5, 2015, pp. 1882-1903.
- [19] R. Piyare, "Internet of Things: Ubiquitous Home Control and Monitoring System using Android based Smart Phone," *International Journal of Internet of Things*, vol. 2, no. 1, 2013, pp. 5-11.

Prospects of Wind Energy Injection in the Brazilian National Interconnected System and Impacts Analysis Through a Quasi-Steady Power Flow

Italo Fernandes*, David Melo, Gabriel Santana, Fernando Brito, Allisson Almeida

Electrical Engineering Department, ISL Wyden International College, 65071-380, Brazil

ARTICLE INFO

Article history:

Received: 31 July, 2018

Accepted: 25 September, 2018

Online: 14 November, 2018

Keywords:

Renewable Energy

Wind Energy

Energy Auctions

Quasi-Steady Power Flow

ABSTRACT

There is no doubt that the wind was the renewable source of energy that had the most significant growth during the last five years, and more importantly, the wind power source in Brazil has always been the cheapest and most competitive of all the others, so it is worth saying that wind have never been so well used. The wind energy generation in Brazil is hugely getting stronger, mostly in Northeast region where for the first-time energy auctions trough hydroelectric has been surpassed wind ones. Besides, there are still those who want to benefit from this significant advance, making Brazil the first country to collect royalties from the wind. In this paper a reviewed about Brazilian wind energy scenario and prospects will be done, enumerating the main impacts caused by this kind of power injection in a static analysis. In addition, a Quasi-Steady Power Flow (QSPF) will be simulated to show the impacts of loss, and voltage fluctuation created by the intermittence of wind resource. Numerical evaluation was performed in IEEE-30 bus benchmark system. Computer results, demonstrated the needing of control to make electrical variations smoothly on different periods of the day.

1. Introduction

The idea of generating electricity from wind came precisely with the development of electricity. With the development of coal and oil, in parallel, the wind was left aside as a source of electricity. However, from the petroleum crises on 70s, wind energy was once again used in the generation of electricity. The principle of conversion is very simple: kinetic energy are provided by rotating blades, where the force of wind turns its on and consequently, the machine rotors, multiplied by a gearbox to the generator. The mechanical power extracted from the wind by the wind turbine is given by half the products of the air density (ρ), area swept by the rotor wind turbine (A), wind speed (V) power three and wind turbine power coefficient (C_p), as shown in (1) [1]

$$P = \frac{1}{2} \rho A C_p V^3 \quad (1)$$

This energy is clean and renewable, there is no cost associated with obtaining a raw material. It is considered the cleanest source of energy in the world. It can be obtained in several places, including at sea, by offshore installations. These advantages of wind power have led several countries to establish incentives, regulating and driving financial investments to stimulate wind power generation [2,3].

Wind power is in high growth and countries that do not have land area to install or expand their wind farms are opting for offshore installation. This is not the end of Brazil that has a large area to be explored and with better winds.

Wind energy is gaining its space in the Electric Energy Trading Chamber (CCEE in Portuguese), enabling the purchase and sale of energy throughout the country. CCEE promotes discussions focused on the evolution of the market, always guided by the pillars of isonomy, transparency and reliability. It acts as an institution responsible for offering the viability of purchase and sale of energy throughout the National Interconnected System [4,5]. This system of generation and transmission of electric power, is unique worldwide in its size and characteristics, encompassing the five regions of Brazil, with the purpose of efficiently generating power, giving the ability to choose which plants will generate energy and where this energy goes, this system of support and priority for wind power plants connected to the grid [6].

In this paper is shown a review about the wind energy context in Brazilian power system, enumerating the main impacts caused by its intermittence factor. Only static analysis is performed, and then parameter as harmonics injection, frequency and angular stability are not cover in the context. Besides, a briefly explanation about the national electrical energy corporation and

* Corresponding Author: Italo Fernandes, ISL Wyden International College, Brazil, Email: italo.fernandes@ieee.org

its structure are treated. Finally, a Quasi-Steady Power Flow (QSPF) is simulated to present the main impacts of wind generation on voltage and losses.

This paper is structured as follows: Section 2 presents the general aspects of wind energy in the world scenario, as well as its trends and perspectives for generation. Section 3 explains the organizational structure of the Brazilian electrical sector. Section 4 shows the market aspects for wind energy in the Brazilian scenario. Section 5 handle numerical results obtained through simulations. Finally, section 6 presents the conclusions.

2. General Aspects of Wind Energy

Wind energy, among renewable energy technology, had shown the greatest growth in Brazilian system and around the world. In northeast side of the country the presence of these source is now greater than hydroelectric, which in terms of total generation reach more than 70% of injection [7-9].

2.1. Wind Energy around the World

Wind energy price is cheaper than the price of thermal energy. However, according to the Global Wind Energy Council (GWEC) report, there was a decrease in the amount of new capacity installed in 2017 compared to the previous two years, this is justified by the elimination of subsidies and institutional support in several countries that undermined investments, at a time when the wind industry is in transition to a system based on the rules of the market [7].

In Brazil, installed capacity in 2017 was 2 GW, representing 4% of the world and the accumulated capacity reached 2% of the global total with 12.8 GW. The country surpassed Canada in the world ranking in accumulated installed capacity going to 8th place. However, given its territorial, population and economic dimensions, Brazil has done little to produce wind energy. [10]

On world stage, wind energy is advancing strongly on the high seas. Navigant Research reports that the global wind industry has installed 3.3 GW of offshore capacity by 2017, attracting approximately 17 GW of global total accumulated. In Germany, the industry had its first "no-subsidy" auction this year, with proposals for more than 1 GW of new capacity. In the next five years, the global market for offshore wind energy must install more than 24 GW of new capacity, and by the end of 2022 should reach the accumulated of 40 GW [8].

2.2. Brazilian context on Wind Energy

According to the GWEC, report of 2015 shows that Brazil has the best winds in the world, it has a wind power potential three times higher than the country's electricity needs, and that the capacity factors Brazilians are above the global average. While Brazil goes from 50% and in times of best winds 70%, the other countries average 25% of capacity factor [5].

With the exception of the Amazon region, the potential of the winds is distributed in the national territory, more intense from June to December, coinciding with the months of lower rainfall intensity (less generation of energy in the hydroelectric plants) [11].

Currently, wind energy is the 3rd place in the Brazilian electricity generation matrix. From December 31st, 2016 to December 31st, 2017, the generation of energy by wind power represented about 7% of all electricity generation [6].

2.3. Trends and perspectives for generation in Brazilian scenario

Diversity in an energy matrix guarantees greater security to the system because it is not dependent on a single source. If there is a problem with a specific source, there are others to continue generating. An example of it are the hydroelectric power plants: In some periods the flows of the rivers decrease, thus decreasing the generation of energy, requesting the actions of other sources, such as thermoelectric plants. In a recent case, on the current year, Brazil was over a truckers' strike, which prevented diesel oil from reaching a thermal plant, affecting directly the interconnected system and consequently the energy price. Not being dependent on just a few sources is advantageous for system reliability [6].

Incentive Program for Alternative Energy Sources (PROINFA *in Portuguese*) encourage the development of renewable sources in the energy matrix and paved the way for the fixation of the component industry and wind turbines in the country. At the end of 2009, the 2nd Brazilian Energy Reserve Auction (ERA) was held, which was the first auction to sell energy exclusively from wind power sources, contracting the amount of 1.8GW [12].

In August 2010, the 3rd ERA and the Alternative Source Auction (ASA) were held, where 2GW of wind power was contracted. These auctions no longer worked exclusively with the wind model, but included several renewable sources competing with each other to negotiate their energy in the auction. The price of wind energy was lower than biomass energy price and Small Hydroelectric Power Plants (SHP) [13].

In 2011, the A-6 auction became the second cheapest energy, soon after the hydroelectric plants. About 49 wind farm projects were contracted, with a physical guarantee of 776.6 average MW and a power of 1,386.9 MW, shown itself to be more competitive than biomass and SHPs [13].

In addition to PROINFA and specific auctions mentioned before, the wind power also sells its energy, on a smaller scale, in the Free Market where contractual conditions are freely negotiated bilaterally.

With better winds, institutional incentives and receptiveness to investments from abroad, Brazil has the cheapest wind energy in the world. By 2020, wind power should have a 12% of presence in country's energy matrix [13].

3. Organizational structure of Brazilian energy sector

Brazilian electricity generation and transmission system is predominantly hydrothermal with continental dimensions and huge amounts of hydroelectric power plants. The National Interconnected System (SIN *in Portuguese*) is made up of four subsystems: South, Southeast/Midwest, Northeast and North [6].

The interconnection of the electrical systems, through the transmission grid, facilitates the interchange between subsystems, allows obtaining synergistic gains and explores the different

hydrological regimes of rivers. The integration of generation and transmission resources allows supplying the market with security and economy [14].

CCEE is a non-profit civil association, whose purpose is to enable the commercialization of electric energy in the Brazilian energy market. It brings together companies of generation of public service, independent producers, self-producers, utilities, commercializes, importers and exporters of energy [13].

The contracting of energy can occur through two environments: (1) the Regulated Contracting Environment (ACR *in Portuguese*) that should meet the demand of distribution utilities who can only buy energy through the auctions held by CCEE. In these auctions, generators and sellers compete with each other for provide demand of distribution utilities; and (2) the Free Contracting Environment (ACL *in Portuguese*) where, generators, consumers and sellers negotiate bilaterally buying and selling energy, i.e., they are free to negotiate.

The Short-Term Market accounts the differences between what the agents contracted and what was actually produced or consumed in both environments (ACL and ACR). The Settlement Price of Differences (PLD *in Portuguese*) is the reference price of the short-term market, used to set a price on what was generated and what was consumed by all market participants.

Several factors influence the computational models used in the calculation of PLD: the occurrence of rainfall in the areas where hydroelectric reservoirs are located, which is the amount of water that can be transformed into electric energy, and the behavior of the load that can be influenced by temperatures. PLD is calculated weekly by the CCEE for four submarkets (subsystems) and three load levels (light, medium, and heavy) and is used to assess differences in the short-term market [4].

The Mechanism of Compensation for Wastes and Deficits has application exclusively on Energy Trading Contracts in the Regulated Environment. It promotes energy and power transfer associated with sellers among distribution agents that have leftover energy for distribution agents with energy deficits; it minimizes or eliminates any penalties for insufficient energy ballast [15].

4. Market aspects for wind energy in Brazilian scenario

The northeastern subsystem is directly affected by the hydrological shortage due to geo-climatic conditions, and as a consequence, it is in the long and bitter energetic instability that has been overcome in the last five years. Renewable energies such as wind power were fundamental to avoid the shortage of the Northeast, and the significant relevance of this was the lack of use of reserve energies for this purpose.

According to Steve Sawyer, GWECs General Secretary, wind power is the most competitive option to add more capacity to power grid in many growing markets such as Africa, Asia and Latin America. By 2030, wind power could reach 2,110 GW and account for up to 20% of world energy [16].

The New Energy auctions, handled by CCEE, are executed and presented through two phases, among which, the first one aims to reduce the risks to generators and buyers, where it is constituted by a given qualification question per bid price, considering the capacity of SIN. The second one consists of a continuous phase for the enterprises classified in the first phase,

where the selection criterion is for a lower price, with four different products per source.

It is valid to report that, during the auction of the New Energy A-4 auction, after the one-hour period with a stabilized current price, hydro and wind sources returned to decrease values, dropping more than 30%, where energies such as wind and solar has more than 60% reduction in price [9].

Fiscal incentives such as the installation of a wind farm in the city of Paulino Neves (state of Maranhão) in 2017 have contributed to Brazil's growth in the ranking of wind energy producers. Through an investment of around BRL 1.5 billion, Maranhão is another state that adheres to the global trend of sustainable energy production at the pole of energy produced through the winds. It is worth noting that the country exceeded Canada and ranks eighth in the world ranking of wind energy producers. The consequence of all this investment in wind power production in one more state of the Northeast subsystem may reflect positively on the cost of energy to consumers, mainly because most of the parks are located in the same, being considered as the best location for capitation of winds in the world. Recalling also, that the Northeast Region only escaped rationing last year because of the wind turbines. In the critical phase of hydroelectric plants, wind farms supplied 11% of Brazil and 60% only in the Northeast.

One resource, increasingly adopted in several countries, is the storage of the cheapest energy available in the system, possibly part of that generated by intermittent sources, to be used in the absence of this generation. Thus, part of the intermittent generation or other sources would have their consumption postponed to times of greater need, and could then receive higher prices, which would compensate for losses incurred in filling and emptying the energy reservoirs, be they hydraulic, batteries or other, less usual. Another advantage of the accumulation is that it can be made with energy generated by the plants based on renewable sources [17,18].

5. Numerical Results

Simulations were performed in a MATLAB® R2015a (8.5.0.197613) environment on an Intel (R) Core i7-4770S 3.1 GHz, 4-cores and 8GB of RAM with the aid of statistics and Machine learning toolbox. To obtain the results, the IEEE-30 System was used, containing 30 bus, 6 generators, 41 lines and a total load of 283.4 MW and 126.2 MVar of active and reactive power respectively. The proposed methodology involved the simulation of a QSPF, applying 100 intervals of time in a load curve modeled by a normal type PDF. Three buses of the System were chosen for inclusion of wind farms. Generators configurations or turbines are not a problem concerned in this simulation, as they are modeled just as injection active power. The criterion for selection of buses was those with higher density of load that did not contain generation. It is clear that in a real situation the installation of wind farms depends on the energy availability among several other factors that could influence the investment. A penetration percentage of 10%, divided proportionally by the amount of load of each of the buses chosen, was considered. The generation of energy through the wind farms was modeled from a Weibull PDF. For the proposed methodology,

wind farms do not have reactive control, being modeled only as active power injections.

Trough numerical results presented in all figures, it could be said that all electrical variables in the system will be affect by the intermittent feature of wind source. The most evident technical impact can be noticed on voltages profile, especially in the WF bus and its neighbor's bars.

The generation profile for the 100 periods executed in the quasi-steady power flow are recorded in Figure 1.

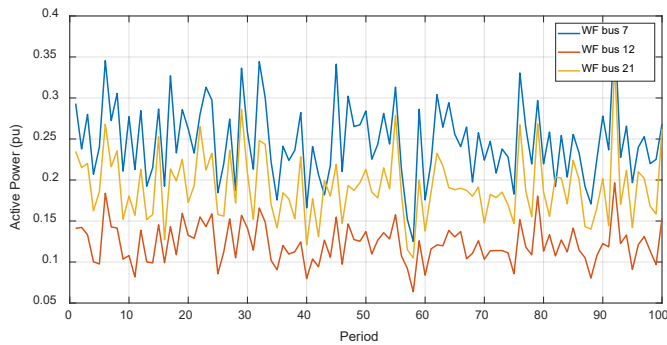


Figure 1: WF profile of active power injection

Voltage profile for WF buses and neighbor buses are shown in Figure 2-4.

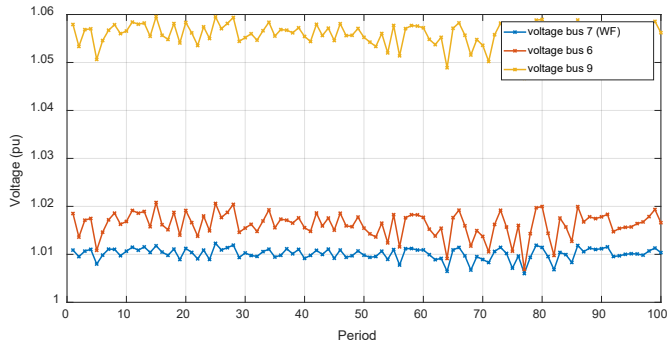


Figure 2: Voltage profile for buses near WF located in bus 7

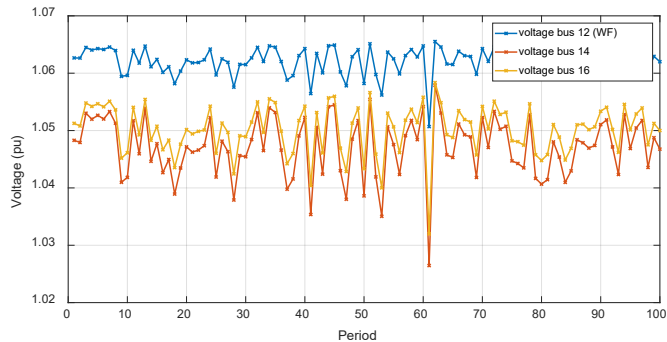


Figure 3: Voltage profile for buses near WF located in bus 12

System losses according to all 100 periods are shown in Figure 5.

For example, in Figure 3, voltage in bus 14 goes from about 1.02 to 1.06 in a short period of the QSPF. In other situation these

variations can be smooth, but simulations show that the bigger the wind power injection, bigger will be these variations, especially when the system presents weak profiles of voltage. This variability makes the operator of the system to improve its control technics, for example, using FACTS devices to make it smoother voltages variations [19].

Another factor that can be mentioned is the systems losses. This analysis is general for all systems that involves generation near the load (distributed generation). The explanation can be done in a very simple way: As an amount of energy is being generated close to a charge, those loads can consume this power,

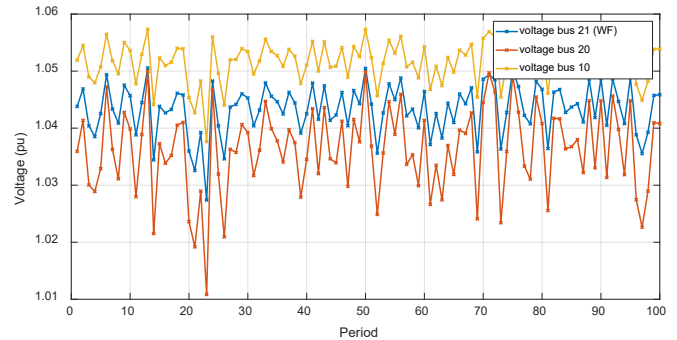


Figure 4: Voltage profile for buses near WF located in bus 21

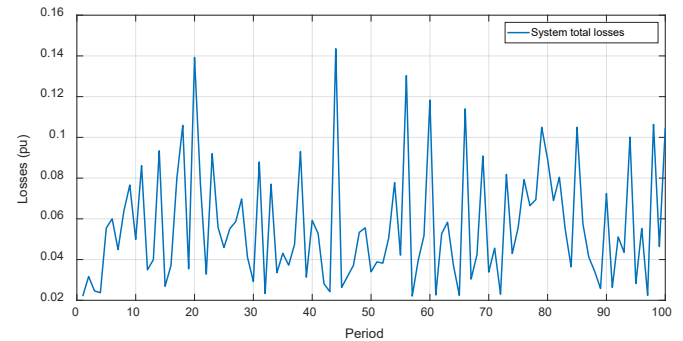


Figure 5: WF profile of active power injection

and the amount of energy that was coming from a distant source it will no longer be needed. By this way, a smaller current will pass through transmission lines, decreasing RI^2 and XI^2 losses. The situation get clearer looking at Figure 5. When the system losses are greater (period 20, 43, 56, 60 and 67) it means that wind power injections are lower, and for the lower losses (period 1, 2, 58 and 71), wind are almost down [20].

6. Conclusion

The future of electric energy is that: A clean and renewable generation system with distributed generation across the entire grid. It is worth mention that the growth and application of those technologies are exponentially noticed. Brazilian investors already realize that wind is potentially the future base of power system and those investments are taking huge proportions. National Interconnected system have to be ready to absorb greater amount of intermittent power. Commercially wind energy have to overcome the reliability problem trough precise forecasting resources to supply the variability of the source. Numerical results on section 5 shows the needing of control to handle the

changeability of voltage. Besides, it is clear from simulations that the fact of including distributed generation in a power system, improve overall grid losses.

Acknowledgment

We thank ISL Wyden International College for technical support, which provided a good environment to develop this research.

References

- [1] CHENG, M.; ZHU, Y. The state of the art of wind energy conversion systems and technologies: A review. *Energy Conversion and Management*, v. 88, p. 332–347, Dezembro 2014.
- [2] Impacts and Challenges in operating Planning”, in Portuguese, Bsc. Thesis, Federal University of Rio de Janeiro, 2014
- [3] CAILLÉ, A. et al. 2007 Survey of Energy Resources. United Kingdom: World Energy Council, 2007. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.481.5707&rep=rep1&type=pdf>>. Acesso em: 3 ago. 2015.
- [4] CCEE Documents “Marketing Rules”, Version 2018.1, available online on: www.ccee.org.br/
- [5] M.H. Nascimento, “Impacts of wind farms in electrical market”, in Portuguese, Msc. Thesis, UFI, 2005
- [6] J. E. Ferreira, “Impacts of wind energy generation in market prices”, in Portuguese, Msc. Thesis, IS CET Business School, 2016
- [7] Global Wind Energy Council, “Global Wind Report”, Annual Market Update 2017, Available online on: www.gwec.net
- [8] GWEC. Global Wind Energy Outlook 2014. [s.l.] GWEC, 2014. Disponível em: <http://www.gwec.net/wp-content/uploads/2014/10/GWEO2014_WEB.pdf>. Acesso em: 9 set. 2015.
- [9] ABEEólica-Brazilian association of wind energy, “Annual generation bulletin 2017”, available online on: <http://abeeolica.org.br/wp-content/uploads/2018/04/Boletim-Anual-de-Geracao-2017.pdf>
- [10] A. P. Correa, “Wind energy in Market”, in Portuguese, Msc. Thesis, Thesis, IS CET Business School, 2014
- [11] ONS Documents, “MONTHLY WIND GENERATION BULLETIN”, August 2017, Available online on: ons.org.br/
- [12] ABEEÓLICA. Boletim de Dados - Julho 2015. São Paulo: ABEEólica, jul. 2015. Disponível em: <<http://www.portalabeeolica.org.br/index.php/dados.html>>.
- [13] CCEE. Consolidated results of auctions - 03/2015. [s.l: s.n.]. Available on: <http://www.ccee.org.br/ccee/documentos/CCEE_347805>.
- [14] M.S. Simas, “Wind Energy and sustainable development in Brazil”, in Portuguese, Msc. Thesis, USP, 2012
- [15] J. Pessanha, V. L. O. Castellani, V. A. Andrade “Short-Term Wind Power Forecasting Based On Quantile Regression”, *Brazil Windpower*, 2017
- [16] S. O. Nunes 1, E. O. Teles 2, E. A. Torres 3, “Risk Analysis in the Wind Energy Market: A Review”, *Brazil Windpower*, 2017
- [17] Kim, K., Park, H., & Kim, H. (2017). Real options analysis for renewable energy investment decisions in developing countries. *Renewable and Sustainable Energy Reviews*, 75, 918–926.
- [18] Li, F., Liu, Z. C., Jia, X. X., Zeng, M., & Li, N. (2012). Investment Risk Assessment Model for Wind Power Projects Based on Full Life-Cycle Theory. *East China Electric Power*, 40(4), 0531–0535.
- [19] AYODELE, T. R.; OGUNJUYIGBE, A. S. O. Mitigation of wind power intermittency: Storage technology approach. *Renewable and Sustainable Energy Reviews*, v. 44, p. 447–456, abr. 2015.
- [20] BLACK, M.; STRBAC, G. Value of Bulk Energy Storage for Managing Wind Power Fluctuations. *IEEE Transactions on Energy Conversion*, v. 22, n. 1, p. 197–205, mar. 2007.

Exploring the use of Manual Liquid Based Cytology, Cell Block with Immunomarkers p16/ki67, VIA and HPV DNA Testing as a Strategy for Cervical Cancer Screening in LMIC

Nandini Nandish Manoli^{*1,2}, Devananda Devegowda², Ashoka Varshini¹, Pushkal Sinduvadi Ramesh², Sherin Susheel Mathew³, Nandish Siddappa Manoli⁴

¹Department of Pathology, JSS Medical College, JSS Academy of Higher Education & Research, 570015, India.

²Centre of Excellence in Molecular Biology & Regenerative Medicine, Department of Biochemistry, JSS Medical College, JSS Academy of Higher Education & Research, 570015, India.

³Department of Pathology, Dr Somervell Memorial Church of South India Medical College, 695504, India

⁴Department of Obstetrics and Gynecology, JSS Medical College, JSS Academy of Higher Education & Research, 570015, India

ARTICLE INFO

Article history:

Received: 05 September, 2018

Revised: 30 September, 2018

Accepted: 14 November, 2018

Keywords:

Manual Liquid Based Cytology

Cervix

Immunomarkers

ABSTRACT

Cervical cancer is the 4th most common cancer in women in low to middle income group countries (LMIC). Various methods for screening cervical cancer are practiced, such as the Conventional Pap Smear (CPS), Liquid Based Cytology with its ancillary techniques like Cell Block with immunocytochemistry. VIA is another method which is being advocated as a primary screening tool. Molecular diagnostics such as use of HPV DNA testing has been at the forefront of the screening programs. In the present study, we have utilized all the above methods by using cost effective in-house procedures to explore their possible utility in the clinical settings. We found them useful with need for more work and training of personnel for better diagnosis of cervical cancer.

1. Introduction

Even though Cervical Cancer (CC) is the leading cause of death in women in the modern world, the incidence varies in developed and developing countries. India accounts for quarter of cervical cancer incidence in the world with 20.2 per 100,000 new cases of CC diagnosed and 11.1 per 100,000 deaths annually [1].

Screening by various methods can bring down the incidence of cervical cancer which has been successfully implemented in developed countries. But, the lack of infrastructure, national policies, poverty, lack of financial resources, inadequately trained health care providers, improper implementation of the screening programs with lack of education of women especially in the rural population has prevented the control of cervical cancer incidence in countries like India. This renders a huge load leading to a major global impact on health care of women [2, 3, 4].

Screening tests for cervical cancer include:

- Conventional exfoliative cervico-vaginal cytology i.e. the cervical (Pap) smear.
- Manual liquid-based cytology

- Fluid sampling technics with automated thin layer preparation (liquid-based cytology)
- Cell block technique with immunomarker study
- HPV DNA testing
- Polar probe
- Laser induced fluorescence
- Visual inspection of cervix after applying Lugol's iodine (VILI) or acetic acid (VIA)
- Speculoscopy
- Cervicography

1.1. Exfoliative cytology (conventional Pap smear)

Conventional Pap smear has been the standard screening method for cervical cancer screening from the past several years. It includes sampling of material from the junction between the ecto- and endo-cervix using an Ayre's spatula. The material obtained is spread on to the clean glass slide which is stained by routine pap stain and studied by the Cytopathologist. It has limitations due to many errors either due to sampling (5-10%), interpretation and obscuring factors like blood which hamper the

*Corresponding Author: Nandini Nandish Manoli, Department of Pathology, JSS Medical College, JSSAHER, +919448978276 & nandinimanoli65@gmail.com

accurate diagnosis. Only 20% of the sample taken gets spread on to the slide which hampers the sensitivity [5,6].

1.2. Manual liquid-based cytology

Liquid based cytology is a technique wherein cells are arranged in a single or monolayer on a clean glass slide. It is a method which helps to remove obscuring factors seen in conventional method. There are two automated methods followed in developed countries: the SurePath and ThinPrep. These methods improve the sensitivity and specificity and also help in using the remaining material for ancillary techniques like cell block with IHC and HPV DNA testing. But these automated methods have their limitations as they are expensive to be used in LMIC.

To overcome these limitations, we have utilized an in-house cost-effective method of Manual Liquid Based Cytology (MLBC). We used simple machines like centrifuge and the sample collected in liquid fixative which was processed in a polymer solution prepared in the laboratory [7].

1.3. HPV Testing

Epidemiological studies and mechanistic evidence has led to the conclusion that 70% of cervical cancer cases are attributed to HPV-16 & HPV-18, the high-risk subtypes of HPV. Cytology based cervical screening has led to the reduction in the incidence and mortality rates of cervical cancer and evidence suggests that inclusion of HPV testing could further refine the screening programs. Also, HPV testing has enormous potential to be used as a cost-effective primary screening module, to identify women with greater risk of disease progression and as a test of cure of disease [3,8, 9].

1.4. Cell Block Technique

Cell block technique is a method wherein the residual material in a sample can be processed to form a tissue block. The process employed can be varied from alcohol, formalin, agarose or thromboplastin to form a cell pellet which is processed like a tissue. The advantages are, multiple sections can be taken and can be used for immunocytochemistry (ICC) to improve the diagnosis of cervical lesions. Cell block can be used as histopathology tissue for controls in cytopathology laboratories. Only limitation is the time required for processing a cell block and the extra cost for the preparation [10].

1.5. Visual inspection tests

Visual inspection tests with 3-5% acetic acid (VIA) and/or Lugol's iodine (VILI) appear to be a satisfactory alternative screening approach to cytology. These tests have been used since the 1990s, mainly in poor resource settings. They are simple, cost-effective with relative ease of use and may be performed by different healthcare workers (physicians, nurse, midwives and technicians). Moreover, this approach does not require high technology or infrastructure and has been shown to reduce mortality in developing countries [8,11,12].

2. Material and Methods

This was a prospective study carried out from January 2017 to June 2018 in a tertiary care hospital of South India. The study was conducted after obtaining Institutional Ethical clearance from the committee. A total of 100 subjects within the age group of 20-70 years attending the OBG department with gynecological

complaints were recruited for the study. Subjects with a history of abnormal Pap tests were included in the study and subjects who did not give consent to participate were excluded from the study. Samples were collected from the recruited subjects only after obtaining signed informed consent.

2.1. Sample collection and processing

Liquid based cervical cytology samples were collected using an endo-cervical cell collection device (Cervex-Brush®-Rovers medical devices). All the 100 samples were subjected to conventional Pap smear testing and manual liquid-based cytology analysis which was subjected to ancillary tests i.e HPV testing and cell block processing. Due to the inadequacy of the sample for DNA extraction, only 68 cases were subjected to HPV DNA detection by Polymerase chain reaction. A total of 25 cases of cell block with HPV correlation were subjected to immunocytochemical analysis with p16 and ki67 markers.

2.2. DNA extraction & Polymerase Chain Reaction for HPV detection

Samples are collected using cyto brush was transferred to a sterile capped container and immediately transported to the hospital's molecular laboratory. Samples were centrifuged at 10,000 rpm for 5min to pellet down the cells and the cells were subjected to DNA extraction. DNA from the samples was extracted using HiPura™ Multi-sample DNA purification kit (HiMedia, India) according to the manufacturer's instructions. PCR was performed using consensus MY09/MY11 primers that targets a 450bp region in L1 gene of the HPV. PCR was performed on Eppendorf's Mastercycler gradient as described in [13]. The positive DNA samples from the PCR was then subjected to type specific PCR with specific primers. Previously described primers described by [14] were utilized for the PCR experiments.

2.3. p16 and ki67 immunocytochemistry with cell blocks

Immunocytochemistry was performed on the formalin-fixed and paraffin-embedded cell block sections by DAB chromogen method. Mouse monoclonal anti-p16 antibody was used. Scoring was done as following: Negative (no staining or <3 positively stained cells), 1+(3-10 positively stained cells), 2+(>10 positively stained cells). Along with the cell number, staining intensity was also taken into consideration. For ki67 only nuclear staining with less than or more than 10 cells were taken as weak or strong positivity.

2.4. Visual inspection with acetic acid (VIA)

We conducted a study with a primary health centre for correlation of VIA with CPS on 100 cases at healthcare camps. The study took into consideration whether the VIA test was adequate or inadequate with visualization of the squamocolumnar junction or not. Pap smears with Ayre's spatula were taken by trained health workers at these camps. 5% acetoacetic acid was applied to the cervix which was visualized by a Gynecologist or the health worker

3. Results

A total of 68 samples were subjected to HPV detection and genotyping by PCR in our study. Out of 68 cases, 12 were positive for HPV (8.16%) and out of these 12 HPV positive cases 7 were positive for high risk HPV type 16. None of the samples were HPV

18 positive. Since we did not have positive controls for other high risk HPV subtypes, we could not test the status of these samples for other HPV subtypes.

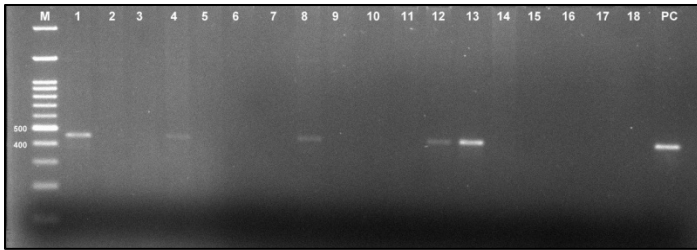


Figure 1: Representative gel image showing amplified HPV gene product of 450bp (M= Marker, PC= Positive control)

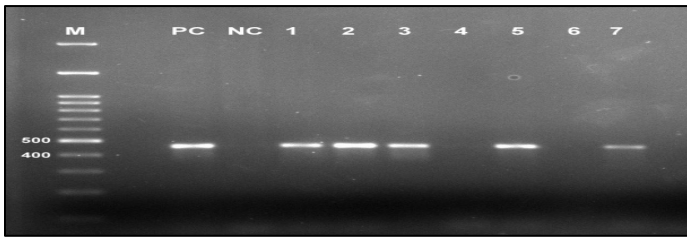


Figure 2: Representative gel image showing amplified HPV 16 gene product of 468bp (M= Marker, PC= Positive control, NC= Negative control)

Twenty-five cases of cell block with HPV correlation showed 6 cases with HPV positivity. All 25 cases were subjected to p16^{ink4a} IHC, of which 18 cases of chronic cervicitis were negative, two cases of koilocytic atypia were negative, one case of LSIL was weak positive, two cases of HSIL were strong positive and two cases of SCC were also strongly positive.

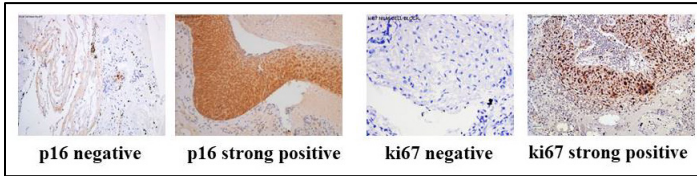


Figure 3: Representative images of IHC stained for p16 and ki67 immunomarkers In cell blocks of NILM(negative) and squamous cell carcinoma(positive)

Table 1: Correlation of histopathological characteristics with HPV DNA positivity

	HPV Positive(n =11)		HPV Negative	Total
	Negative HR-HPV(16/18)	Positive HR-HPV(16)		
Chronic Cervicitis	1	0	6	07
LGSIL	3	3	4	10
HGSIL	0	1	6	07
SCC	0	3	2	05
Total	4	7	18	29

Table 2: Correlation of cell block staining with HPV DNA positivity.

P16 CELL BLOCK	HPV		
	ON	Positive	Negative
	Positive	5	0
Negative	1	19	

Statistical analysis

Correlation of HPV DNA testing with histopathology Correlation of Cell block and HPV DNA testing

Sensitivity: 45%

Sensitivity: 83.3%

Specificity: 85 %

Specificity: 100%

PPV: 91%

PPV : 100%

NPV: 24%

NPV: 95%

Table 3: Correlation of histopathology with cell block and conventional Pap smear

CELL BLOCK	CHRONIC CERVICITIS(10)	LSIL(14)	HSIL(12)	SCC(7)
NILM	10(100%)	0	0	0
KOILOCYTIC ATYPIA	0	2(14%)	2(17%)	0
LSIL	0	12(86%)	1(8%)	0
HSIL	0	0	9(75%)	1(14%)
SCC	0	0	0	6(86%)
CPS				
NILM	10(100%)			3(42%)
KOILOCYTIC ATYPIA		2(13%)		
LSIL		8(53%)	10(84%)	1(16%)
HSIL		4(34%)	1(8%)	
SCC			1(8%)	3(42%)

Table 4: Correlation of VIA with conventional Pap smear

Diagnosis	CPS	VIA
NILM	47	35
LSIL	04	1
HSIL	07	1
SCC	12	2
Endo-cervical carcinoma in-situ	05	-
ASCUS	06	1
High grade carcinoma	03	1
Unsatisfactory	16	4
total	100	47

Of the 63 cases, 20 cases are without any correlation wherein 10 cases were VIA negative. Thus, the percentage of missed cases on VIA was 16%.

4. Discussion

Cervical cancer screening and detection has improved from the days of conventional Pap smear screening to molecular tests in developed countries where government national health care policies have initiated screening programs leading to the reduction in the incidence. In developing country like ours, the screening programs have not got its wings due to various reasons. Attempts to make a low-cost method of early detection lead us to start an MLBC technique which reduces obscuring factors and spreads the cells in a monolayer for a clearer viewing of the cells. Similar technique has been used by many investigators as it has additional advantage of ancillary studies like testing for HPV DNA, preparation of cell block for immune-marker studies [7,15,16].

HPV as a primary screening test has been advocated and being followed in European countries it has also found its feasibility in LMIC countries because of availability of many commercially available kits. They have become a milestone in the more effective

screening of cervical cancer and prolonging the screening interval for patients.

We have standardized our own in-house HPV DNA testing methodology with a turnaround time of one day. The World Health Organization (WHO) recommends targeting HPV screening to women who are 30 years of age and older because of their higher risk of CC, and that priority should be given to screening women aged 30-49 years (WHO screening recommendation update 2014) [3].

4.1. Triage of HPV-Positive Women

HPV-based screening has a low positive predictive value for CC because it does not directly test for cancer, but for HPV infection instead. At the present time, three test methods can potentially be used as triage test: visual methods (VIA/VILLI); cytology; and molecular testing. To date, there is no clear evidence to determine which strategy should be prioritized. Therefore, the choice of test essentially depends on the available resource [12].

4.2. Triage with cytology

Cytology is the most widely recommended test to triage HPV-positive women, where quality-assured cytology is available. HPV-positive women with a cytology diagnosis of ASCUS or worse are referred for colposcopy, and the rest are advised to have repeat HPV testing after 1 year. Cytology performs better in a triaging scenario, since the prevalence of disease is high in the sample and cytologists have a limited number of specimens to evaluate. The current recommendations by the American Society for Colposcopy and Cervical Pathology (ASCCP) are direct referral to colposcopy for HPV 16/18 positive women and repeat testing after 1 year for women positive for other HPV types [17].

4.3. Triaging with Biomarkers

LBC with immunocytochemistry and cell block sections with immunohistochemistry result in enhanced specimen quality, and accurate diagnosis, and diminished false negative cases. LBC has potential as a screening tool for cancer and precancerous lesions in several tissues other than gynecologic organs.

Cell block tissues made from remnants and residual LBC samples, aspirates, and fluid samples may also have applications for practice in the field of cytopathology. We have used a cost effective method by using MLBC in our set up to be used for HPV and cell block with p16 and ki67 as immune markers [18]. These markers are known to highlight the HSIL and squamous cell carcinoma cases of cervix. p16 can also diagnose LSIL cases even though it also gives positivity for endometrial cell tubal metaplasia and squamous metaplasia which will not be given positive by ki67 [15]. Thus, the use of p16ink4a and ki67 on cell blocks will enhance high grade/malignant lesions of the cervix from the non-neoplastic conditions and thus improve diagnostic accuracy as we found in our study.

4.4. Triaging with VIA

VIA which is a good approach for screening and treating in resource poor settings where cytology and HPV testing cannot be done is useful when the skill and knowledge about the technique is good. We found that it has its limitations as found by many workers [17].

5. Conclusion

Cervical cancer screening in low to middle income countries still needs to be refined in terms of affordability and accessibility. Current strategies for cervical cancer screening are not being implemented to its full potential mainly due to the lack of training, high cost and need of well set-up screening centers. There are various methods for screening cervical cancer in LMIC, which are being done in a small scale either in the form of research studies or by NGOs with whom we joined hands and did a study on VIA. In our study we explored the usage of multi-algorithm screening strategy for the screening of cervical cancer in a tertiary care hospital.

There is a need for a uniform policy of screening of women at the primary health care center level with increasing the awareness of the different methods among the public. Also, there is a need for well-trained health workers and Cytopathologists to diagnose and maintain follow up about cervical cancer with a cancer registry.

References

- [1] J.Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012", *INT J CANCER*, 136(5), E359-E386, 2015. <https://doi.org/10.1002/ijc.29210>
- [2] K.S Tewari, A. Agarwal, A. Pathak, A. Ramesh, B. Parikh, M. Singhal, G. Saini, P.V Sushma, N. Huilgol, S. Gundeti and S. Gupta, "Meeting report : First Indian national conference on cervical cancer management-expert recommendations and identification of barriers to implementation", *GYNECOLONCOLRESRACT*, 5(5), 2018. <https://doi.org/10.1186/s40661-018-0061-5>
- [3] R. Catarino, P. Petignat, G. Dongui, and P. Vassilakos, "Cervical cancer screening in developing countries at a crossroad: Emerging technologies and policy choices", *WORLD J CLIN ONCOL*, 6(6), 281, 2015. <https://doi.org/10.5306/wjco.v6.i6.281>
- [4] S. Bobdey, J. Sathwara, A. Jain and G. Balasubramaniam, "Burden of cervical cancer and role of screening in India", *INDIAN J MED PAEDIATR ONCOL*, 37(4), 278, 2016. <http://doi.org/10.4103/0971-5851.195751>
- [5] J. Sherris, "Cervical cancer in the developing world", *WESTERN J MED*, 175(4), 231-233, 2001. <https://doi.org/10.1136/ewjm.175.4.231>
- [6] R.A Kerkar, and Y. V Kulkarni, "Screening for cervical cancer: an overview", *J OBSTET GYNECOL INDIA*, 56(2), 115-122, 2006.
- [7] N.M Nandini, S.M Nandish, P. Pallavi, S.K Akshatha, A.P Chandrashekhar, S. Anjali, and M. Dhar, "Manual liquid based cytology in primary screening for cervical cancer-a cost effective preposition for scarce resource settings", *ASIAN PAC J CANCER P*, 13(8), 3645-3651, 2012. <https://doi.org/10.7314/APJCP.2012.13.8.3645>
- [8] N. Wentzensen, M. Schiffman, T. Palmer, M. Arbyn, "Triage of HPV positive women in cervical cancer screening", *J CLIN VIROL*, 76, 49-55, 2016. <https://doi.org/10.1016/j.jcv.2015.11.015>
- [9] T.C Wright, M.H Stoler, C.M Behrens, A. Sharma, G. Zhang, T.L Wright, "Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test", *GYNECOLONCOL*, 136(2), 189-197, 2015. <https://doi.org/10.1016/j.ygyno.2014.11.076>
- [10] L. Skoog, and E. Tani, "Immunocytochemistry: an indispensable technique in routine cytology", *CYTOPATHOLOGY*, 22(4), 215-229, 2011. <https://doi.org/10.1111/j.1365-2303.2011.00887.x>
- [11] L. Denny, M. Quinn and R. Sankaranarayanan, "Screening for cervical cancer in developing countries", *VACCINE*, 24, S71-S77, 2006. <https://doi.org/10.1016/j.vaccine.2006.05.121>
- [12] U.R Poli, P.D Bidinger, and S. Gowrishankar, "Visual inspection with acetic acid (via) screening program: 7 years experience in early detection of cervical cancer and pre-cancers in rural South India", *INDIAN J COMMUNITY MED*, 40(3), 203, 2015. <https://doi.org/10.4103/0970-0218.158873>
- [13] P.S Ramesh, D. Devegowda, P.R Naik, P. Doddamani, and S.M Nataraj, "Evaluating the Feasibility of Nested PCR as a Screening Tool to Detect HPV Infection in Saliva of Oral Squamous Cell Carcinoma Subjects", 12(7), 2018. <https://doi.org/10.7860/JCDR/2018/34880.11806>
- [14] S. K Bandhary, V. Shetty, M. Saldanha, P. Gatti, D. Devegowda, S.R Pushkal and A.K. Shetty, "Detection of Human Papilloma Virus and Risk Factors among Patients with Head and Neck Squamous Cell Carcinoma Attending a

- Tertiary Referral Centre in South India”, *ASIAN PAC J CANCER P*, 19(5), 1325, 2018. <http://doi.org/10.22034/apjcp.2018.19.5.1325>
- [15] I. Akpolat, D.A. Smith, I. Ramzy, M. Chirala and D.R Mody, “The utility of p16INK4a and Ki-67 staining on cell blocks prepared from residual thin-layer cervicovaginal material”, *CANCER CYTOPATHO*,102(3),142-149, 2004. <https://doi.org/10.1002/cncr.20258>
- [16] R.K. Sherwani, T. Khan, K. Akhtar, A. Zeba, F.A Siddiqui, K. Rahman and N. Afsan, “Conventional Pap smear and liquid based cytology for cervical cancer screening-A comparative study”, *J CYTOL*, 24(4), 167, 2007. <https://doi.org/10.4103/0970-9371.41888>
- [17] P. Basu, F. Meheus, Y. Chami, R. Hariprasad, F. Zhao and R. Sankaranarayanan, “Management algorithms for cervical cancer screening and precancer treatment for resource-limited settings”, *INT J GYNECOL OBSTET*,138, 26-32, 2017. <http://doi.org/10.1002/ijgo.12183>
- [18] H.Sakamoto, M. Takenaka, K. Ushimaru and T. Tanaka, “Use of Liquid-Based Cytology (LBC) and cell blocks from cell remnants for cytologic, immunohistochemical, and immunocytochemical diagnosis of malignancy”, *OJPATHOLOGY*,2(3),58,2012. <https://dx.doi.org/10.4236/ojpathology.2012.23012>.

Guidance Law Based on Line-of-Sight Rate Information Considering Uncertain Modeled Dynamics

Saori Nakagawa*, Takeshi Yamasaki, Hiroyuki Takano, Isao Yamaguchi

National Defense Academy of Japan, Department of Aerospace Engineering, 239-8686, Japan

ARTICLE INFO

Article history:

Received: 28 August, 2018

Accepted: 28 October, 2018

Online: 15 November, 2018

Keywords:

Aerospace Engineering

Autonomous system

Adaptive control and tuning

ABSTRACT

Proportional navigation (PN) is a widely-used guidance law for missile-target engagement. The goal of the missile intercept problem is to reduce the closest distance between the missile and target by diminishing the line-of-sight rate (LOS rate). In general, PN guidance law necessitates information of the LOS rate and missile velocity. The closing velocity (relative approaching speed to the target) instead of the missile velocity is an additional option for effective guidance. However, there are cases where a sensing device for measuring target motions that can be mounted on a missile is limited. In this paper, we propose a novel guidance law on the basis of proportional navigation (PN) using only line-of-sight (LOS) rate information. In this paper, an uncertainty and disturbance estimator (UDE) is applied to estimate such target motions including velocity change or unpredictable movement etc. The UDE works also for compensating uncertain modeled dynamics such as a missile's bearing uncertainty and velocity changes. The proposed guidance law is referred to as uncertainty and disturbance-compensated intercept guidance. Numerical simulations with some engagement scenarios are presented taking account of the velocity changes of the missile to demonstrate the potential of the proposed guidance law.

1 Introduction

Proportional navigation (PN) is a widely-used guidance law for terminal interception of target-missile engagement [1], and the guidance law attracts attention since the first half of the 20th century. Researches have been conducted not only for missile guidance, but also in the fields such as the guidance of vehicles [2], the formation flight of aircraft [3], small UAVs autonomous path-following [4], furthermore the problem of rendezvous of satellites and terminal guidance control to a small moon landing [5, 6].

Proportional navigation (PN) is mainly classified into two; (1) true proportional navigation (TPN) using closing velocity information and (2) the pure proportional navigation (PPN) using the missile's velocity information [7]. Both laws use information on the line-of-sight (LOS) rate (rate of change in the direction of a line connecting a missile and a target) and the velocity of the missile (if it is obtained, using the closing velocity to the target). The PN is widely used as a guidance law for the intercept in two-vehicle engagement

because the implementation only with the LOS rate information is quite simple [1]. In the case of a stationary or non-maneuvering target, without measurement error, dynamics lag, nor lateral acceleration limit on the navigation system of the missile, the PN can completely intercept the target with zero miss distance. For such a reason, researches on missile guidance have been conducted based on the PN or alternative representation such as the zero-effort-miss (ZEM) guidance. For instance, the augmented proportional navigation (APN) compensates a target lateral constant acceleration by adding a modified acceleration term to the PN law [8]. The compensated PN [9] (or velocity change-compensated PN [10]) guides missiles to keep the LOS rate constant at zero by correcting the LOS rate change produced by the missile's axial acceleration.

In [11], optimal theory on intercept guidance provides that the effective navigation constant of the PN set to three is an optimal against the non-maneuvering target in the sense of the least squared integral of the lateral accelerations. The preceding guidance system does not take dynamics-lag into account while an opti-

*Saori Nakagawa, 1-10-20 Hashirimizu, Yokosuka, Kanagawa 239-8686, Japan, TEL:+81-468-41-3810, email: saori.sari.nkgw@gmail.com

mal guidance for a first-order time-lag system is given in [8]. The guidance law in [8] that minimizes the squared integral of the commanded lateral acceleration of the missile uses the target lateral acceleration information. In addition, in this optimal guidance law (referred to as OG in this paper), a variable gain is used instead of the constant value, and the variable gain can be expressed as a function of the time-to-go (the time remaining till interception). In this regard, the sub-optimal guidance (SOG) law was proposed using PN with phase lead compensation by Baba et al. [12]. The optimal or suboptimal guidance laws described so far are effective under the assumption where the closing velocity information is available. Sensing device for target information may depends on the missile's operations. There are cases where the missile cannot install an equipment measuring the closing velocity. Therefore, in this study, we focus on the case where only LOS rate can be acquired as for target maneuvers-related information other than that of the missile itself.

Since the closing velocity cannot be obtained while only LOS rate is used, the usual fixed gain approach in PN may not exhibit sufficient guidance performance. Here, we focus on the LOS rate dynamics in order to enhance or keep the guidance performance under restricted conditions where only information as for the target is to be the LOS rate. We proposed a new guidance law that drives the LOS rate to be zero, estimates and compensates an uncertain group including unmodeled dynamics and external [13]; In [13], the authors did not consider time-lag in system response, so we present results for a time-lag system [14]. Furthermore, this study demonstrates the robustness of the proposed guidance affected by the axial acceleration by the missile thrust and the deceleration due to aerodynamic drag.

The rest of this paper is organized as follows: Section 2 details the setting of the guidance problem with notation assignments for the missile-target engagement, followed by the governing equations. Next, a guidance law, which is based on an assembled simple LOS rate dynamics model, is developed. In Section 3, numerical simulation results are shown to demonstrate the potential of the proposed guidance law. Section 4 summarizes this study.

2 Guidance Theory

2.1 Engagement problems and Governing equations

This section introduces the notations, assumptions, governing equations for the guidance law, and defines the engagement problem.

Figure 1 shows a missile-target geometry and their related notations for deriving the guidance law for the missile-target engagement scenario. Here after, the guided vehicle is referred to as a missile. λ denotes the LOS angle, and R is the LOS range. V , γ and a for each vehicle represents the velocity, flight path angle, and

lateral acceleration, respectively. The subscripts for these variables distinguish the missile from the target; with "m" for the missile, and "t" for the target.

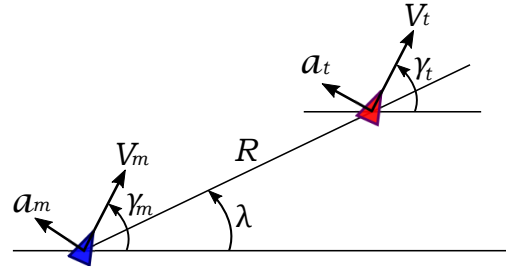


Figure 1: Engagement Geometry

Assumptions are made for the derivation of the proposed guidance law as

- (i) The missile and target are point masses, moving in a plane.
- (ii) The target moves with a constant speed.
- (iii) The LOS rate is available without error.
- (iv) The response lag to the commanded lateral acceleration for the missile can be represented by a first-order-lag system.

Under these assumptions, the governing equations on the missile-target engagement can be summarized as

$$\dot{R} = V_t \cos(\gamma_t - \lambda) - V_m \cos(\gamma_m - \lambda) \quad (1)$$

$$\dot{\lambda} = \frac{1}{R} \{V_t \sin(\gamma_t - \lambda) - V_m \sin(\gamma_m - \lambda)\} \quad (2)$$

$$\dot{\gamma}_t = a_t / V_t \quad (3)$$

$$\dot{\gamma}_m = a_m / V_m \quad (4)$$

$$\dot{a}_m = -\frac{1}{\tau_1} a_m + \frac{1}{\tau_1} a_c \quad (5)$$

$$\dot{V}_m = \frac{T - D}{m} \quad (6)$$

$$\dot{m} = \frac{-T}{g I_{sp}} \quad (7)$$

where a_c is a commanded value of missile's lateral acceleration, and system response a_m is approximated by the output from a first-order-lag system with the time constant τ_1 . Also T , D , I_{sp} and g denote the missile's thrust, drag force, and specific impulse, and the gravitational acceleration, respectively. For additional variables for simulations, the lift denoted by L and the drag-related equations are shown as;

$$L = \frac{1}{2} \rho V_m^2 S C_L \quad (8)$$

$$D = \frac{1}{2} \rho V_m^2 S C_D \quad (9)$$

$$C_D = C_{D_0} + \kappa C_L^2 \quad (10)$$

where ρ is the air density, S is the cross sectional area, C_L, C_D, C_{D_0} , and κ denote the lift coefficient, the drag coefficient, the zero-lift drag coefficient, and the induced drag-related parameter, respectively.

If the LOS angle λ is maintained constant, the missile can intercept the target [1, 8]. That is, the problem for deriving the guidance law on missile-target engagement is to calculate the commanded lateral acceleration for missile, a_c , which will bring the LOS angular velocity (the LOS rate) $\dot{\lambda}$ to zero. Thus, the objective of this study is to find a_c nullifying the LOS rate without using information of the closing velocity.

2.2 Construction of a simplified dynamics model

In designing a guidance system, the construction of a dynamics model plays an important role. An accurate model is indispensable to build a high precision guidance. On the other hand, a designed guidance system with such an accurate model tends to become a complicated and high-order system, and to demand multiple sensors and/or estimators. For this reason, we try to use a simplified guidance model as much as possible aiming at simplifying the design process for a guidance system that uses only information of the LOS rate.

In order to design a guidance system, this study focuses on a mathematical model representing the LOS rate dynamics. By differentiating Eq. (2) and substituting Eqs. (1), (3) and (4), the LOS rate dynamics model is derived in the following differential equation form;

$$\ddot{\lambda} = \frac{1}{R} \{-2\dot{R}\dot{\lambda} - a_m + h_0\} \quad (11)$$

$$h_0 = a_m \{1 - \cos(\gamma_m - \lambda)\} + a_t \cos(\gamma_t - \lambda) + \dot{V}_t \sin(\gamma_t - \lambda) - \dot{V}_m \sin(\gamma_m - \lambda) \quad (12)$$

The term h_0 in the preceding Eq. (12) includes "model uncertainty" arising from mathematical simplifications, and "disturbance" caused mainly from uncertain target accelerations or maneuvers. Hereafter, the term h_0 including such model uncertainty and external disturbance is called an uncertainty and disturbance term (UDT).

In [15], Eq. (11) is used as a system equation for the development of the guidance logic. However, Eq. (11) includes not only the available LOS rate but also the LOS range and the closing velocity ($V_c = -\dot{R}$) that cannot be measured or estimated. Instead of Eq. (11), in order to take account of the restriction where only the LOS rate is available, this study uses the following equation;

$$\ddot{\lambda} = -\frac{a_m}{R_0} + h \quad (13)$$

where R_0 is a given parameter in advance, for example the reachable flight range of the missile or the initial value of the LOS range. The UDT (h_0) of Eq. (11) and

additional uncertain terms caused from this simplification are combined as

$$h = \left(\frac{1}{R_0} - \frac{1}{R} \right) a_m + \frac{1}{R} \{-2\dot{R}\dot{\lambda} + h_0\} \quad (14)$$

The estimation of the UDT (h) is explained in Section 2.4.

In this study, the command lateral acceleration a_c is considered as the input to the guidance system while taking account of the missile dynamics which is approximated as a first-order-lag system. In the guidance system, the LOS rate $\dot{\lambda}$ and the lateral acceleration of the missile a_m are treated as state variables. Thus, the lateral acceleration a_m is treated as an intermediate variable for a backstepping approach. The purpose of this guidance system can be interpreted as "driving the first state variable $\dot{\lambda}$ to zero with the input of a_c ". A method for calculating the input a_c driving the first state variable to zero will be described in the next section.

2.3 Derivation of uncertainty and disturbance-compensated guidance

In this section, we explain the derivation of uncertainty and disturbance-compensated guidance based on the governing Eqs. (5) and (13). The guidance law should drive the LOS rate $\dot{\lambda}$ to zero with the lateral commanded acceleration a_c as described in the previous section. However, the first state variable cannot be directly controlled by the lateral acceleration commanded value a_c . Therefore, this study uses the backstepping method to control the LOS rate $\dot{\lambda}$ via the intermediate state a_m . A candidate of the Lyapunov function is defined as

$$V = \frac{1}{2} \sigma^2 \quad (15)$$

where, for simplicity, the LOS rate $\dot{\lambda}$ is defined as σ . When Eq. (15) is differentiated with time, it becomes

$$\dot{V} = \sigma \dot{\sigma} = \sigma \left(-\frac{a_m}{R_0} + h \right) \quad (16)$$

The estimated value of the UDT h is defined as \hat{h} . Moreover, the desired missile lateral acceleration a_{md} is set as

$$a_{md} = R_0 (k\sigma + \hat{h}) \quad (17)$$

On the other hand, addition and subtraction of a_{md}/R_0 to Eq. (13) and making $(a_{md} - a_m)$ term lead to

$$\dot{\sigma} = -\frac{a_{md}}{R_0} + h + \frac{a_{md} - a_m}{R_0} \quad (18)$$

Substituting Eq. (17) for the first term of Eq. (18) provides

$$\dot{\sigma} = -k\sigma + (h - \hat{h}) + \frac{a_{md} - a_m}{R_0} \quad (19)$$

Substituting Eq. (19) for the first equality in Eq. (16), it is found that

$$\dot{V} = -k\sigma^2 + \sigma (h - \hat{h}) + \frac{\sigma}{R_0} (a_{md} - a_m) \quad (20)$$

If the lateral acceleration a_m converges to the desired lateral acceleration a_{md} , and when the UDT(h) can be estimated without error, that is, $\hat{h} \rightarrow h$, the Eq. (20) is expressed as $\dot{V} \approx -k\sigma^2 \leq 0$ (the equality holds when $\sigma = 0$). Consequently, if k is positive, V can be called as the Lyapunov function. For this reason, the LOS rate σ converges to zero as the lapse of time.

In order to make a_m close to a_{md} , a new candidate Lyapunov function is defined as

$$V_1 = \frac{1}{2}\sigma^2 + \frac{1}{2}(a_{md} - a_m)^2 \quad (21)$$

Differentiating Eq. (21) with time leads to

$$\dot{V}_1 = \sigma\dot{\sigma} + (a_{md} - a_m)(\dot{a}_{md} - \dot{a}_m) \quad (22)$$

From the observation of Eqs. (5), (17) and (22), commanded lateral acceleration can be selected as

$$a_c = a_{md} + \frac{\tau_1\sigma}{R_0} \quad (23)$$

If it can be approximated as $\dot{a}_{md} \doteq 0$ and the estimated value become $\hat{h} \rightarrow h$, Eq. (22) becomes

$$\dot{V}_1 = -k\sigma^2 - (a_{md} - a_m)^2/\tau_1 \quad (24)$$

that shows $\dot{V}_1 \leq 0$.

Equations (17) and (23) represent the guidance law. Excepting the estimated value \hat{h} of the UDT in Eqs. (17) and (23), the guidance law includes terms proportional to the LOS rate without the closing velocity: Eqs. (17) and (23) essentially includes the PN guidance law.

For comparison purpose to the proposed guidance law, the conventional guidance laws are illustrated next, which clarifies what kind of extra information is required for each guidance law.

The PN guidance law that uses missile velocity instead of the closing velocity (what is called 'pure PN (PPN)') is represented as [16]:

$$a_{mPPN} = NV_m\sigma \quad (25)$$

where N denote the navigation constant. The lateral acceleration is assumed to be generated perpendicular to the velocity vector in the PPN guidance law. In this research, the closing velocity information cannot be obtained. However, if this is obtained, it is effective to use the closing velocity V_c instead of the missile velocity V_m of the Eq. (25). Thus, the command lateral acceleration of PN guidance law is

$$a_{mTPN} = N'V_c\sigma \quad (26)$$

where N' is the effective navigation constant [16]. According to the optimal control theory as described in the introduction [11], if the missile has perfect dynamics (it has no response lag), and the target has constant-speed with rectilinear motion, when the effective navigation constant is 3 ($N' = 3$), the input of command lateral acceleration is minimizing the square integral of acceleration.

The augmented PN (APN) guidance law is a well-known optimal guidance law against the constant

maneuvering target. The target's constant lateral acceleration-related term is added to the TPN guidance law as [8]

$$a_{mAPN} = N'V_c\sigma + \frac{N'a_t}{2} \quad (27)$$

where a_T indicates the target maneuver (lateral acceleration which is assumed to be normal to the LOS) [15].

Furthermore, the optimal guidance law (OG) is derived based on the optimal control theory for a liner system in case where the response of the system can be approximated by a first-order-lag;

$$a_{mOG} = \frac{N'}{t_{go}^2}[y + \dot{y}t_{go} + 0.5a_t t_{go}^2 - a_c\tau_1^2(e^{-x} + x - 1)] \quad (28)$$

where

$$x = \frac{t_{go}}{\tau_1} \quad (29)$$

Also, y indicates a component perpendicular to the initial LOS. y is subtraction of the missile position component from the target one with respect to the initial LOS. If y is sufficiently small, the relationship of trigonometric ratio provides

$$\lambda \doteq \sin \lambda = \frac{y}{R} \quad (30)$$

The following relationship can be obtained by differentiating both sides of Eq. (30).

$$V_c\dot{\lambda} \doteq \frac{y + \dot{y}t_{go}}{t_{go}^2} \quad (31)$$

The numerator in the right hand side in Eq. (31) is called ZEM (the miss distance with no guidance force applied). The variable gain in Eq. (28) is derived as

$$N' = \frac{6x^2(e^{-x} - 1 + x)}{2x^3 + 3 + 6x - 6x^2 - 12xe^{-x} - 3e^{-2x}} \quad (32)$$

Using Eq. (31), Eq. (28) is expressed as

$$a_{mOG} = N'V_c\sigma + \frac{N'a_t}{2} - N'a_c x^{-2}(e^{-x} + x - 1) \quad (33)$$

In the simulation of this paper, these commanded acceleration (Eqs. (25) ~ (27) and Eq. (33)) are used to compare with the proposed guidance law.

k of the proposed guidance law (Eq. (17)) is a positive constant to be tuned. Some design policy can be helpful. In this study, the gain is determined as compared with PN guidance law of Eq. (25) as

$$k = NV_{m0}/R_0 \quad (34)$$

where R_0 is a pre-specified value; it could be the initial LOS range or the possible flight range. V_{m0} and N denote the assumed average velocity of the missile and

the navigation constant, respectively. In this proposed guidance law, N is the main tuning parameter.

2.4 Estimation of the uncertainty and disturbance term

We describe the estimation method [15, 17, 18] for UDT h in this section.

As for the UDT estimation, assembled term for uncertainty and disturbance in a focused system dynamics are estimated based on the system dynamics. For example, a discrete-type method, so-called time delay control (TDC) [19, 20] was proposed for an uncertainty and disturbance compensator. The TDC technique, however, arose a potential problem; since the time derivative in TDC is approximated with numerical differentiations, instability of the system may occur. In [17], a frequency-domain-based approach that can be treated in continuous system were proposed by Zhong. In this method, the stability of the disturbance is guaranteed. Hence, this study follows the Zhong's idea for the UDT estimator with additional extensions for the measurement limit.

In [18] and [15], the authors also proposed missile guidance law using an uncertainty and disturbance compensator under the condition where the measurements of the LOS range and the range rate as well as the LOS rate can be obtained. The guidance system proposed here imposed further limitations, that is, the proposed system can use information of the LOS rate only.

The UDT estimation methodology using the LOS rate without the LOS range nor the closing velocity is developed.

The first step begins with solving Eq. (7) for h

$$h = \dot{\sigma} + \frac{a_m}{R_0} \quad (35)$$

The right hand side of the Eq. (35) contains the differential term of the LOS rate which is a measured variable. Using the derivative of the measured variable (or estimated value) may cause numerical problem since the differentiation is sensitive to measurement noise and to the characteristics of state estimator. Thus, using the derivative of the LOS rate should be avoided.

Similar to [18] and [15], assuming that a UDT h is the input to some strictly proper virtual filter $G(s)$, and setting its output denoted by \hat{h} as the estimate of h , lead to the following representation in the Laplace domain form as

$$\hat{H}(s) = G(s)H(s) \quad (36)$$

In fact, it is impossible to extract $h(t)$ as a signal. In this paper, we merely assume that it is the input to the virtual filter for the sake of developing the UDT estimator. Although various kinds of filters can be the candidates for the virtual filter, the following first-order-lag filter is applied in order to facilitate the design.

$$G(s) = \frac{1}{\tau s + 1} \quad (37)$$

where τ indicates the time constant and that is a design parameter to be tuned. The time constant τ characterizes the convergence rate. By using the first-order-lag filter shown in Eq. (37), Eq. (36) is expressed in the time domain form as follows:

$$\tau \dot{\hat{h}}(t) + \hat{h}(t) = h(t) \quad (38)$$

where we assume $\hat{h}(0) = 0$ without loss of generality. Substitute the left hand side of Eq. (38) for the left hand side of Eq. (35) and rearrange it to derive

$$\begin{aligned} \tau \dot{\hat{h}} + \hat{h} &= \dot{\sigma} + \frac{a_m}{R_0} \\ \tau \dot{\hat{h}} &= \dot{\sigma} - \hat{h} + \frac{a_m}{R_0} \end{aligned} \quad (39)$$

After integration of Eq. (39) with intervals $[0, t]$ and solving for \hat{h} , the following estimator can be obtained.

$$\hat{h} = (\sigma - \sigma(0) + \omega) / \tau \quad (40)$$

$$\dot{\omega} = -\hat{h} + \frac{a_m}{R_0}, \omega(0) = 0, \hat{h}(0) = 0 \quad (41)$$

The estimator, denoted by Eqs. (40) and (41) is consisted only of the measurable LOS rate, the lateral acceleration of missile and the design parameter. Therefore, if the LOS rate is measurable, the UDT estimator is designed using one design parameter τ .

The convergence rate of the UDT can be enhanced by decreasing the design parameter τ . However, in practically, it has minimum limitation to attenuate the influence of noise. Since Zhong [17] explains the convergence property of the estimate, it is omitted in this paper.

2.5 Stability of the Uncertainty and Disturbance Term

In this section, we explain about the convergence of the uncertainty and disturbance estimator. It is assumed that the UDT (h) is continuous (the time differentiation of h is bounded). Estimated error is defined as

$$\tilde{h} \equiv h - \hat{h} \quad (42)$$

Considering the following non-negative-definite function.

$$V_h = \frac{\tau}{2} \tilde{h}^2 \quad (43)$$

and performing the time derivative of Eq. (43) yields

$$\dot{V}_h = \tau \tilde{h} \dot{\tilde{h}} \quad (44)$$

Substituting Eq. (13) into Eq. (39) provides

$$\dot{\hat{h}} = \frac{1}{\tau} \tilde{h} \quad (45)$$

Then, the differentiation with time of Eq. (42) derives

$$\dot{\tilde{h}} = \dot{h} - \dot{\hat{h}} \quad (46)$$

Substituting Eq. (46) for Eq. (45) leads to

$$\dot{\tilde{h}} = \dot{h} - \frac{1}{\tau} \tilde{h} \quad (47)$$

Noting Eq.(47), Eq. (44) can be rewritten as

$$\dot{V}_h = \tilde{h}(\tau\dot{h} - \tilde{h}) \quad (48)$$

where \dot{h} is assumed to be bounded, considering $|\dot{h}| \leq C$ (C : the positive constant), $\dot{V}_h < 0$ is guaranteed when $|\tilde{h}| > \tau|\dot{h}|$ and the estimation error with elapse of time ($t \rightarrow \infty$) is ultimately bounded within the range of

$$|\tilde{h}| \leq \tau|\dot{h}| \leq \tau C \quad (49)$$

It can be seen from Eq. (49) that the estimation error can be reduced by decreasing τ . Furthermore, when the UDT is constant, that is $\dot{h} = 0$, the uncertainty and disturbance term h is estimated without error ($\hat{h} \rightarrow h$).

2.6 Discussion of uncertainty and disturbance-compensated intercept guidance with LOS rate measurements

From the UDT estimator of Eqs. (40) ~ (41) and the guidance command of Eq. (35), the proposed UDT is summarized as follows:

$$\begin{aligned} \dot{\omega} &= -\hat{h} + \frac{a_m}{R_0}, \omega(0) = 0, \hat{h}(0) = 0 \\ \dot{\hat{h}} &= (\sigma - \sigma(0) + \omega)/\tau \\ a_m &= R_0(k\sigma + \hat{h}) \end{aligned}$$

When the lateral acceleration command value is not limited and there is no time-lag in the guidance system, the lateral acceleration in Eq. (41) equals to the desired lateral acceleration of Eq. (17) as described in the preceding equation. In this case, Eq. (17) is substituted into Eq. (41) with $a_m = a_{md}$ then one obtain

$$\dot{\omega} = k\sigma, \omega(0) = 0 \quad (50)$$

As shown above, the UDT estimator proposed by the authors is almost equivalent to the one composed only proportional-integral filter of the LOS rate, and the design parameters k , R_0 and τ . That is, in comparison with the existing guidance laws demanding extra information such as the closing velocity, the LOS range, the time-to-go and the lateral acceleration of target, the proposed guidance law using only the LOS rate is useful from the view point of implementation simplicity.

3 Simulation

Guidance simulations have been performed to confirm the effectiveness of the proposed guidance law. When the missile velocity is sufficiently larger than the target velocity, there is no significant difference between the proposed guidance law and PN guidance law. Therefore, this paper presents an example case; the target velocity is larger than the missile velocity where the

performance difference appears noticeably by using the proposed guidance law. In the simulations, the equations of motion Eqs. (1) ~ (7) are used. The setting values of initial states and design parameters are summarized in the Table 1.

The limitation load of the missile is set to forty times of the gravitational acceleration (G), that is, $40G$. When the commanded lateral acceleration exceeds the limit, the limited value is used as a command in the simulation. The target acceleration set as $4G$ to be constant. For comparison purposes, the PPN guidance law using the missile velocity in Eq. (25); the TPN guidance law using the closing velocity in Eq. (26); the APN guidance law adding the extra term which considered the maneuvering target in Eq. (27); and the OG of using t_{go} information in Eq. (33) are compared by the proposed guidance law in the simulations.

Table 1: Initial kinematics and design parameters

Parameters	Values
V_m	1,500 [m/s]
x_m	0 [m]
y_m	0 [m]
γ_m	20 [deg]
$\max\{a_m\}$	$40G$ [m/s^2]
m	100 [kg]
$\min\{m\}$	60 [kg]
S	0.01267 [m^2]
I_{sp}	300 [s]
G	9.806 [m/s^2]
V_t	3,500 [m/s]
x_t	55,000 [m]
y_t	9,000 [m]
γ_t	-175 [deg]
a_t	$4G$ [m/s^2]
k	NV_{m0}/R_0 [1/s]
N	3 [-]
N' (TPN, APN)	3 [-]
τ	0.3 [s]
τ_1	0.5 [s]
R_0	$R(0)$ [m]

The simulations with such comparative guidance laws are made because the simulation setting of guidance laws with the effective navigation constant of $N' = 3$ in TPN and APN provide optimal solution in the sense that the square integral of lateral acceleration becomes minimum when there is no system response delay. The proposed guidance law assumes that only LOS rate information is used, whereas the TPN, APN and OG guidance laws use the closing velocity that cannot be used in this assumption. Hence, simulation results PPN using the missile velocity are also

shown as for another comparison guidance law with the same measurement condition. When simulations are performed for TPN, APN and OG guidance law, information on the closing velocity and the target lateral acceleration can be used without error.

Table 2: Resulting miss distance is shown. For comparison, we also show the results when V_m is constant, that is, when the affect of the axial acceleration by the thrust or of the deceleration due to the aerodynamic drag is not considered.

guidance method	miss distance [m]	
	$V_m = \text{const.}$	V_m changes
Proposed	$< \delta$	$< \delta$
PPN	565.76	813.19
TPN	0.18	8.43
APN	0.0028	0.02
OG	$< \delta$	0.0096

$$\delta := 0.001$$

The flight trajectories of the missile and the target are shown in Fig.2. The red solid line drawn from the upper right to the lower left is the trajectory of the target, the red broken line denotes the missile trajectory with the PPN guidance law, the green broken dotted line is made with the TPN guidance law, the magenta dotted line is made with the APN guidance law, the red broken line is made with the OG law, and the blue solid line is made with the proposed guidance law. The resulting miss distance value using each guidance law is shown in Table 2. For comparison purpose, the results of the case where the missile velocity is constant is also shown. In this case, when the miss distance is 0.001[m] or less, it is expressed as " $< \delta$ ".

From enlarged view of the lower right of Fig. 2 (the equivalent aspect ratio), it is observed that the trajectories with the APN, OG and proposed guidance laws have similar trajectories. Simultaneously, we can see that the APN, OG and the proposed guidance law approach to the target in a shape close to a straight line, but the PPN and TPN guidance law draws a curve towards the target. It is considered that this causes large miss distance values as shown in Table 2. Additionally, for the case where the missile velocity changes with time, the miss distance values of the PPN, TPN, APN and OG guidance law become large, whereas the proposed guidance law keeps in the same level. Note that the trajectory with the proposed guidance law is almost similar to the trajectories with the APN guidance law and OG law using the closing velocity and the lateral acceleration, and the miss distance value is smaller than the other guidance laws.

The curvature of the missile trajectory in Fig. 2 becomes clear by observing the histories of the lateral acceleration applied to the missile. Figure 3 shows the time history of the missile's lateral acceleration (so-called latex) when using the each guidance law as well

as the time history of the LOS rate in Fig. 4. In Fig. 3, the value of lateral accelerations of the APN guidance and OG laws converge to almost zero with exception at the time of intercept. In the proposed guidance law, since the target lateral acceleration is considered as a part of the UDT, it turned out that the missile lateral acceleration tends to converge to a value close to the target lateral acceleration at 4G. In addition, it can be seen from Fig. 3 that the PPN and TPN guidance laws have reached the limited value of 40G before the intercept. The PPN guidance law is considered to generate a large miss distance value as shown in Table 2 because it reached this limit earlier than that of the TPN guidance law. It can also be observed from Fig. 4, since the LOS rate with the proposed guidance law converges to a value close to zero at the impact time, a small miss distance value shown in Table 2 was obtained. Consequently, under the condition shown in Table 1, that is, when the target velocity is larger than the missile velocity, the proposed guidance law is considered to exhibit desirable performance.

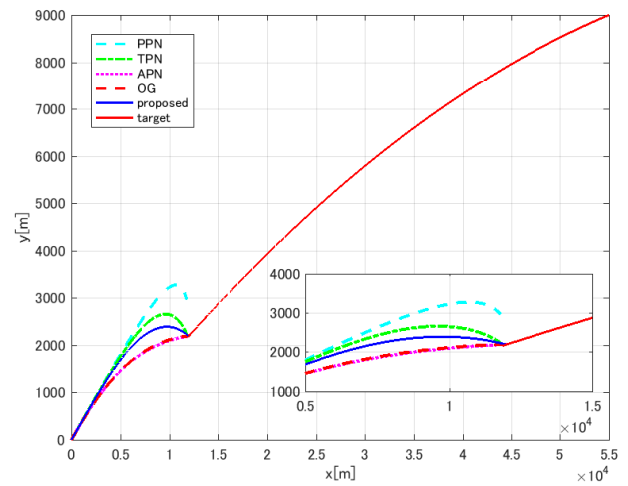


Figure 2: Flight trajectories of the missile and the target. The lower right is an enlarged view.

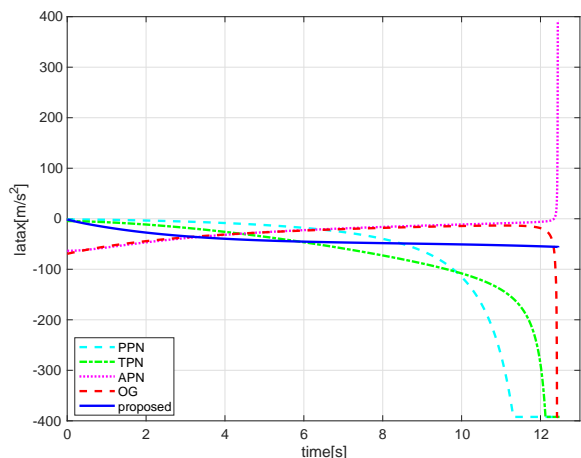


Figure 3: The time history of the missile's lateral acceleration in each guidance law is shown. The limitation load factor of the missile is set 40G this is forty times of the gravitational acceleration.

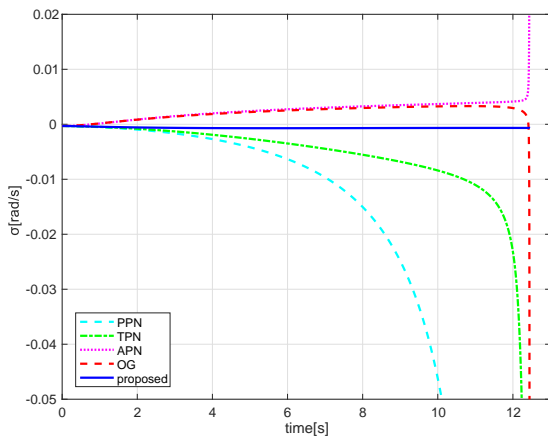


Figure 4: LOS rate vs time is shown. It is desirable for the LOS rate to be zero.

Figure 5 and 6 shows the time histories of the missile velocity and the missile mass. In this simulation, since the minimum value of the missile mass is limited to 60[kg], the missile intercepted the target before reaching that value.

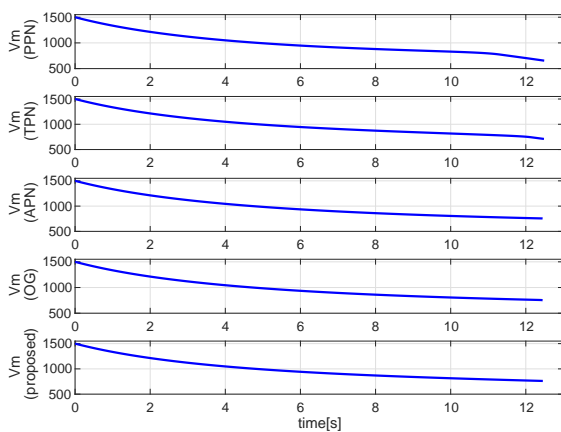


Figure 5: The missile velocity of each guidance considering axial acceleration by the thrust force, and deceleration due to aerodynamic drag vs time.

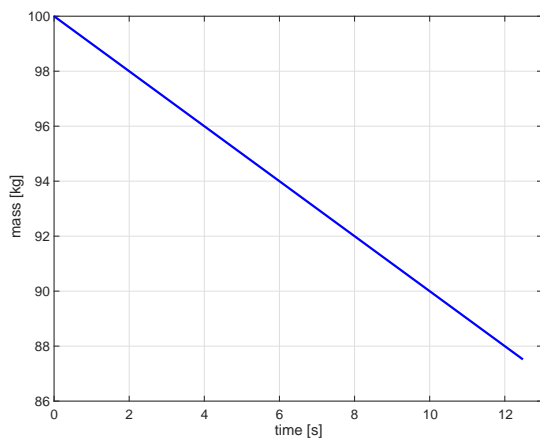


Figure 6: Mass of missile decreasing with time.

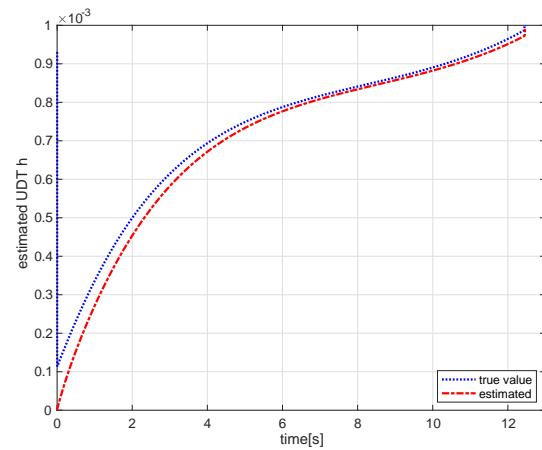


Figure 7: The time history of the estimated UDT h and the true value.

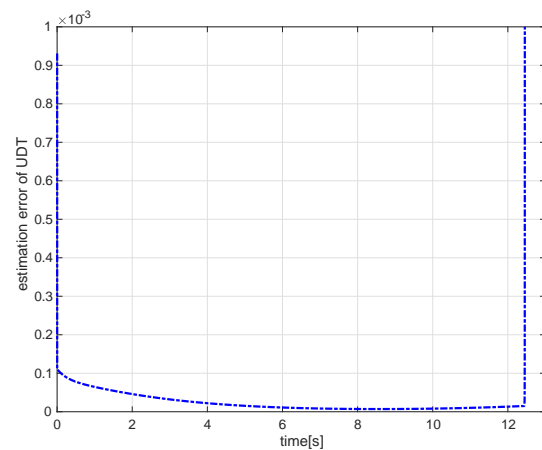


Figure 8: UDT error vs time. This figure shows $(true) - (estimate)$ values in Fig.7.

Finally, to demonstrate the performance of the UDT estimator, the time history of the UDT h which is a true value and that of estimated value \hat{h} are shown in Fig. 7, in addition, the time history of the estimation error is also shown in Fig. 8. From these figures, it can be seen that the estimator using the time constant of $\tau = 0.3[s]$ can be estimated so that the estimated value approaches the true value about 2 to 6 seconds. By reducing the value of the time constant τ , it can be expected to increase the convergence rate to the true value, but the response of the lateral acceleration tends to fluctuate. Therefore, the time constant is too small is not desirable. The reason that the estimation error increases at the intercept time in Figs. 7 and 8 is that numerical divergence occurs for the real system where the distance R in Eq. (14) converges to zero while the acceleration converges to the target acceleration that are none zero value. Therefore, the proposed guidance law should be deactivated or hebetated at the end of the interception to obtain better performance. This is the future work for this study since the LOS range is not measured. Regardless of the difficulties, the result-

ing miss distance as shown in Table 1 demonstrates the excellent performance of the proposed guidance law.

4 Conclusions

A new guidance law for missile-target engagement using information of LOS rate only is developed. From the simulation study, the proposed guidance law exhibits excellent performance almost equal to the OG law where the OG law uses additional information of the LOS range, the closing velocity and the lateral acceleration of the target, whereas the proposed guidance law uses only LOS rate information even if uncertain modeled dynamics such as velocity change and directional error is included. Hence, even with a guidance system that cannot measure the closing velocity or maneuvering target, we can anticipate that almost the same performance as that of conventional excellent guidance with using the closing velocity or target maneuver can be obtained by using the proposed guidance law.

For future works, we will apply the proposed guidance against various targets or initial settings, study a blind sight determination method (or deactivation technique in the proximity of the target), and investigate the influence of the time-lag constant changes and observation noise to the design parameters on guidance performance.

References

- [1] R.E. Machol, W.P. Tanner, Jr, and S.N. Alexander, *System Engineering Handbook*, Chap.19 Guidance, R.E. Hill, 1965.
- [2] Y. Yavin, R.De. Villiers, Proportional navigation and the game of two cars: the case of a pursuer with variable speed, *Computers & Mathematics with Applications*, Vol.18, No.1-3, pp.69-76, 1986.
- [3] M.J. Tahk, C.S. Park, and C.K. Ryoo, Line-of-Sight Guidance Laws for Formation Flight, *Journal of Guidance, Control, and Dynamics*, Vol.28, No.4, pp.708-716, 2005.
- [4] Y. Sato, T. Yamasaki, et al., Trajectory Guidance and Control for a Small UAV, *International Journal of Aeronautical and Space Sciences*, Vol.7, No.2, pp.137-144, 2006.
- [5] M. Matsuda, H. Takano, T. Yamasaki and I. Yamaguchi, A Note on the Optimal Spacecraft Guidance, APISAT, 7.4.1.pdf, 2012.
- [6] P.J. Yuan and S.C. Hsu, Rendezvous Guidance with Proportional Navigation, *Journal of Guidance, Control, and Dynamics*, Vol.17, No.2C pp.409-411, 1994.
- [7] C.D. Yang, C.C. Yang, A Unified Approach to Proportional Navigation, *Transactions on aerospace and electronic system*, IEEE, Vol.33, No.2, pp.557-567, April 1997.
- [8] P. Zarchan, *Tactical and Strategic Missile Guidance*, Sixth Edition, AIAA, ch.8, pp.163-185, 2007.
- [9] P. Zipfel, *Modeling and Simulation of Aerospace Vehicle Dynamics*, Second Edition, AIAA, 2007.
- [10] T. Kuroda, and F. Imado, Advanced Missile Guidance System against a Very High Speed Maneuvering Target, *Proceedings of AIAA Guidance, Navigation and Control Conference*, pp.3445, 1989.
- [11] A.E. Bryson, Jr., and Y.C. Ho, *Applied Optimal Control, - Optimization, Estimation, and Control-*, Hemisphere Publishing Corporation, New York, pp.154-155, 1975.
- [12] Y. Baba, K. Inoue, and R.M. HOWE, Suboptimal guidance with line-of- sight rate only measurements, *Proceedings of AIAA Guidance, Navigation and Control Conference*, AIAA 88-4066-CP, Minneapolis, pp.122-129, 1988.
- [13] S. Nakagawa, T. Yamasaki, H. Takano, and I. Yamaguchi, Disturbance- Compensated Intercept Guidance Using Line-of-Sight Rate Information, 11th Asian Control Conference, IEEE, pp.388-393, December 2017.
- [14] S. Nakagawa, T. Yamasaki, H. Takano, and I. Yamaguchi, Comparison study on Disturbance-Compensated Intercept Guidance Using Line-of-Sight Rate Information against Optimal Guidance, *Multi-Symposium on Control System*, 2018.
- [15] T. Yamasaki, S.N. Balakrishnan, and H. Takano, Sliding mode-based intercept guidance with uncertainty and disturbance compensation, *Journal of Franklin Institute*, Vol. 352(11), pp. 5145-5172, 2015.
- [16] C.D. Yang, and C.C. Yang, A Unified Approach to Proportional Navigation, IEEE, *Transactions on aerospace and electronic system*, Vol.33, No.2, pp.557-567, April 1997.
- [17] Q.C. Zhong, Comments on gA time delay controller for systems with Uncertain Dynamics, 27th January 2013.
- [18] S.B. Phadke, and S.E. Talole, Sliding Mode and Inertial Delay Control Based Missile Guidance, *IEEE Transactions on Aerospace and Electronic Systems*, Vol.48(4), pp.3331-3346, 2012.
- [19] K. Youcef-Toumi, and O. Ito, Controller design for systems with unknown nonlinear dynamics, *Proceedings of the American Control Conference*, Minneapolis, pp.836-845, 1987.
- [20] K. Youcef-Toumi, and O. Ito, A time delay controller for systems with unknown dynamics, *Journal of Dynamic System, Measurement, and Control*, - Transactions of the ASME, Vol.112(1), pp.133-142, 1990.

Attitude Control Simulation of a Legged Aerial Vehicle Using the Leg Motions

Yoshiyuki Higashi^{*1}, Soonki Chang²

¹Kyoto Institute of Technology, Faculty of Mechanical Engineering, 6068585, Japan

²Kyoto Institute of Technology, Division of Mechanodesign, 6068585, Japan

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 22 October, 2018

Online: 15 November, 2018

Keywords:

Attitude Control

Aerial Vehicle

Modeling

ABSTRACT

To gather information rapidly in disaster sites, a lot of search/rescue robots have been developed. It is difficult to correspond to the complicated environment composed of fields requiring various locomotion strategies, because most of these robots have only one type of locomotion device. To increase the available search routes under such conditions with the aim of gathering information more efficiently, we have previously proposed a legged aerial vehicle. The vehicle has tandem rotors to fly in the air and four legs to walk on the ground. The particular feature of this robot is that it has fewer actuators than the sum of those required for controlling a quadrupedal robot and a tandem-rotor helicopter individually. This paper presents modeling of the robot and development of an attitude control system that uses the leg motions. The behavior of the vehicle with the proposed attitude control is simulated using Multibody Dynamics (MBD) simulation software.

1 Introduction

Accidents and disasters such as building collapses, fires, earthquakes, and floods are responsible for massive loss of life. According to an interview survey about the Great Hanshin-Awaji Earthquake (also known as the Kobe Earthquake) of 1995, the most important contribution to rescue activities following such an event is rapid information gathering. Therefore, search robots are attracting attention as rapid means of gathering such information. In late years, a lot of researchers have been developed varied robots, and there has been remarkable progress in unmanned aerial and ground vehicles. For example, to save the electric power and to extend the inspection time on aged bridges, a UAV with a magnetic adsorption device was proposed and developed by Akahori et al [1]. The quadrotor helicopter-based UAV has an adsorption device using electro permanent magnets (EPM). And two cameras are set on the camera arms to carry out the close visual inspection.

Oliver et al. reported about a quadrotor helicopter with a tilt mechanism in [2]. The proposed quadrotor helicopter can fly horizontally while maintaining a steady attitude to conduct aerial inspection with stable

condition. The development of unmanned multirotor helicopters has led to expand applications, such as (i) inspecting aerial power lines for maintenance, (ii) constructing platforms with which to rescue people, and (iii) undertaking construction on inaccessible areas [3]. On the ground, walking robots are able to cross irregular ground that is impossible to wheeled and crawler robots, leading to the development of experimental dynamic walking robot[4], [5]. Also, unmanned vehicles have been developed that can both terrestrial locomote and fly in the air. This is because in search scenarios it is often necessary to both fly quickly over extended areas and search carefully on the ground. Pratt and Leang developed the dynamic underactuated flying-walking (DUCK) robot, which combines a quadcopter helicopter with passive-dynamic legs to create a various system that can fly and walk [6].

To coordinate walking and flying abilities, in [7], the authors designed and constructed the legged air vehicle shown in Fig. 1. The frame, legs and the body are constructed from pipes and plates of carbon-fiber-reinforced polymer for high stiffness and low weight. Two brushless motor and propellers are equipped as the thruster. And four legs composed of servomotors and parallel link mechanisms are used for walking and

^{*}Corresponding Author: Yoshiyuki Higashi, Matsugasaki, Sakyo-ku, Kyoto, +81-75-724-7364 & higashi@kit.ac.jp

controlling the attitude.

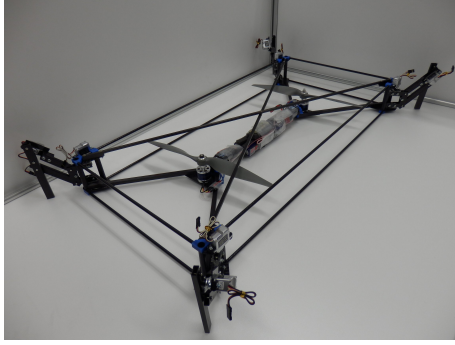


Figure 1: A prototype of the proposed legged aerial vehicle.

Table 1: Specifications of the legged aerial vehicle.

Length [mm]	702-870
Width [mm]	400-560
Height [mm]	120-200
Weight [g]	830

The flight attitude of the robot is controlled by changing of the center of gravity (CoG) caused by motion of the legs. The control using CoG can reduce the number of actuators than the sum of those required for controlling a quadrupedal robot and a tandem-rotor helicopter individually. With respect to flight control using change of CoG, there is a quadrotor helicopter controlled by an inverted pendulum. Miwa developed the quadrotor helicopter equipped an inverted pendulum on top of the body, and achieved control the attitude by using the tilt angle of the pendulum [8]. The present paper describes construction of mathematical models of the legs, rotors, and development of attitude control system based on them to simulate the motion of the robot in the air. And the behavior of the robot in the air is simulated using the constructed model and a physical model constructed in ADAMS. ADAMS is an analysis software to simulate the behavior of moving parts and distribution of loads and forces to the whole robot without solving the equations of motion analytically.

2 Model of a Leg and Parameter Identification

The inputs of the physical model constructed by ADAMS are forces such as the driving torques and thrusts of the legs and rotors, but the inputs of the actuator are voltage signals that depend on reference profiles of the leg angles and rotor speeds. Therefore, this section presents a mathematical model of the legs and identifies the relevant parameter values. The model is built to clarify the relation between the servomotor input and the drive torque. Figure 2 shows the servomotors for driving links around a roll axis and a yaw axis, and a parallel link mechanism of links.

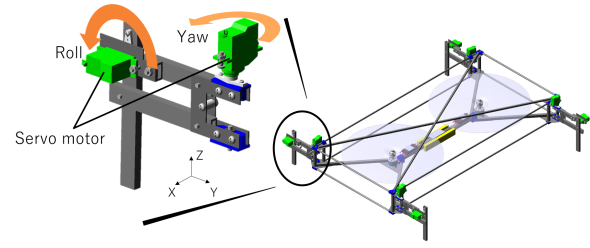


Figure 2: Servomotors and a parallel link mechanism of the leg

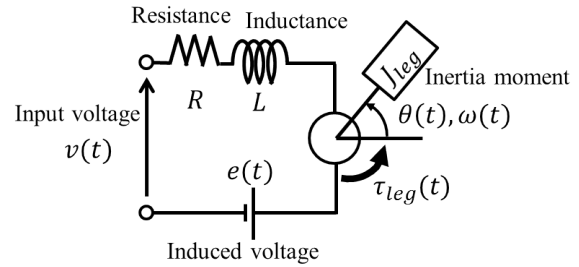


Figure 3: Electro-mechanical model of the direct current (DC) motor in one of the servomotors, including the effect of the moment of inertia, friction, and gravity.

We assumed that the leg could be modeled as a servomotor model that includes the influences of the moment of inertia, friction, and weight of the leg about each axis. Figure 3 shows the model of a direct current (DC) motor in servomotors. The driving torque of a the leg $\tau_{leg}(t)$ driving the leg is presented as the sum of the torques $\tau_m(t)$, $\tau_f(\omega(t))$, and $\tau_g(\theta(t))$. Here, $\tau_m(t)$, $\tau_f(\omega(t))$, and $\tau_g(\theta(t))$ are torque of the servomotor, friction, and gravity, respectively. Thus

$$\tau_{leg}(t) = \tau_m(t) + \tau_f(\omega(t)) + \tau_g(\theta(t)), \quad (1)$$

where t , $\omega(t)$ and $\theta(t)$ are time, the angular velocity and angle of the leg from a chosen reference.

The transfer function from the differential voltage to the motor $v_e(t) (= v(t) - e(t))$ to servomotor torque $\tau_m(t)$ the [9] is expressed as

$$\frac{T_m(s)}{V_e(s)} = \frac{K_t}{Ls + R}, \quad (2)$$

by using the coil inductance L , the motor resistance R and the torque coefficient K_t . And $T_m(s)$ and $V_e(s)$ are the Laplace transforms of $\tau_m(t)$ and $v_e(t)$.

The friction torque $\tau_f(\omega(t))$, which is a function of the angular velocity $\omega(t)$ of the leg, is given below[10]:

$$\tau_f(\omega(t)) = \begin{cases} -\tau_{act}(t) & \omega(t) = 0 \wedge |\tau_{act}(t)| < \tau_s \\ -\tau_s \text{sgn}(\tau_{act}(t)) & \omega(t) = 0 \wedge |\tau_{act}(t)| \geq \tau_s \\ \tau_a(\omega(t)) & \text{otherwise} \end{cases} \quad (3)$$

$$\tau_a(\omega(t)) = -\text{sgn}(\omega(t))\{a_1 \exp(-\text{sgn}(\omega(t))\omega(t)) + \text{sgn}(\omega(t))a_2 + a_3\omega(t)\}, \quad (4)$$

where τ_s is the static friction torque, $\tau_{act}(t)$ is the sum of $\tau_m(t)$ and the gravity torque $\tau_g(t)$, and a_1, a_2 , and a_3 ($a_1 + a_2 = \tau_s$) are positive. Therefore, the Laplace transforms of the leg angle $\Theta(s)$ is presented as

$$\Theta(s) = \frac{T_{leg}(s)}{J_{leg}s^2}, \quad (5)$$

Here, T_{leg} and J_{leg} are the Laplace transforms of $\tau_{leg}(t)$ and the moment of inertia of links and servomotors.

The servomotor controller is taken to be a proportional-derivative (PD) controller [9] as given by

$$v(t) = K_{sp}(\theta_{ref}(t) - \theta(t)) + K_{sd}(\omega_{ref}(t) - \omega(t)), \quad (6)$$

where K_{sp} and K_{sd} are the proportional gain and differential gain, respectively. $\theta_{ref}(t)$ is the reference input of the leg angle.

In generally servomotors are controlled by PD controller to earn rapid response. The controller does not include an integral (I) controller. Therefore the measurement angle has slight angle error to the reference as shown in Figs. 4 and 5. The drive angle of a DC motor in the servomotor is measured by an installed potentiometer or a rotary encoder. Measured value has some electric noise, however, the noise does not cause terrible error because these signal passes a low pass filter. The unknown parameters A, B and C are estimated using the Levenberg-Marquardt method which is one of the iterative technique that locates the minimum of a function that is expressed as the sum of squares of nonlinear functions. To identify the parameters, the response of the leg about each axis was measured by using following the three steps below:

1. Attach two light-emitting diodes (LEDs) to the leg (see Figs. 4 and 5).
2. Record the motion of the leg to a video (30 fps) and the reference angle to a micro computer.
3. Apply image processing to each frame of the video to calculate the coordinates of each LED.
4. From those coordinates, calculate the angle of the leg for each frame.

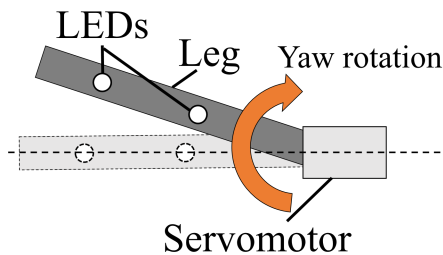


Figure 4: Leg motion in angle measurement experiment for the yaw rotation using LEDs.

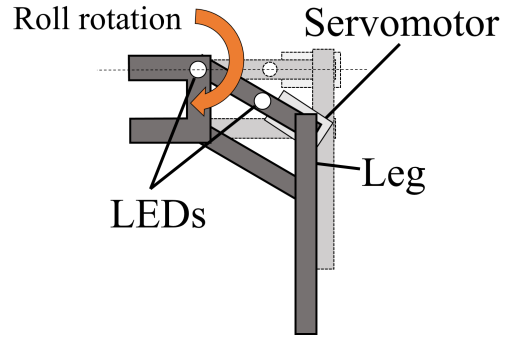


Figure 5: Leg motion in angle measurement experiment for the roll rotation using LEDs.

We determine the parameter values by minimizing the squared difference between the measurements and the simulation values using the models shown in Figs 4 and 5. Here, the simulation frequency was 1 kHz. Table 2 lists the identified parameter values, and Figs. 6 and 7 show the simulation results for each axis accompanied by the respective measurements. These simulation results shows that the behavior of the leg model is close to the measured behavior.

Table 2: Result of Parameter Identification

Symbol	Identified value
K_t	0.2016
a_3	0.035
K_{sp}	0.912
K_{sd}	0.005
a_2	0.070

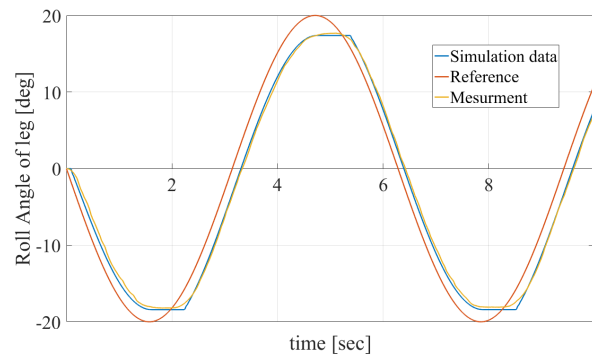


Figure 6: Simulation results of roll angle of the leg. Blue line is simulation result using identified parameters. Mearment data is measured roll angle through the experiment shown in Fig. 5.

3 Model of Rotor

We explain the model of a rotor composed of a propeller and a brushless motor assumed as a DC motor in this section. The brushless motor is assumed as a DC motor. Thus, as with Eq. (2), the transfer function from the differential voltage to the motor torque is given as

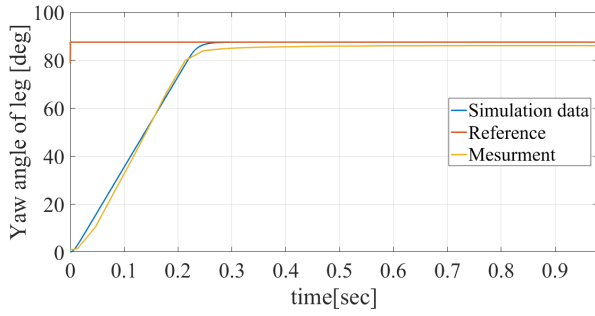


Figure 7: Simulation results of roll angle of the leg. Blue line is simulation result using identified parameters. Mearment data is measured yaw angle through the experiment shown in Fig. 4.

$$\frac{T_{bm}(s)}{V_{be}(s)} = \frac{K_{bt}}{L_b s + R_b}, \quad (7)$$

where each variable corresponds to the respective one in Eq. (2).

The thrust $T(t)$ generated by the propeller and the counter-torque $\tau_c(t)$ acting on the propeller due to the air are known to be proportional to the square of the rotor velocity in the steady state. Therefore, $T(t)$ and $\tau_c(t)$ are calculated using approximation curves based on the data sheet of the propeller on the manufacturer APC's web site [11]. Those thrust and torque are given as

$$T(t) = 3.026n^2(t), \quad (8)$$

$$\tau_c(t) = 5.039n^2(t), \quad (9)$$

where $n(t)$ is the rotation velocity in units of revolutions per minute (rpm); Eqs. (8) and (9) are plotted in Figs. 8 and 9, respectively. Moreover, the relationship between the rotation velocity $\psi(t)$ in units of radians per second (rad/s) and the torque $\tau_p(t) (= \tau_{bm}(t) - \tau_c(t))$ is

$$\frac{\Psi(s)}{T_p(s)} = \frac{1}{J_p s}, \quad (10)$$

where J_p is the inertia torque of the propeller, and $\Psi(s)$ and $T_p(s)$ are the Laplace transforms of $\psi(t)$ and $\tau_p(t)$. The response of the rotor is simulated based on these equations.

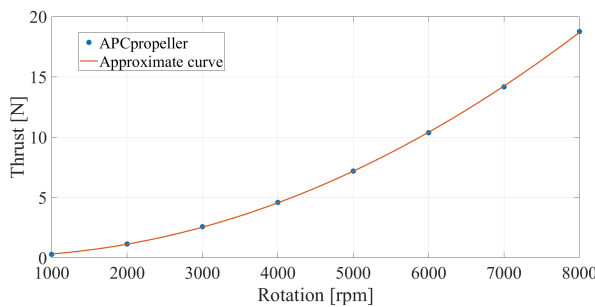


Figure 8: Thrust data from manufacturer's site [11] and approximation curve.

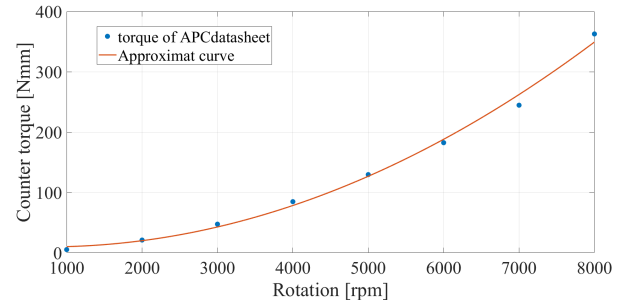


Figure 9: Counter-torque data from manufacturer's site [11] and approximation curve.

4 Attitude Control using the Leg Motion

4.1 Modeling of CoG and Leg Angles

To develop a control system for CoG control, we clarify relation between the CoG and the leg angles. The variables and link names are defined as shown in Figs. 10 and 11. The masses of links B, U, L, and F are m_b , m_u , m_l , and m_f , respectively, and those of the leg and the vehicle are m and M , respectively. The matrix $\theta \in \mathbb{R}^{2 \times 4}$ of legs angles is defined as

$$\theta = \begin{bmatrix} \theta_{FLy} & \theta_{FRy} & \theta_{RLy} & \theta_{RRy} \\ \theta_{FLr} & \theta_{FRr} & \theta_{RLr} & \theta_{RRr} \end{bmatrix} \quad (11)$$

$$= \begin{bmatrix} \theta_y \\ \theta_r \end{bmatrix} = \begin{bmatrix} \theta_{FL} & \theta_{FR} & \theta_{RL} & \theta_{RR} \end{bmatrix}. \quad (12)$$

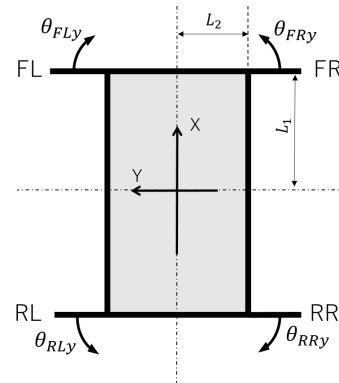


Figure 10: Definition of leg angles around the yaw axis.

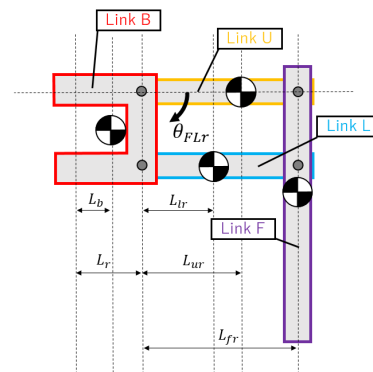


Figure 11: Definition of angles and link parameters on the front-left (FL) leg.

By the definitions, the robot CoG $\mathbf{G}(\boldsymbol{\theta}) = \{G_X, G_Y\}^T$ is expressed following Eq. (13) in terms of the various leg CoGs, namely $\mathbf{G}_{FL}(\boldsymbol{\theta}_{FL})$, $\mathbf{G}_{FR}(\boldsymbol{\theta}_{FR})$, $\mathbf{G}_{RL}(\boldsymbol{\theta}_{RL})$, and $\mathbf{G}_{RR}(\boldsymbol{\theta}_{RR})$:

$$\mathbf{G}(\boldsymbol{\theta}) = \frac{m}{M}(\mathbf{G}_{FL}(\boldsymbol{\theta}_{FL}) + \mathbf{G}_{FR}(\boldsymbol{\theta}_{FR}) + \mathbf{G}_{RL}(\boldsymbol{\theta}_{RL}) + \mathbf{G}_{RR}(\boldsymbol{\theta}_{RR})). \quad (13)$$

Each leg CoG is determined from the respective leg angle. The CoG of the front-left leg (FL) is given by

$$\mathbf{G}_{FL}(\boldsymbol{\theta}_{FL}) = \begin{bmatrix} \frac{1}{m}(mL_1 + ((m_u + m_l + m_f)L_r + m_bL_b + (m_uL_{ur} + m_lL_{lr} + m_fL_{fr})\cos\theta_{FLr})\sin\theta_{FLy}) \\ \frac{1}{m}(mL_2 + ((m_u + m_l + m_f)L_r + m_bL_b + (m_uL_{ur} + m_lL_{lr} + m_fL_{fr})\cos\theta_{FLr})\cos\theta_{FLy}) \end{bmatrix}. \quad (14)$$

The vehicle CoG is calculated from Eqs. (13) and (14) and the various leg angles.

However, the reference leg angles cannot be determined uniquely from the reference CoG because of $\boldsymbol{\theta} \in \mathbb{R}^{2 \times 4}$ and $\mathbf{G} \in \mathbb{R}^2$. Assuming that leg motion around the yaw axis and roll axis cause movement of the CoG along X and Y axis each, the conditions for the leg yaw angles are

$$\begin{cases} \theta_{RLy} = \theta_{RRy} = 0, \\ \mathbf{G}_{FLX}(\theta_{FLy}, \hat{\theta}_{FLr}) = \mathbf{G}_{FRX}(\theta_{FRy}, \hat{\theta}_{FRr}) \quad (G_X \geq 0) \\ \theta_{FLy} = \theta_{FRy} = 0, \\ \mathbf{G}_{RLX}(\theta_{RLy}, \hat{\theta}_{RLr}) = \mathbf{G}_{RRX}(\theta_{RRy}, \hat{\theta}_{RRr}) \quad (G_X < 0) \end{cases} \quad (15)$$

by using the estimated leg angle matrix $\hat{\boldsymbol{\theta}}$. And those for the leg roll angles are

$$\begin{cases} \theta_{FLr} = \theta_{RLr} = 0, & (G_Y \geq G_Y(\hat{\boldsymbol{\theta}}_y, 0)) \\ \theta_{FRr} = \theta_{RRr} = 0, & (G_Y < G_Y(\hat{\boldsymbol{\theta}}_y, 0)). \end{cases} \quad (16)$$

Also, when $G_Y \geq G_Y(\hat{\boldsymbol{\theta}}_y, 0)$ is satisfied, the additional conditions are

$$\begin{cases} G_{FRY}(\hat{\theta}_{FRy}, \theta_{FRr}) = G_{RRY}(\hat{\theta}_{RRy}, \theta_{RRr}) \\ (\{G_{FRY}(\hat{\theta}_{FRy}, 0) \geq G_{RRY}(\hat{\theta}_{RRy}, 0)\} \\ \wedge \{G_Y \geq \frac{m}{M}(G_{FLY}(\hat{\theta}_{FLy}, 0) + G_{RLY}(\hat{\theta}_{RLy}, 0) + 2G_{RRY}(\hat{\theta}_{RRy}, 0))\}) \\ \theta_{RRr} = 0, \\ G_{FRY}(\hat{\theta}_{FRy}, \theta_{FRr}) = \frac{M}{m}G_Y - G_{FLY}(\hat{\theta}_{FLy}, 0) - G_{RLY}(\hat{\theta}_{RLy}, 0) - G_{RRY}(\hat{\theta}_{RRy}, 0) \\ (\{G_{FRY}(\hat{\theta}_{FRy}, 0) \geq G_{RRY}(\hat{\theta}_{RRy}, 0)\} \\ \wedge \{G_Y < \frac{m}{M}(G_{FLY}(\hat{\theta}_{FLy}, 0) + G_{RLY}(\hat{\theta}_{RLy}, 0) + 2G_{RRY}(\hat{\theta}_{RRy}, 0))\}) \\ G_{FRY}(\hat{\theta}_{FRy}, \theta_{FRr}) = G_{RRY}(\hat{\theta}_{RRy}, \theta_{RRr}) \\ (\{G_{FRY}(\hat{\theta}_{FRy}, 0) < G_{RRY}(\hat{\theta}_{RRy}, 0)\} \\ \wedge \{G_Y \geq \frac{m}{M}(G_{FLY}(\hat{\theta}_{FLy}, 0) + G_{RLY}(\hat{\theta}_{RLy}, 0) + 2G_{FRY}(\hat{\theta}_{FRy}, 0))\}) \\ \theta_{FRr} = 0, \\ G_{RRY}(\hat{\theta}_{RRy}, \theta_{RRr}) = \frac{M}{m}G_Y - G_{FLY}(\hat{\theta}_{FLy}, 0) - G_{RLY}(\hat{\theta}_{RLy}, 0) - G_{FRY}(\hat{\theta}_{FRy}, 0) \\ (\{G_{FRY}(\hat{\theta}_{FRy}, 0) < G_{RRY}(\hat{\theta}_{RRy}, 0)\} \\ \wedge \{G_Y < \frac{m}{M}(G_{FLY}(\hat{\theta}_{FLy}, 0) + G_{RLY}(\hat{\theta}_{RLy}, 0) + 2G_{RRY}(\hat{\theta}_{RRy}, 0))\}). \end{cases} \quad (17)$$

When another condition is to be satisfied, the additional conditions are given with suffixes representing the right and left sides of the robot. Thus, the reference leg angles are determined based on the relation between a leg CoG and this angle in the conditions.

4.2 Calculation of Reference Position of CoG

The reference CoG for attitude control is calculated from the error in the attitude angle trajectory. In the first step, the pitch-roll-yaw torque around the CoG acting the entire aircraft $\boldsymbol{\tau}_{CoG}(t) = \{\tau_p(t), \tau_r(t), \tau_y(t)\}$ is calculated from a PD controller shown below:

$$\boldsymbol{\tau}_{CoG}(t) = \mathbf{K}_p \circ (\boldsymbol{\phi}_{ref}(t) - \boldsymbol{\phi}(t)) + \mathbf{K}_d \circ (\dot{\boldsymbol{\phi}}_{ref}(t) - \dot{\boldsymbol{\phi}}(t)) \quad (18)$$

by using the pitch-roll-yaw $(\phi_p(t), \phi_r(t), \phi_y(t))$, $\boldsymbol{\phi}(t) \in \mathbb{R}^3$, and their reference $\boldsymbol{\phi}(t)_{ref} \in \mathbb{R}^3$. Where $\mathbf{K}_p, \mathbf{K}_d \in \mathbb{R}^3$ are the proportional gain and differential gain, $\dot{\ast}$ is differential of the \ast , and \circ is Hadamard product. In the second step, the thrusts and the center of thrust (CoT), which is the position of the torque around the CoG balancing caused by the thrust, is estimated from the input values using the rotor model. The estimated thrusts and CoT are represented by $\hat{\mathbf{T}}(t) = \{\hat{T}_1(t), \hat{T}_2(t)\}^T$ and $\mathbf{C}\hat{\boldsymbol{\sigma}}\mathbf{T}(t) = \{C\hat{\sigma}T_X(t), C\hat{\sigma}T_Y(t)\}^T$, respectively. The final step determines the reference CoG $\mathbf{G}_{ref}(t) = \{X_{ref}, Y_{ref}\}^T$ from the torque, the thrust, and the CoT given by the first and second steps by the

following equation:

$$\mathbf{G}_{ref}(t) = \frac{1}{\hat{T}_1(t) + \hat{T}_2(t)} \{\tau_p(t), \tau_r(t)\}^T - \mathbf{C}\hat{\mathbf{o}}\mathbf{T}(t). \quad (19)$$

Meanwhile, reference rotor speed is determined by $\tau_y(t)$ and the reference of the sum of the thrusts is determined using Eqs. (10) and (9).

5 Attitude Control Simulation

The motion of the robot was simulated based on the physical model constructed by ADAMS and the control system constructed by a numerical analysis software MATLAB. Simulation parameters are the same values to those of the prototype. The sampling frequency of simulations was 1 kHz. And Table 4 shows reference angles in the simulations.

Figures 12 and 13 show the pitch angle for simulation pattern 1 and the roll angle for simulation pattern 2. Fig. 14 shows the result of the yaw angle for simulation pattern 3. Results in Figs. 12, 13, and 14 corresponded with the respective reference lines, and showed that the developed control system was effective through the simulations.

Figures 15 - 19 show detail of the pitch, the roll and the yaw angle in several reference angles including large reference. When the pitch reference is 20 deg shown in Fig. 15, the pitch angle follows the reference rapidly, although the roll angle changes like a sine wave in this simulation, the control performance is enough effective because the range is very narrow between -0.5 and 0.5deg. In case of the roll reference is 20 deg, the pitch angle did almost not change as shown in Fig. 16. The reason that the change of the pitch angle in Fig. 16 is smaller than the change of the roll in Fig. 15, is difference of the moment of inertia and variable range of the CoG

The moment of inertia of the robot around the pitch axis is larger than one around the roll axis. And the range of change of CoG to control the pitch angle is also wider than one of the roll control. Therefore the stability of pitch angle is better than the performance of the roll angle.

The effect appeared in the yaw angle too. However the change of except for target angle is very small, and these results showed validness of the developed controller. Especially the roll control worked in large reference 25 deg (Fig. 17) which was enough for flight control. The yaw control also generated slight change in other axes when reference are low (Fig. 18) and high (Fig. 19).

Table 3: Proportional-derivative (PD) controller gains

	K_p	K_d
Pitch	1.0	0.5
Roll	2.0	1.3
Yaw	0.1	0.55

Table 4: Reference of attitude angles in control simulations

	Pitch [deg]	Roll [deg]	Yaw [deg]
Pattern 1	5.0	0.0	0.0
Pattern 2	0.0	5.0	0.0
Pattern 3	0.0	0.0	20.0
Pattern 4	20.0	0.0	0.0
Pattern 5	0.0	20.0	0.0
Pattern 6	0.0	25.0	0.0
Pattern 7	0.0	0.0	10.0
Pattern 8	0.0	0.0	30.0

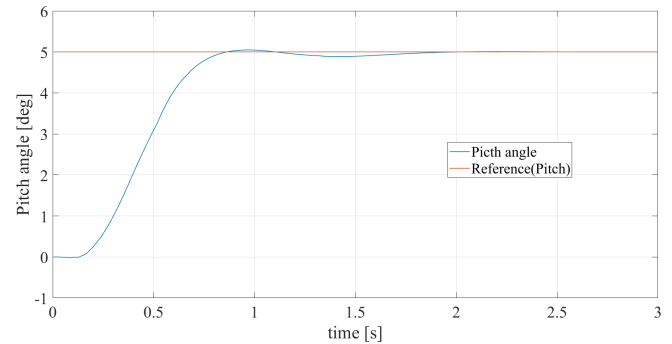


Figure 12: Step response of pitch-angle controller (pattern 1, reference pitch:5.0, roll:0.0, yaw:0.0).

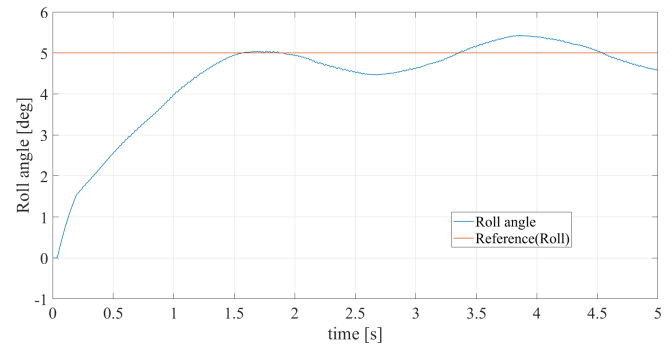


Figure 13: Step response of roll-angle controller (pattern 2, reference pitch:0.0, roll:5.0, yaw:0.0).

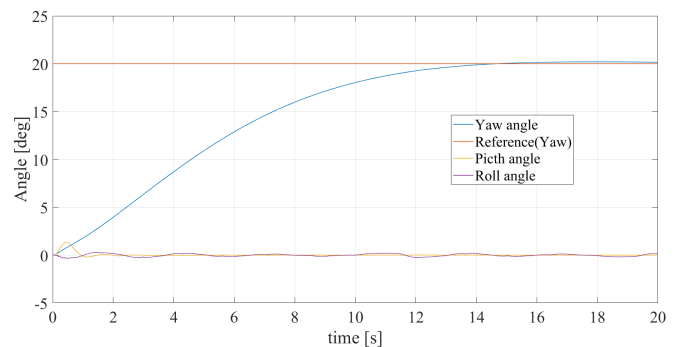


Figure 14: Step response of yaw-angle controller (pattern 3, reference pitch:0.0, roll:0.0, yaw:20.0)

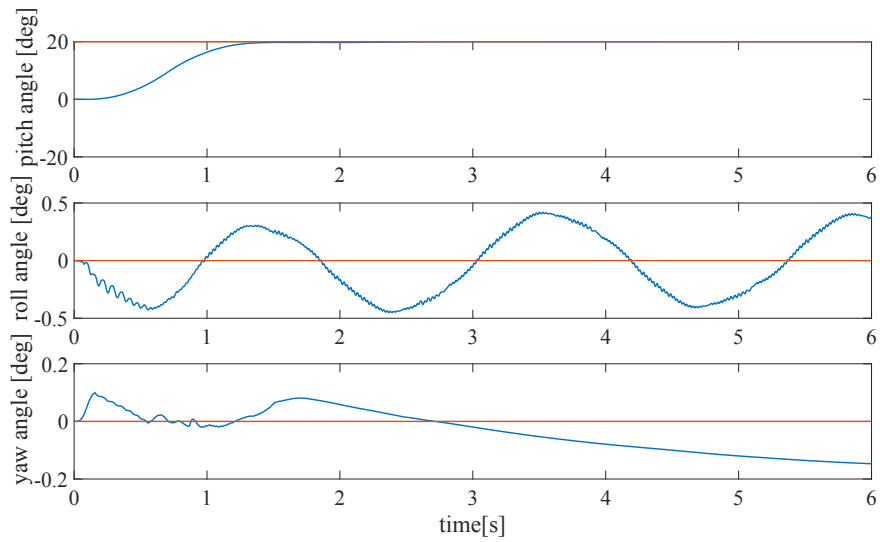


Figure 15: Response of three axes (pattern 4, reference pitch:20.0, roll:0.0, yaw:0.0)

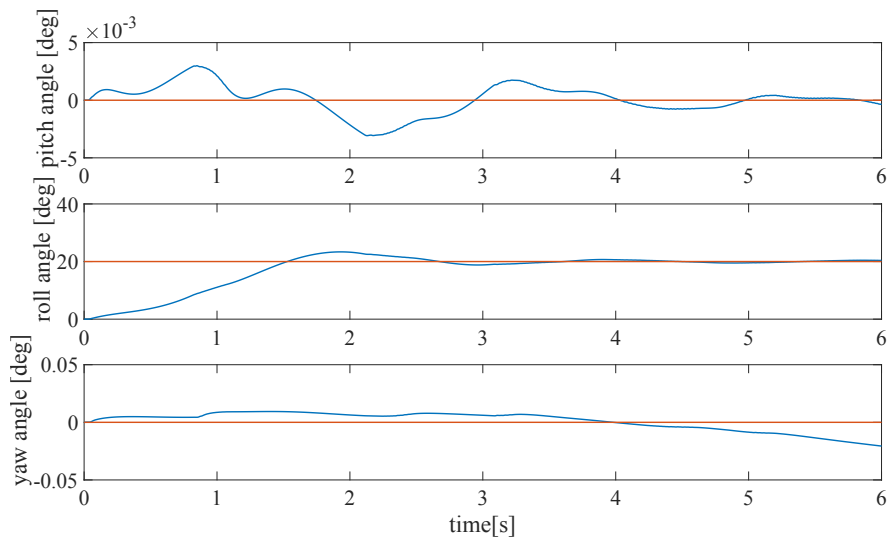


Figure 16: Response of three axes (pattern 5, reference pitch:0.0, roll:20.0, yaw:0.0)

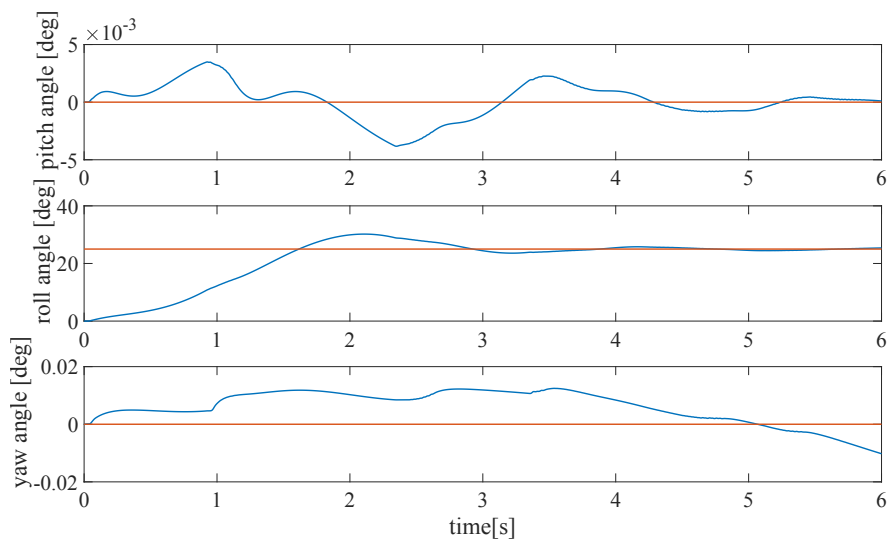


Figure 17: Response of three axes (pattern 6, reference pitch:0.0, roll:25.0, yaw:0.0)

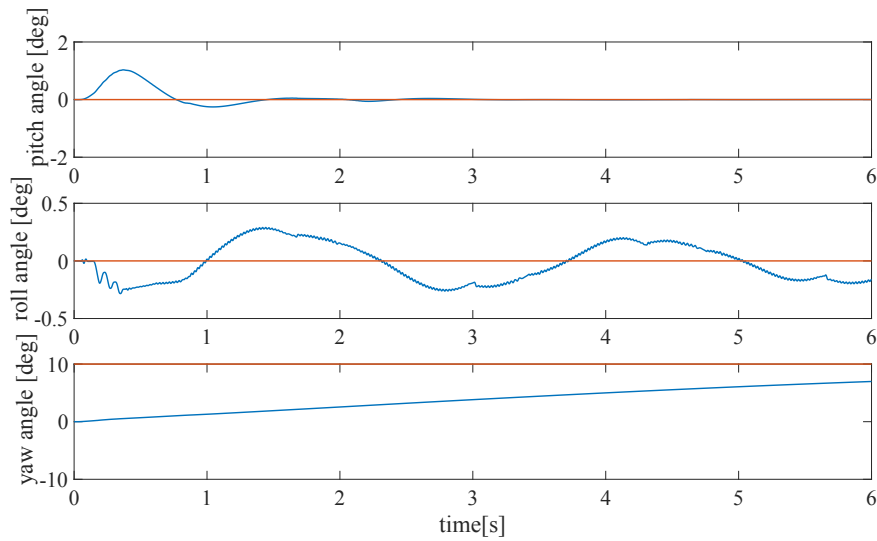


Figure 18: Response of three axes (pattern 7, reference pitch:0.0, roll:0.0, yaw:10.0)

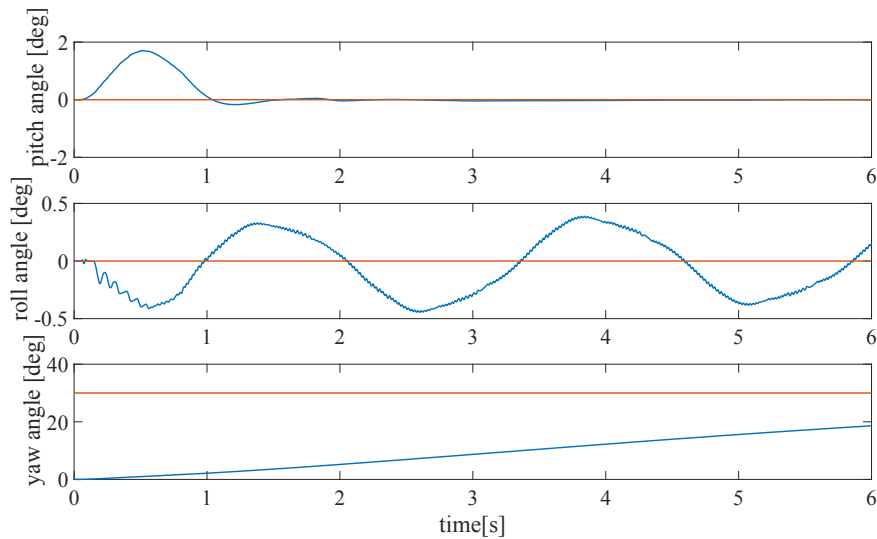


Figure 19: Response of three axes (pattern 8, reference pitch:0.0, roll:0.0, yaw:30.0)

6 Conclusions

The paper has described modeling of the leg and the rotor including the propeller, and a control method using the change of the CoG caused by the leg motion. Flight attitude was simulated for some reference angles by the developed control system. These simulation results shows that (i) response of the leg model corresponded to the experimental data and (ii) the attitude angles of the robot could follow their references using the attitude control system. Moreover the difference of moment of inertia and range of CoG in each axis appeared in the results, it was shown that the constructed model could reflect the characteristic of the robot. However, the results also showed that the transient response is not yet fast enough. Particularly convergence of the yaw angle for the yaw reference was slow, even though the yaw angle could follow the reference at last. Future work will involve implementing the control system on the developed legged aerial robot to evaluate the control performance.

References

- [1] S. Akahori, Y. Higashi, A. Masuda, Development of an Aerial Inspection Robot with EPM and Camera Arm for Steel Structures, Proceedings of the International Conference, pp. 3546-3549, 2016.
- [2] Russell Oliver, Sui Yang Khoo, Michael Norton, Scott Adams, Abbas Kouzani, Development of a Single Axis Tilting Quadcopter, Proceedings of the Proceedings of the International Conference, pp. 1851-1854, 2016.
- [3] A.E Jimenez-Cano, J. Braga, G. Heredia, A. Ollero, Aerial Manipulator for Structure Inspection by Contact from the Under-side, Proceedings of the International Conference on Intelligent Robots and Systems, pp. 1879-1884, 2015.
- [4] S. Kitano, S. Hirose, G. Endo, Design and Development of Quadruped Robot TITAN-XIII, Journal of Japan Society for Design Engineering, volume 51, number 12, pp. 875-884, 2016.
- [5] M. Raibert, K. Blanespoor, N. Gabriel, R. Playter, BigDog, the Rough - Terrain Quadruped Robot, Proceedings of the 17th IFAC World Congress, Vol, 41, Issue 2, pp. 10822-10825, 2008.
- [6] C. J. Pratt and K. K. Leang, Dynamic Underactuated Flying-Walking (DUCK) Robot, Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 3267-3274, 2016.

- [7] Y. Higashi and R. Okada, Suggestion of Legged Air Vehicle for Locomotion on ground and in air, Proceedings of the 8th International Conference on Intelligent Unmanned Systems, pp. 228-231, 2012.
- [8] M. Miwa, S. Kunou, S. Uemura, A. Imamura, H. Niimi, Attitude Control of Quad-Rotor Helicopter with COG Shift, Journal of JSEM, vol. 13, Special Issue, pp. s102-s107, 2013.
- [9] Ishikawa M., Kitayoshi R., Wada T., Maruta I., Sughie S., Modeling of R/C Servo Motor and Application to Underactuated Mechanical Systems, Transactions of the Society of Instrument and Control Engineers Vol.46, No.4, pp. 237-244, 2010.
- [10] B. Armstrong, Friction: Experimental Determination, Modeling and Compensation, Proceedings of the 1988 IEEE International Conference on Robotics and Automation, p.1422, 1988.
- [11] J. J. More, The Levenberg–Marquardt algorithm: Implementation and theory Proceedings of Conference on numerical analysis, CONF-770636-1, 1997.
- [12] APC Propeller Performance Data, http://www.apcprop.com/v/downloads/PERFILES_WEB/datalist.asp, 2017/01/27

iSensA - A System for Collecting and Integrating Sensor Data

João Manuel Leitão Pires Caldeira¹, Vasco Nuno da Gama de Jesus Soares^{1,*}, Pedro Miguel de Figueiredo Dinis Oliveira Gaspar², Joel José Puga Coelho Rodrigues³, Ricardo Manuel Valentim Fontes⁴, José Luís Lopes Silva⁴

¹Instituto Politécnico de Castelo Branco, Instituto de Telecomunicações, Portugal

²Universidade da Beira Interior, Portugal

³National Institute of Telecommunications (Inatel), Brazil; Instituto de Telecomunicações, Portugal; University of Fortaleza (UNIFOR), Brazil

⁴Instituto Politécnico de Castelo Branco, Portugal

ARTICLE INFO

Article history:

Received: 28 September, 2018

Accepted: 26 October, 2018

Online: 10 November, 2018

Keywords:

iSensA

Arduino

Sensors

Monitoring system

Web and mobile applications

Internet of Things

ABSTRACT

The idea of monitoring several types of parameters in various environments has been motivating significant research works in Internet of Things (IoT). This paper presents the design and construction of iSensA, a system for integrating and collecting information from sensors. The solution implements a multi-sensor monitoring system and then expands the monitoring concept to an IoT solution, by employing multi-network access, Web services, database and web and mobile applications for user interaction. iSensA system is highly configurable, enabling several monitoring solutions with different types of sensors. Experiments have been performed on real application scenarios to validate and evaluate our proposition.

1. Introduction

This paper is an extension of work originally presented in conference CISTI 2018 [1].

Nowadays, industry faces new challenges and paradigms derived from the current industrial revolution, the 4th one, named as Industry 4.0. This new era of industrial development aims to provide an expedite answer to the fast and dynamic requirements of production, and promoting its operational effectiveness and efficiency, and thus contributing to higher productivity. As described by [2] and [3], the most relevant features of Industry 4.0 are related with information and communication technologies & electronics (ICT&E), advanced algorithms, added value and knowledge management. These features can be summarized as: (1) digitization, optimization, and production customization; (2) automation and adaptation; (3) human machine interaction (HMI); (4) value-added services and businesses, and (5) automatic data exchange and communication. To accomplish these objective, Industry 4.0 relies on several technologies, being mobile computing, cloud computing, big data, and the Internet of Things (IoT) the key technologies of Industry 4.0 [3-7]. Among these concepts, the IoT paradigm has been extensively focused in the last years, since it covers a wide scope of applications in various fields.

* Vasco Soares, Email: vasco.g.soares@ipcb.pt

It brings ubiquitous intelligence through the interconnection of equipment (sensors, devices, etc.) to the Internet.

Results from a market and literature survey of IoT-based monitoring show that some developments have been reported in implementation of such systems for example for monitoring: water level [8,9] and water quality [10,11], power consumption [12,13], smart agriculture [14,15], garbage [16,17], solar energy production [18], air quality [19], equipment working status [20]. Instead of focusing on a specific problem and its solution, the work presented in this paper proposes a generic IoT data acquisition solution, that can be widely applied to different scenarios. It addresses the following main challenges:

- Applicable to different environments, with distinct requirements;
- Send and receive data to and from remote servers over Ethernet, Wi-Fi or GSM/GPRS;
- Configuration of operation and control rules based on parameters and specific requirements;
- Statistical tools to support real-time or historical data management and analysis, trend analysis, and data correlation;
- Configuration of reports, alarms/notifications, which can be sent by e-mail and/or SMS;

- High scalability;
- Online platform accessible via computer, tablet or smartphone;
- Complete and secure backup in the cloud;
- Implemented using open source technology.

The main contributions of this work are as follows. Firstly, an intelligent data acquisition system, called iSensA [31], is proposed. It is based on Arduino controllers and sensors, Web and mobile applications supported by a cloud architecture with a relational database and Web services. Secondly, the effectiveness of this system for monitoring, analyzing real-time or historical data and ensuring incident detection is tested. The proposed system is experimentally evaluated and validated in several real-world case studies of different sectors with specific requirements. Thirdly, it is assessed if the system facilitates resource management and contributes to reduce operational expenses.

The remainder of this paper is organized as follows. Section 2 introduces iSensA system for remote monitoring. Section 3 presents the test and validation of iSensA in real scenarios, and discusses resource consumption. Section 4 concludes the paper and points further research directions.

2. iSensA

iSensA system was developed using a three-tier architecture [21] that comprises a presentation tier, a business tier, and a data tier. The three layers in the three-tier architecture are as follows. The client layer contains the user interface part of the application. Data is presented to the user or input is taken from the user. The business layer acts as an interface between the client layer and the data access layer. It is responsible for business logic like validation of data, calculations, data insertion, etc. Finally, the database is located in the data layer.

Figure 1 presents the iSensA architecture. The Server #1 is responsible for data processing. It calculates indirect values such as total, average, minimum, and maximum values. These values are calculated from direct values sent by the controllers. Indirect values are used for charts and for reporting purposes. The Server #2 is also used due to the large amount of data potentially collected from the controllers. It allows dividing the load between web / mobile users and controllers. The platform architecture was conceived as a modular system suitable to be integrated into third-party middleware.

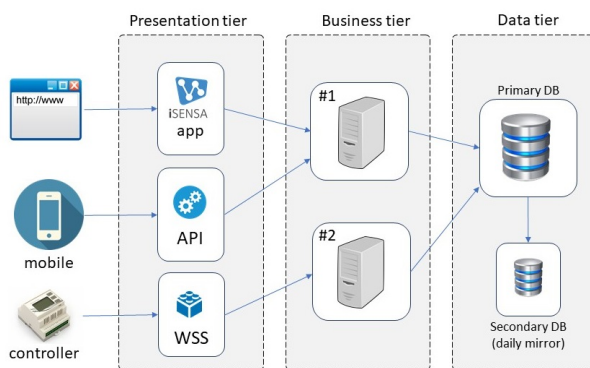


Figure 1. Overview of the iSensA system.

iSensA development required: (1) the construction of a data acquisition and control module, performed through a controller, sensors and electronic circuits; (2) the implementation of a data acquisition and processing interface, Web and mobile applications, supported by a cloud architecture with a relational database and Web services. The following subsections present the controller, the Web application, and the mobile application.

Due to space limitations it is not possible to present the features of the system and applications in detail. The interested reader may contact the authors for detailed information and request access to test it.

2.1. Controller

The iSensA controller is an Arduino-based hardware platform [22]. The adoption of this platform in development environments allows rapid prototyping of functional solutions. Characterized by the use of Atmel AVR microcontrollers, with intuitive programming based on the C programming language, this platform is becoming a true competitor to more complex control systems currently used by companies [23]. A great advantage of this type of platform is its flexibility. Providing multiple inputs and outputs, digital and analog, these platforms become adaptable to the acquisition of any type of signals. With the increasing success and the low-cost of these platforms, several sensors are available for measurement and collection of various parameters. Measuring sensors, which lately only existed for industrial controllers, are now available for use on low-cost platforms such as Arduino. Modularity is another great feature of these platforms, so they can be built according to the specific needs of a particular application.

Taking advantage of all these features, the iSensA controller is based on an Arduino platform. Depending on the requirements of each application scenario, an iSensA controller can be coupled to a communication module using Ethernet, wireless (Wi-Fi) or mobile Global System for Mobile Communications (GSM). These modules are required for communication between the controller and the iSensA Server #2. Figure 2 presents the iSensA controller in several of its different configurations.

The operation of iSensA controllers is based on collecting data from the attached sensors and sending them to the iSensA Server #2 (Figure 1). For that, the controller starts by trying to get Internet connection thru the DHCP service. After, it enters into an infinite loop divided into three main modules, “Register Control”, “Collect Data” and “Sensor Input”. The first module to be executed is “Register Control” and it is executed only once. This module is responsible for the registration of controllers with the iSensA Server #2. The registration procedure improves security to the system by only allowing pre-approved controllers to communicate with iSensA Server #2. Therefore, each controller before start sending data must be known by iSensA Server #2. After successful registration, the controller starts executing “Collect Data” and “Sensor Input” modules periodically. “Collect Data” reads the real-time values from the attached sensors while “Sensor Input” performs all the tasks needed to send the collected data to Server #2. Figure 3 presents the flowchart of the firmware executed by iSensA controllers.

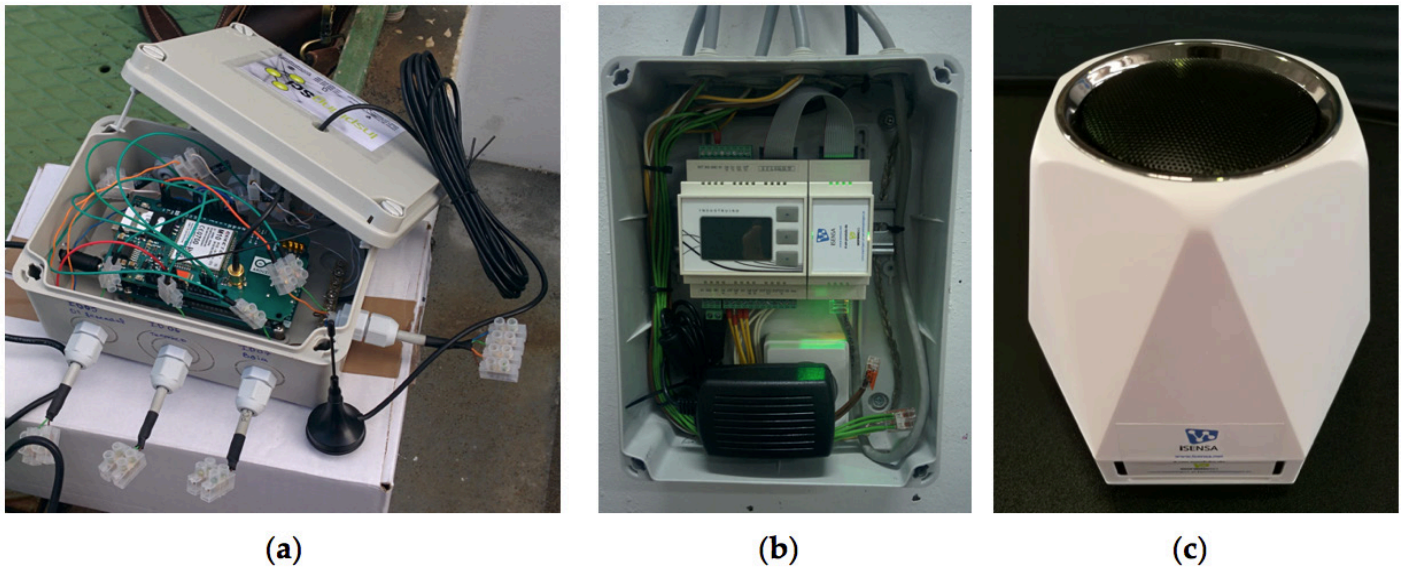


Figure 2. iSensA controller: (a) Initial prototype of the controller with a GSM module; (b) Current version of the controller with an Ethernet module; (c) Controller with a Wi-Fi module.

All data exchanged between iSensA controllers and the Server #2 is encrypted using a proprietary algorithm to ensure data integrity and privacy. iSensA Server #2 provides two secure Web services for controllers' communication – “WSRegisterControl” and “WSSensorInput”. The first one is used by “Register Control” module. The information sent to this Web service is the following: MAC address of the controller; Time interval for sending data; Controller model; Sensor list with identification (ID) of each sensor and type. “WSSensorInput” is used by “Sensor Input” module. The information sent to this Web service is the following: MAC address of the controller; Sensor list with identification of each sensor and the corresponding value. These Web services ensure that all the data collected by controllers reaches the iSensA server with the proper context. The MAC address identifies the controller that sent the data and the sensor ID the sensor to which the value corresponds.

To ensure continuous operation of the controllers some error recovery routines were included into the algorithm. If controllers cannot communicate with iSensA Server #2 within a number of attempts, then controllers reset themselves. This may occur whenever the Internet connection has been lost and must be restarted again. Besides that, all the firmware execution is shielded by a reset trigger associated to a timeout. If for some reason the firmware gets stuck at some point of its execution, beyond the timeout, then the reset trigger is activated and makes the controller to reset itself. These mechanisms guarantee robustness to the system, allowing it to recover from most uncontrolled situations.

The first tests of the system in real-world deployments revealed several limitations and challenging issues. For example, there were lots of problems concerning network communications, in particular, GSM communication. It was observed that the controller would stop working after losing GSM connection, which occurred very often. An analysis of this problem revealed that the native Arduino library procedure responsible for getting a link to the GSM network was blocking. Therefore, a timeout recovery algorithm was implemented in the GSM library to prevent the firmware to get infinitely stuck waiting for a GSM link.

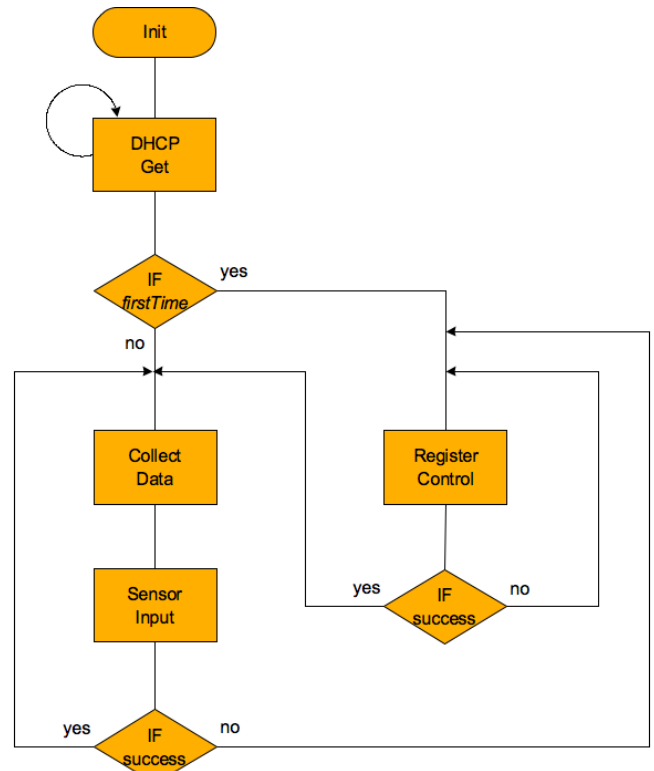


Figure 3. iSensA controller firmware flowchart.

Another issue that occurred on some iSensA real-world deployments, was caused by the signal loss in cabling between sensors and controllers. After thorough electrical analysis and several tests, it was concluded that this problem resulted from two main reasons. The first one due to long cables. This was solved by decreasing the pullup resistors used in the sensors connections, thus allowing the controllers to receive the correct values from sensors. The second one resulted from cables that were installed in noisy environments, near to pumps, electrical links, etc. In such situations, the problem was solved by using shielded cables to connect sensors to controllers.

2.2. Web Application and Relational Database

Open-source technologies were used to develop an intuitive and user-friendly Web application that can be used at any time, from anywhere, in different electronic devices (computers, tablets, mobile phones, etc.). PHP runs on the server side, being embedded in HTML. JavaScript is a scripting language executed on the client side, interpreted by the web browser, and allows the creation of interactive pages. jQuery is a quick, small and feature rich Javascript API, making HTML and event handling, animations and Ajax much simpler and easier to use. On the client side Bootstrap was also used. It is a front-end framework that allows agile development in responsive web projects. It allows integration of jQuery with other technologies such as CSS and LESS, for the dynamism desired in the interaction with the user. Bootstrap is also used in the development of platform design. Google Charts framework is used to plot data into rich and interactive charts.

Due to the characteristics of the Web application to be developed, the ICONIX methodology was selected. The analysis, design and implementation phases of the application were supported by its processes.

The Web application shown in Figure 5 provides, among others, the following functionalities:

- Secure Web access available anytime anywhere;
- Accessible through any of the major Web browsers on any device;
- Configurable dashboards;
- Easy and intuitive navigation;
- Centralized management of facilities;
- Setting access levels for different users;
- User defined alarms that can be received by email or cell phone text message;
- User defined reports that can be received by email;
- Real-time data visualization;
- Historical data analysis;
- Data export in various formats.

The database presented in Figure 6, is implemented in a MySQL fully normalized schema that allows for rapid complex querying. HTTPS was set up to protect the integrity of the Web application, and the security of its users. It allows users to securely connect to the Web application.

2.3. Mobile Application

The mobile application named MobiSensA [1] aims to extend the range of features of iSensA system to mobile devices. It allows monitoring the installations, showing the status and information of controllers and sensors in real time, without having to depend on the Web application and email. Furthermore, it notifies users when alerts or anomalies occur.

ICONIX methodology was also used for the development of the mobile application. MobiSensA was developed for smartphones with Android OS, using the Android Studio integrated development environment, and the SQLite Database Browser tool for database management. To implement the application and organize the interface, a set of specifications and

guidelines of the Material Design language, referring to usability and accessibility was used. The Git tool was used to enforce versioning. Firebase technology was used to keep the application's local repository updated. A robust and secure Web Services API provides access to data from the iSensA database. Used protocols and technologies are illustrated in Figure 4.

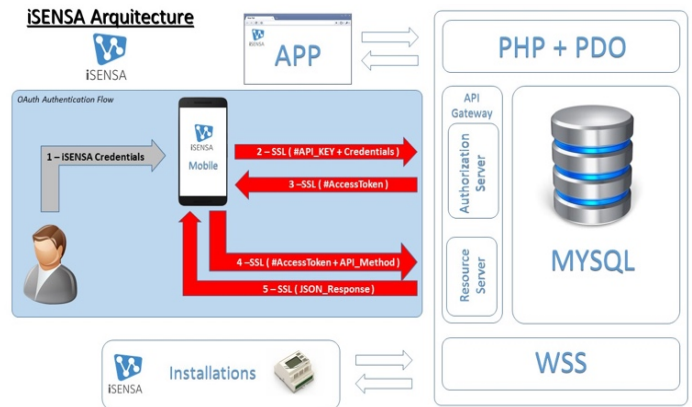


Figure 4. System's architecture.

The following are the requirements for the mobile application.

Non-functional requirements:

- The application is developed for Android 4.4 or later.
- The application limits size of stored data.
- The application works without an internet connection (offline).

Functional requirements:

- The user must login to use the application;
- The application shows the user's installations, together with the status of the controllers;
- The application allows visualizing the entire structure of the user data (installations, controllers and sensors);
- The application allows visualizing a sensor details, readings and alert history;
- The application icon must show the number of installation alerts and/or anomalies;
- The user may control the update frequency of the controllers/sensors;
- The user may decide the number of days the information is kept in history;
- The application notifies the user when an anomaly or an alert occurs;
- The application shows the number of anomalies and alerts that have occurred.

Some screenshots of the mobile application are shown in Figures 7-9. Figure 7 a) shows the start interface with a login option; b) displays the user's installations and controllers. Figure 8 a) shows detailed information of a controller; b) detailed information of a sensor that is connected to a controller. Figure 9 a) shows the sensor alerts; b) notification of controller anomalies.



Figure 5. iSensA Web Application.

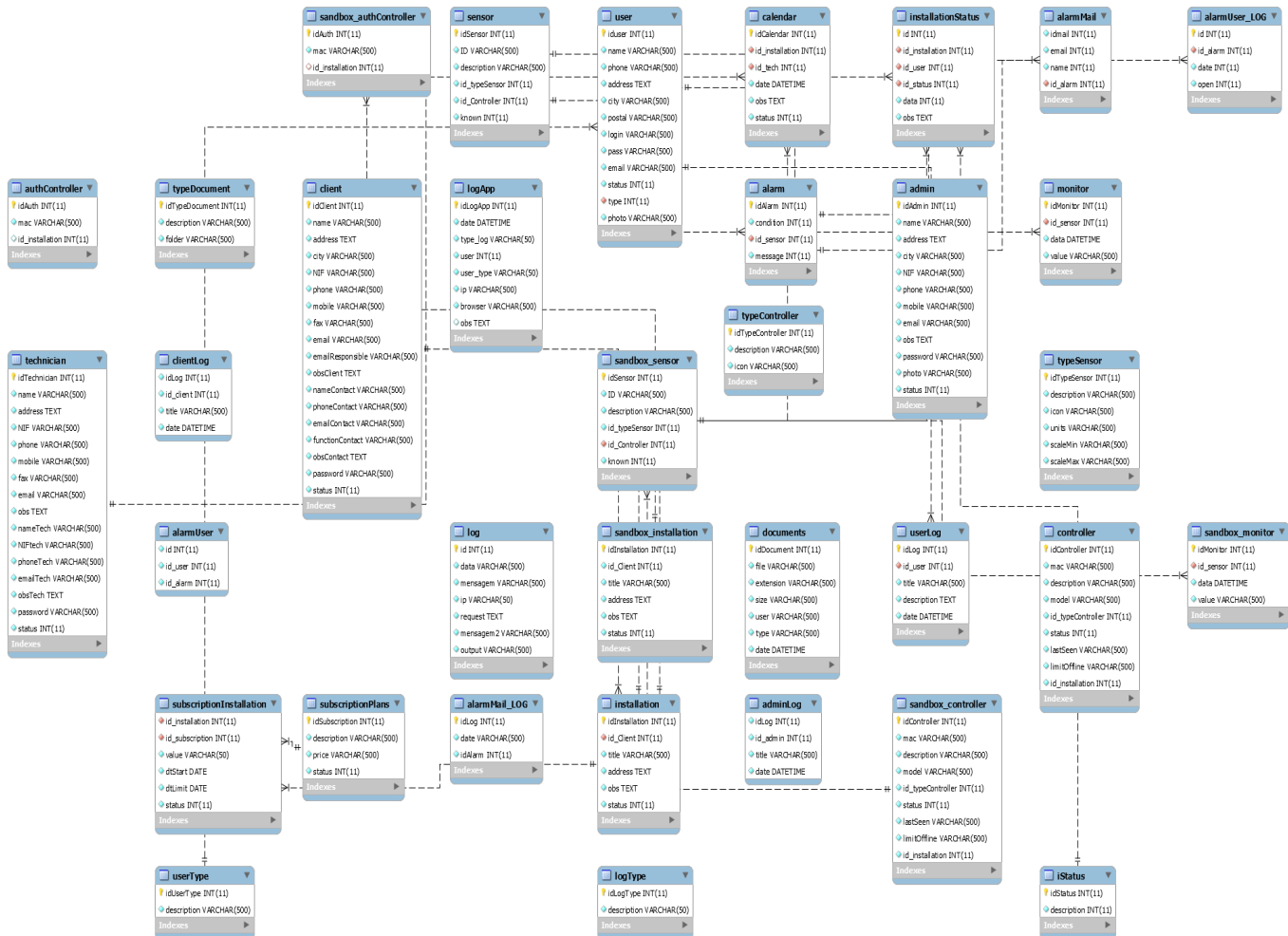
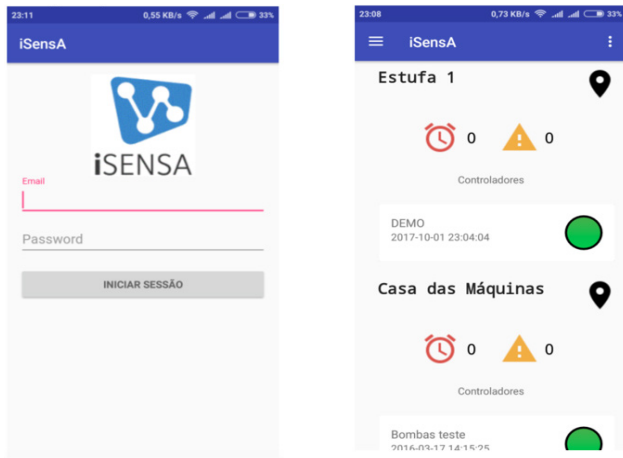


Figure 6. iSensA Relational Database.



a) b)
Figure 7. Screenshots: a) login; b) user's installations.



a) b)
Figure 8. screenshots: a) detailed information of a controller; b) detailed information of a sensor.

This section presents two real application scenarios that were used for testing and validation of the iSensA system. The evaluation of the iSensA resource consumption is also discussed.

ALBIGEC [24], a company responsible for the management of the swimming pool complex in the city of Castelo Branco, Portugal, wanted to monitor and remotely control the swimming pools water temperature, the bath waters temperature, and the female and male bathrooms temperature and humidity. iSensA's first prototype was deployed there. After a detailed study of the site and the requirements meeting, it was concluded that it would be necessary to install two iSensA controllers. One of these controllers was installed next to the swimming pool. Using three temperature-humidity sensors it monitors the temperature and humidity of the bathrooms and the central building. In order to monitor the water temperature of the two swimming pools and the sanitary waters, another controller was installed in the technical zone, next to the water tanks.

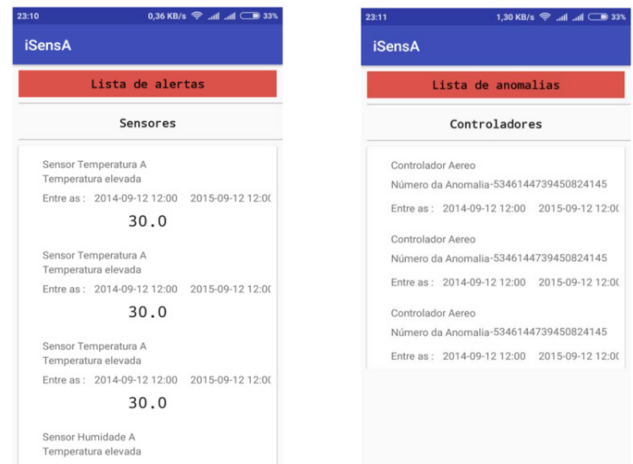


Figure 9. Screenshots: a) sensor alerts; b) controller anomalies.

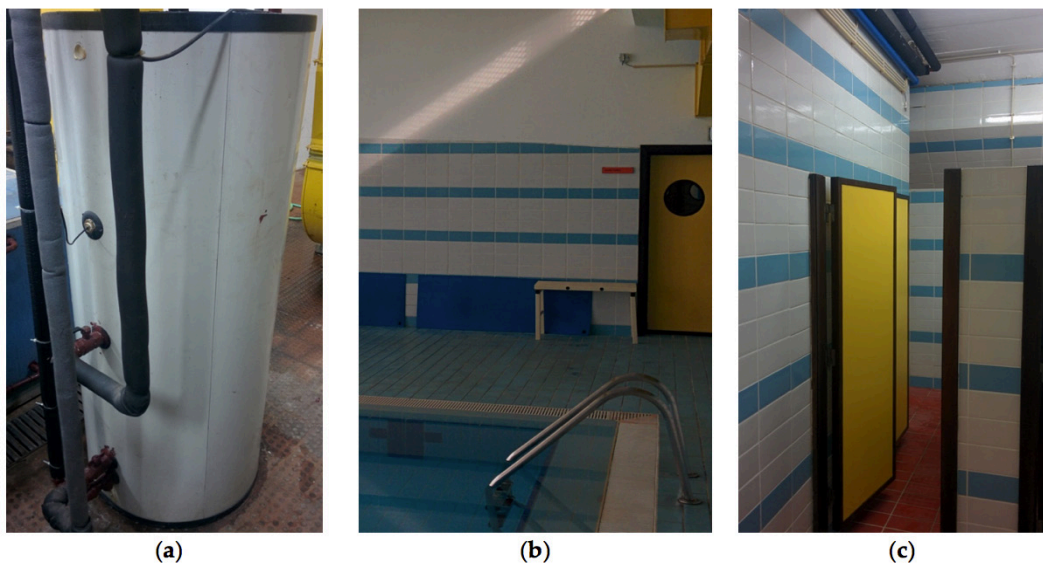


Figure 10. Sensors for swimming pool monitoring: (a) Sensing the bath waters temperature; (b) Sensing the temperature and humidity of the environment in the main pool; (c) Sensing the temperature and humidity in a bathroom.

In this case, four temperature sensors were placed in the outlet pipes that go from the water tanks to the swimming pools, and from the sanitary water tanks to the bathrooms. Both controllers communicate with the iSensA Web services using the building's Ethernet LAN.

Following the initial success, ALBIGEC extended the use of iSensA to another swimming pool complex under its management, located in Alcains village, Portugal. This was a simpler scenario where only one controller was required. In this case, the parameters to be monitored were the temperature and humidity in two bathrooms (female and male), the temperature of the swimming pool water and the temperature of the bath waters. Since an Ethernet connection wasn't available, a GSM modem was connected to the controller to allow the communication with the iSensA Web services.

The Municipality of Pampilhosa da Serra, Portugal [25], has also successfully adopted iSensA for controlling their swimming pool complex. Figure 10 shows the installation of iSensA in this scenario. Figure 5 displays the iSensA Web platform showing the values collected in this installation.

The functionalities provided by iSensA Web application allowed both companies to program alarms with upper and lower limits for all measured parameters. These alarms trigger alerts notified by e-mail and allow companies to react accordingly. These notifications can help prevent any downtime from the swimming pool complex due, for example, to low water temperatures. iSensA also provides automated statistical reports essential to effective management. Before using iSensA, only a limited amount of data was collected on paper forms by the maintenance staff and entered into the computer at a later date. Therefore, it was difficult to sort through and analyze it in an efficient manner.

3.2. Monitoring and Control of Sewer Pumping Stations

SMAS [26], a company responsible for the municipal services in the city of Castelo Branco, Portugal, proposed as a challenge the use of iSensA to monitor a sewer pumping station. Sewer pumping stations are used to move wastewater to higher elevations in order to allow transport by gravity flow. Sewage is fed into and stored in a wet well. When the sewage level rises to a predetermined level, an electric pump is started automatically to lift the sewage upward through a pressurized pipe system from where it is discharged into a gravity manhole again. This process is repeated until the sewage reaches a treatment plant. No tool exists currently for remotely monitoring SMAS stations. Their workers visit each of the sewer pumping stations once a day, to check for problems or failures. If a fault occurs between visits, it will not be detected, and it will lead to overflows of wet wells that must be avoided due to potentially posing health and environment hazards.

iSensA was deployed in one sewer pumping station to allow for control and remote monitoring of events that occurred there. Figure 11 shows the installation in this scenario. The deployed system was comprised of an iSensA controller connected to three amperometric clamps and a level sensor. The current clamps are used to measure the instantaneous current consumption of each phase of the electric pump (in this case a three-phase electric pump). The level sensor measures the height of the water in the wet well in real time. In addition, the controller has three relays

connected. When the well float drops, an overflow will occur, and the first relay is triggered. When there is a malfunction in the electric pump soft-start, the second relay is triggered. Finally, a differential protection relay can also be triggered. The second and third relays can help detecting if the pump did not start. Figure 12 presents the electric block diagram of the installed controller.

An SMS notification is sent to the telephone numbers pre-programmed on the controller, each time one of the relays is triggered. This allows workers to be notified immediately. Again the Web application allows to set alert values for both the values measured by the amperometric clamps and the water level in the well, and these alarms are then notified by e-mail. In the absence of a cabled network connection, a GPRS/GSM module was connected to the controller. This allows sending data to iSensA Web services and also sending SMS messages. SMAS workers can now react more quickly and resolve the problems sooner.

3.3. Performance and Resource Usage

The above-described scenario of the swimming pool complex of the Municipality of Pampilhosa da Serra was used for assessing iSensA system performance and resource consumption. The characterization of workload and resource consumption was performed based on server logs, including both data exchanged between iSensA controllers and the Server #2 (Figure 1) and resource consumption of the Web services requests, as well as the volume of data stored in the database.

The two iSensA controllers deployed in this scenario, acquire sensor data and send it to the server every 15 minutes. For example, during the month of October 2017, 14233 data messages were sent from the controllers to the server. The average size observed for register and data messages was about 210.18 octets. The collected sensor data over this period of time required approximately 18.06 MB of storage space in the database. Total bandwidth consumption by the Web services was 2995794 octets. These results allow to conclude about the effectiveness and efficiency of iSensA system.

4. Conclusions

It is predicted that Industry 4.0 will accelerate industry to achieve remarkable levels of operational efficiencies and thus largely increase productivity [27,28]. Among technologies, IoT is considered as a key enabler for the new-generation advanced manufacturing, Industry 4.0 [29]. IoT can be seen as a mean of connecting physical objects to the Internet as a ubiquitous network that enables objects to collect and exchange information. Thus, it blends and expands the traditional automation systems and industrial informatics systems into a much broader context [30].

In this paper, a system for reading and acquiring sensor data - iSensA - was presented. iSensA is based on Arduino controllers, web and mobile applications supported by a cloud architecture with a relational database and web services. iSensA enables the monitoring and analysis of data, environments and devices in a wide range of areas. The testing and validation of iSensA in real scenarios, show the soundness of the presented system to adapt to different application requirements. It contributes efficiently to incident detection, resource management, and cost reduction.

Future work includes: (1) using the gathered historical data to develop predictive models to help anticipate potential future events /incidents and to identify energy saving opportunities; (2)

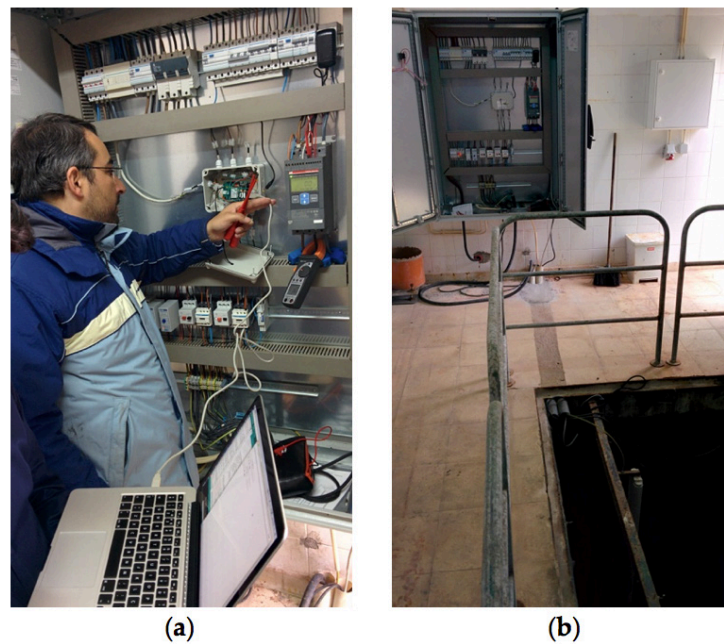


Figure 11. Monitoring a sewer pumping station: (a) Installation of the controller; (b) Wet well.

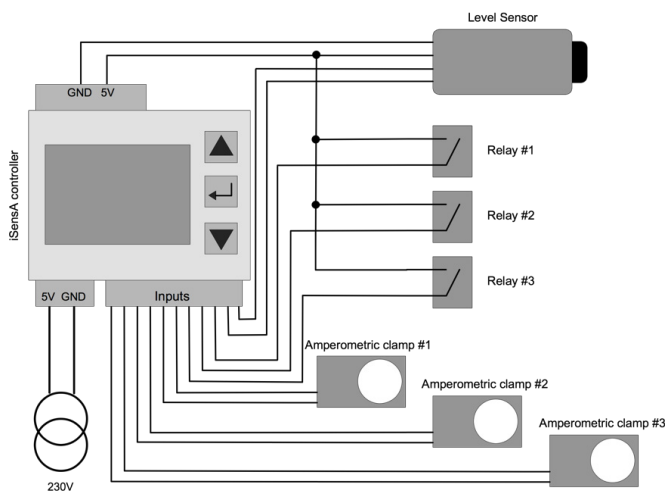


Figure 12. Electric block diagram of the installed controller.

assessing iSensa system for safety critical applications; (3) testing and maintenance of the mobile application.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work is funded by FCT/MEC through national funds and when applicable co-funded by FEDER – PT2020 partnership agreement under the project UID/EEA/50008/2013, and by Finep, with resources from Funttel, Grant No. 01.14.0231.00, under the Centro de Referência em Radiocomunicações - CRR project of the Instituto Nacional de Telecomunicações (Inatel), Brazil. The authors would like to acknowledge the companies InspiringSci, Lda, Enérgico Balanço Unipessoal Lda, ALBIGEC - Empresa de Gestão de Equipamentos Culturais, Desportivos e de Lazer, E. M., and SMAS - Serviços Municipalizados de Castelo Branco, as well as the Municipality of Pampilhosa da Serra, for their interest and

valuable contribution to the successful development of the iSensa system.

References

- [1] J. Silva, R. Fontes J. M. L. P. Caldeira, V. N. G. J. Soares, and P. D. Gaspar, "MobiSensa: Desenvolvimento de uma Aplicação Móvel para a Plataforma iSensa," in 13th Iberian Conference on Information Systems and Technologies (CISTI2018), Cáceres, Spain, June 13-16, 2018.
- [2] Posada, J., Toro, C., Barandiaran, I., Oyarzun, D., Stricker, D., et al, "Visual computing as a key enabling technology for Industry 4.0 and industrial Internet," IEEE Computer Graphics and Applications, 35(2), pp. 26-40, 2015.
- [3] Roblek, V., Meško, M., Krapež, A, "A Complex View of Industry 4.0," SAGE Open 2016, 6(2), 2016.
- [4] Gruber, F.E, "Industry 4.0: A best practice project of the automotive industry," Digital Product and Process Development Systems, Springer Berlin Heidelberg, pp. 36-40, 2013.
- [5] Vijaykumar, S., Saravanakumar, S.G., Balamurugan, M, "Unique sense: Smart computing prototype for Industry 4.0 revolution with IOT and bigdata implementation model," Indian Journal of Science and Technology, 8(35), 2015.
- [6] Wan, J., Tang, S., Shu, Z., Li, D., Wang, S., et al, "Software-defined industrial Internet of Things in the context of Industry 4.0," IEEE Sensors Journal, 16 (22), pp. 7373-7380, 2016.
- [7] Lu, Y., "Industry 4.0 - A survey on technologies, applications and open research issues," Journal of Industrial Information Integration, doi: 10.1016/j.jii.2017.04.005, 2017.
- [8] Pragati, D., Kirtikumar, S. V., "IoT based Water Monitoring System: A Review," International Journal of Advance Engineering and Research Development, 4(6), 2017.
- [9] Sukriti, Gupt, S., Indumathy, K, "IoT based Smart Irrigation and Tank Monitoring System," International Journal of Innovative Research in Computer and Communication Engineering, 4(9), 2016.
- [10] Vaishnavi, D., Gaikwad, M, "Water Quality Monitoring System Based on IoT," Advances in Wireless and Mobile Communications, 10(5), pp. 1107-1116, 2017.
- [11] Geetha, S., Gouthami, S, "Internet of things enabled real time water quality monitoring system," International Journal for @qua – Smart ICT for Water, 2(1), 2017.
- [12] Kurde, A., Kulkarni, V, "IoT Based Smart Power Metering," International Journal of Scientific and Research Publications, 6(9), 2016.
- [13] Pandit, S., Mandhre, S., Meghana, N, "Smart Energy Meter using Internet of Things," Vishwakarma Journal of Engineering Research, 1(2), 2017.
- [14] Dharti, V., Borole, A., Singh, S., "Smart Agriculture Monitoring and Data Acquisition System," International Research Journal of Engineering and Technology, 3(3), 2016.

- [15] Athani, S., Tejeshwar, C., Patil, M., Patil, P., Kulkarni, R., "Soil moisture monitoring using IoT enabled arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka – India," Proceedings of the International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017), India, 10-11 February, pp. 43-48, 2017.
- [16] Kumar, N., Vuayalakshmi, B., Prarthana, R., Shankar, A., "IoT based smart garbage alert system using Arduino UNO," Proceedings of the 2016 IEEE Region 10 Conference (TENCON), Singapore, 22-25 November, 2016.
- [17] Revathy, S., Anandhi, A., Hemalatha, K., "IoT Based Smart Bin Monitoring Using Sensor and GSM for Smart Cities," International Journal of Research in Computer Science, 4(1), 2017.
- [18] Patil, S., Vijayalashmi, M., "Solar Energy Monitoring System Using IoT," Indian Journal of Scientific Research, 15(2), pp. 149-155, 2017.
- [19] Fioccola, G., Sommese, R., Tufano, I., Canonico, R., Ventre, G., "Polluino: An efficient cloud-based management of IoT devices for air quality monitoring," Proceedings of the IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), Bologna, Italy, 7-9 September, 2016.
- [20] Patel, K., Patel, S., "IoT Based Data Logger for Monitoring and Controlling Equipment Working Status and Environmental Conditions," International Journal of Innovative Research in Computer and Communication Engineering, 4(4), 2016.
- [21] Bouhai N. (Editor), S. Imad (Editor), "Internet of Things: Evolutions and Innovations Volume 4," Wiley: NJ, USA, ISBN 9781119476573, 2017.
- [22] Arduino. Available online: <https://www.arduino.cc/> (accessed on 7 November 2017).
- [23] Reneker, D., "Arduino vs. PLC for industrial control." In Control Design for Machine Builders, Jul 13, 2017. Available online: <https://www.controldesign.com/articles/2017/arduino-vs-plc-for-industrial-control/> (accessed on 7 November 2017), 2017.
- [24] ALBIGEC. Available online: <http://www.albigec.pt/> (accessed on 21 September 2018).
- [25] CM Pampilhosa da Serra. Available online: <http://www.cm-pampilhosadaserra.pt/> (accessed on 21 September 2018).
- [26] Serviços Municipalizados de Castelo Branco. Available online: <http://www.sm-castelobranco.pt/> (accessed on 21 September 2018).
- [27] Hermann, M.; Pentek, T.; Otto, B., "Design principles for Industry 4.0 scenarios," Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS), Institute of Electrical and Electronics Engineers - IEEE-, IEEE Computer Society, Kauai, Hawaii, 5-8 January, pp. 3928-3937, 2016.
- [28] Thames, L., Schaefer, D., "Software-defined cloud manufacturing for Industry 4.0," Proceedings CIRP, 52, pp. 12-17, 2016.
- [29] Trappey, A.J.C., Trappey, C.V., Govindarajan, U.H., Chuang, A.C., Sun, J.J., et al., "A review of essential standards and patent landscapes for the Internet of Things: A key enabler for Industry 4.0," Adv. Eng. Informat, 2016.
- [30] Queiroz, D.V., Alencar, M.S., Gomes, R. D., Fonseca, I.E., Benavente-Peces, C., "Survey and systematic mapping of industrial wireless sensor networks," Journal of Network and Computer Applications, 2017.
- [31] iSensa. Available online: <http://www.isensa.net/> (accessed on 21 September 2018).

A Novel Fair and Efficient Resource Allocation Scheduling Algorithm for Uplink in LTE-A

Havva Esra Bilisik*, Radosveta Sokullu

Engineering Faculty, Department of Electrical-Electronics Engineering, Ege University, 35040, İzmir, Turkey

ARTICLE INFO

Article history:

Received: 28 August, 2018

Accepted: 30 October, 2018

Online: 10 November, 2018

Keywords:

Long Term Evolution – A

Uplink Transmission

Radio Resource Allocation

Algorithms

Network Throughput

User Fairness

ABSTRACT

With the introduction of new services and new more sophisticated mobile devices the radio network operators are faced with new challenges to increase the system performance. The most recent standards introduced by 3GPP for the new architectures of the Long Term Evolution network address these issues and outline possibilities for optimizing network performance and user QoS. A major instrument in that respect is the resource scheduling and allocation procedure. So far many different algorithms have been proposed. Uplink resource allocation however is less covered, because it poses additional constraints which make it difficult to balance the optimization between channel state information, system throughput and user perceived throughput. In this paper we propose a novel algorithm for resource allocation which balances the advantages of two previously suggested ones, specifically Round Robin and Best-CQI. We also define a new parameter, the user ratio, which allows us to explicitly quantify the trade-off between fairness, system throughput and user throughput for different channel conditions.

1. Introduction

In recent years the number of smart mobile devices as well as the number of various applications they support has increased in unprecedented proportions. In turn this has increased immensely the network traffic and has changed its characteristics posing many new challenges for the network engineers and network operators. A major instrument which regulates the relation between user demands and network traffic is the network adopted procedure for radio resource allocation, which is defined in the respective network standard.

The Long-Term Evolution (LTE) standard proposed by 3GPP and its latest version Long Term Evolution – Advanced (LTE-A) are the most recent telecommunication standards introduced to meet increasing user demands in terms of high data rates and better quality of service. LTE provides flexible deployments allowing low latency and supporting up to 300 Mbps of data transmission in downlink and up to 75 Mbps throughput for uplink. [1, 2, 3] The standard defines two separate radio access methods for the transmission in the downlink (Base Station to user) and the transmission in the uplink (user to Base Station). Orthogonal frequency division multiple access (OFDMA), selected for the downlink is not suitable for uplink transmissions mainly due to its high Peak to Average Power Ratio (PAPR). Another multiplexing

method, namely single carrier-frequency division multiple access (SC-FDMA) is proposed for the uplink [1, 2].

According to the LTE-A architecture, the Base Station known as “Evolved Node B (eNodeB)”, regulates the resource allocation process in both transmission directions. [2]. Functions and algorithms for allocating network resources for the downlink and the uplink are part of the Medium Access Control (MAC) layer at the eNodeB. Since this work is focused on uplink resource allocation, from here on the discussion will concentrate on the specifics of uplink transmission and resource allocation.

In the uplink (UL) the modified, pre-coded form of the OFDMA known as SC-FDMA is adopted in order to reduce cell interference and Peak to Average Power Ratio (PAPR). It is well known that the user equipment’s (UE) battery life is quite limited so SC-FDMA fits well the major requirement for the access method used in the uplink - to be power efficient. However, despite its obvious advantages, there are some additional constraints which make it more difficult to allocate resources in the uplink than in the downlink. These constraints include above all singularity, contiguity, and transmit power constraints [5]. The constraint defined as “singularity” mandates that a given resource block (RB) can be allocated only to a single user. The constraint defined as “contiguity” implies that all RBs allocated to a given user must be contiguous. The third constrained, defined as “transmit power constraint” in its turn requires that the maximum transmit power

* Havva Esra Bilisik, Email: bilisik.h@gmail.com

for any user should be less than or equal to 23 dBm [5]. Regarding the transmission modes LTE-A standard defines both time division (TDD) and frequency division duplex (FDD). In the time domain, the transmissions are realized in “time domain frames”. Each frame has duration of 10 ms and is respectively divided into 10 consecutive sub-frames with durations of 1 ms each, the so called Transmission Time Interval (TTI). In its turn, a given sub-frame is subdivided into two slots with durations of 0.5 ms each, carrying 7 OFDM symbols. Regarding the frequency domain, the available system bandwidth is divided into sub-carriers of 15 kHz. Finally, the resources according to the LTE-A standard are defined both in time and frequency domain in terms of Resource Block (RBs). Each RB consists of 12 subcarriers, with a total bandwidth of 180 kHz and last for 0.5 ms in the time domain. The number of available resources (i.e. RBs) will change for different system bandwidths.

LTE is an all-IP packet-based technology. The network architecture defined for LTE and LTE-A comprises the Evolved Node B (eNodeB), the Evolved Packet System (EPS) and the User Equipment (UE) [1,2]. The eNodeB is the entity between the UE and the network. That is why the eNodeB is responsible for allocating resources to the UEs both in the downlink and uplink. In the MAC layer of the eNodeB itself resides the so-called Packet Scheduler (PS), which controls the resource scheduling and allocation process. Its major task is to allocate RBs to all the UEs for every TTI with duration of 1 ms. Every TTI, each UE sends a Sounding Reference Signal (SRS) to its serving eNodeB. Based on the received SRS the eNodeB assigns a metric, called Channel Quality Indicator (CQI), with values ranging from 1 to 15 for each UE. Using this metric, the eNodeB defines each UEs channel quality value and generates a matrix called the “channel gain matrix”. The rows in that matrix represent the UEs and the columns are the available RBs. By allocating resources to UEs using this matrix, the spectral efficiency is maximized.

In this article, a novel Fair and Efficient Resource Allocation (FERA) algorithm for the uplink resource allocation scheduling is proposed and evaluated. The rest of the paper is organized as follows. Section 2 summarizes related works on uplink scheduling algorithms. Section 3 gives details on resource allocation procedures in LTE-A uplink. In section 4, the proposed algorithm, the system model and the metrics used to evaluate and compare the proposed algorithm with other well-known scheduling algorithms are detailed. Section 5 discusses the simulation scenarios that were conducted; Section 6 presents results and discussion followed by the conclusion.

2. Related Works

Resource allocation scheduling algorithms for LTE UL have been discussed by many scientists [2, 4-11]. Due to the specific constraints of singularity, contiguity, and transmit power, the resource allocation process is more complicated in the uplink as compared to the downlink. The several most popular scheduler algorithms so far are Round Robin (RR), Best CQI (B-CQI) and Proportional Fair (PF).

RR is the simplest resource allocation algorithm compared to other scheduling algorithms. This algorithm is channel-unaware and aims to achieve high fairness. The assumption is that the channel conditions do not change during the transmission time,

and therefore the channel conditions are not taken into account by the allocation algorithm. RBs are assigned to users one by one and fairly among all users [6]. This approach is often used because it guarantees high overall fairness in the resource allocation. Its main disadvantage is that the overall throughput is quite low compared to other algorithms. In [7] the authors define and use the RR algorithm in a time-based fashion, ensuring RB allocation sequentially to the users in a circular manner.

Unlike RR, the Best-Channel Quality Indicator (B-CQI) scheduling algorithm is an example of a channel-aware scheduling algorithm. It takes into account the current state of the channel while allocating resources to UEs [8]. B-CQI scheduling algorithm allocates RBs to any UE depending on how good the channel quality of that UE is. Any UE that has data to send transmits its CQI value to its serving eNodeB. A high CQI value reflects good channel conditions for the given UE. Therefore, the UE that has higher CQI values can be allocated more RBs as compared to UEs that has lower CQI values. Thus the B-CQI’s main objective, to increase the overall throughput, is achieved. However, because the RB allocation is done solely based on the CQI value of an UE, the resources might not be distributed fairly among all users. B-CQI algorithm increases the throughput of some users however some users, especially the UEs located close to the cell edge, might be left to starve.

The third most popular algorithm, known as the Proportional Fair (PF) is a combination of the above described RR and B-CQI algorithms [9]. The goal of this algorithm is to achieve sufficient throughput while maintaining fairness among users. The main idea of the PF algorithm is to determine users with relatively good channel conditions and assign resources to them.

In [7], the authors proposed a new scheduling algorithm “Modified RRBC” which is based on a different combination of RR and B-CQI algorithms. The newly proposed algorithm uses RR algorithm in the first time slot to allocate RBs to UEs, and in the second time slot allocates RBs to UEs with higher CQI value and lower RB allocation in the previous slot. The proposed algorithm has been compared to RR and B-CQI in terms of fairness, throughput and average queuing delay. Results show that the proposed algorithm performs better in terms of throughput than RR algorithm and slightly better in terms of fairness and the average delay as compared to B-CQI.

Besides these major resource allocation algorithms, there a number of other suggestions which try to balance the optimization criteria between high fairness (low throughput, low complexity) and high throughput (low fairness, complexity). The authors in [4], propose a new Mobility Aware scheduling algorithm that takes advantage of the simplicity of RR and B-CQI algorithms and aims to reduce their disadvantages taking into consideration the mobility of the users. The performance of the proposed algorithm is compared to RR and B-CQI in terms of fairness, throughput and block error rate (BLER). Results show that proposed algorithm performs similar to RR in terms of fairness and similar to B-CQI in terms of throughput for the downlink. In [9], the authors studied the throughput-fairness trade-off considering three different uplink scheduling algorithms: RR, Maximum Throughput (MT) and First Maximum Expansion (FME). The results show that considering fairness among users RR gives better performance than the other

two algorithms, so with VoIP and Video flows RR is the best scheduler. On the downside however, when considering the best effort flows RR algorithm shows the worst performance.

The Approximate Maximum Throughput (AMT) algorithm was proposed by the authors in [10], aiming to solve the resource allocation problem in a more computationally efficient way than the Optimal Maximum Throughput (OMT), which was proposed before. The AMT algorithm maximizes the throughput by using a heuristic approach. AMT algorithm allocates RBs to UEs with good SNR and the UEs with low SNR value are not served at all. The authors compare the algorithm in terms of throughput and BLER with B-CQI and Kwan Maximum Throughput (KMT) algorithms. Results show that the proposed AMT algorithm performs similar to B-CQI and better than KMT in terms of throughput. BLER parameter is compared using two users and the results show that AMT and B-CQI algorithms show similar values but KMT algorithm shows lower BLER value.

In [11] the authors propose a novel resource allocation algorithm for the uplink, which they call Opportunistic Dual Metric scheduling based on Quality of service and Power Control (ODM-QPC). The algorithm aims to maximize the system throughput and Quality of Service (QoS) while keeping track of the fairness among users. The algorithm allocates the maximum number of RBs to each UE in accordance with its target QoS. Once RBs are allocated to all UEs, Power Control (PC) regulations are applied to the UE's emission power without affecting the user's throughput. The proposed resource allocation algorithm achieves high aggregated throughput and quite reasonable fairness while meeting the satisfaction of the user QoS and the reduction of the user's power depending on his channel conditions. The authors compare their algorithm with B-CQI and PF in terms of throughput and fairness respectively. The results show that ODM-QPC achieves throughput close to that obtained by B-CQI for all simulated cases. In terms of fairness ODM-QPC performs similar to PF algorithm.

Excluding RR, most of the other algorithms discussed above are quite computationally demanding. Some of them require more than one-step iteration, which means that they are also time demanding. Thus, the focus of this work was to design a fast, simple algorithm that can adapt to different channel conditions and provide users with similar performance balancing between high throughput and high fairness independent of the different channel conditions.

3. Uplink Resource Allocation

In modern wireless communications the abundance and specifics of applications and devices creates differences in the traffic created in the downlink and uplink direction. These differences are adequately reflected in the LTE-A standard, where, as mentioned in section 1, two different multiple access methods are specified. Naturally these reflect on the specific resource allocation procedure, that is why in this section we first elaborate on these differences and the specifics of uplink resource allocation before proceeding with the proposed algorithm.

The first major difference is limited power and computation resources available. Compared to the eNodeB the handheld mobile device operates on limited battery life and considerably restricted

computational resources. This mandates very efficient data transmission. The solution provided by the standard is the SC-OFDMA which ensures low PAPR but introduces a number of additional constraints (singularity, contingency) to the resource allocation procedure. The second originates from the abundance and variety of applications which can be initiated by the user and the difficulty to predict the number of resources a user will need to exchange data with the eNodeB. Additional constraints of contiguity and singularity also add up to make uplink resource allocation more complicated than downlink.

The goal of the network is to serve its users in accordance with their needs providing, low latency in transmission, allocating the required resources and distributing resources fairly among different UEs. The major optimization criteria for the uplink can be defined based on the requirements of the two sides involved: the user and the network. From the user point of view, the major criterion is fairness while from the network point of view throughput is the most important one. A good uplink resource allocation algorithm will provide the right balance and trade-off between these criteria. On the other hand it has to be simple and fast to execute.

The details of uplink resource definitions can be found in [2]. As explained before in Section I resources are defined in terms of frequency and duration. For the uplink, the smallest resource unit is the resource element. Each resource element corresponds to a square in the resource grid (Figure 1), and is defined by the index pair (k,l) where k and l are the indices in the frequency and time domains respectively. An uplink physical channel corresponds to a set of resource elements carrying information originating from higher layers. The transmitted signal within each slot is defined by a resource grid of $N_{RB}^{UL} N_{SC}^{RB}$ subcarriers and N_{Symb}^{UL} SC-FDMA symbols. The N_{RB}^{UL} parameter is based on the bandwidth set in cells and must satisfy the following relation:

$$N_{RB}^{min,UL} \leq N_{RB}^{UL} \leq N_{RB}^{max,UL}$$

According to the standard the smallest and the largest bandwidths supported are for the uplink are $N_{RB}^{min,UL} = 6$ and $N_{RB}^{max,UL} = 110$ respectively.

A physical resource block on the other hand is defined as N_{Symb}^{UL} : consecutive SC-FDMA symbols in the time domain and N_{SC}^{RB} consecutive subcarriers in the frequency domain. Thus, the definition of a Resource Block corresponds to $N_{Symb}^{UL} \times N_{SC}^{RB}$ resource elements or equivalently to one slot in the time domain and 180 kHz in the frequency domain.

The procedures of allocating resources are part of the Radio Resource Management (RRM), located in the eNodeB's Medium Access Layer (MAC) [14]. The specific algorithm specifying resource allocation in time and frequency domain is carried out in every TTI. During every TTI, each UE has to send the so-called Sounding Reference Signal (SRS) to the serving eNodeB, which will allow it to assign a metric for the quality of the channel between the two, the CQI as explained before [1]. The metric is used by the eNodeB to create a channel gain matrix for all UEs, where a row will correspond to a UE, and the columns define the available RBs. This channel gain matrix is used by the resource

allocation algorithm and allows it to maximize the overall efficiency of the system [15].

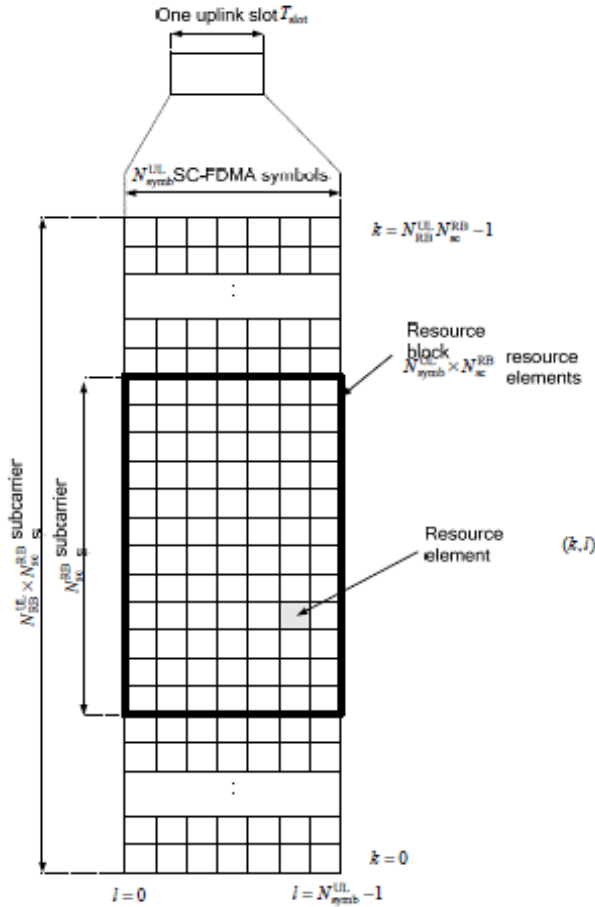


Figure 1: Uplink Resource Grid [2]

4. Proposed Algorithm and System Model

4.1. System Model

In this work, one eNodeB and varying numbers of UEs are used (10, 20 and 30). In order to clearly show the differences between examined algorithms, three different types of channel models for wireless channel between eNodeB and the UE are considered. These channel models are, Typical Urban, Rural Area and Pedestrian Channel B. The 3GPP proposes the Typical Urban channel model (TU) to represent densely populated areas with highest path loss. The Rural Area (RA) channel model is suggested for less populated areas [16]. The Pedestrian channel model (Ped) is for mobile users with low-mobility, with speed less than 30 km/h. Details of the specific parameters for these channel models are given in section 5.

The traffic model for all UEs is assumed that of full-buffer. The scenarios were conducted using different types of channel models, scheduling algorithms, system bandwidth and number of UEs.

4.2. Notations and Performance Metrics

In this section the notations and the metrics used in this article are detailed.

The number of users in the system at a given moment is defined as N. The throughput of the i^{th} user is given as:

$$T_i = \text{throughput of } i^{th} \text{ user}$$

While the average throughput for the system is:

$$T_{avg} = \text{average system throughput.}$$

As the algorithms are examined in terms of varying system bandwidth the specific bandwidth for a given scenario is defined as S_{BW} . The total simulation time is:

$$D_{sim} = \text{total simulation time}$$

$$S_{BW} = \text{system bandwidth.}$$

Since the suggested algorithm takes into consideration the user data stream in terms of bits we also define:

$$BA_i = i^{th} \text{ users received bit.}$$

While most available comparisons are done for throughput and fairness only in our work we introduce one more metric called ‘user ratio’. Fairness defines how fairly resources are allocated among users, while throughput defines the amount of transmitted bits. Since to compare the performance of different scheduling algorithms and to keep track of the trade-off between throughput and fairness at the same time is difficult, a new metric is proposed called ‘user ratio’. It allows us to observe and compare throughput and fairness from users’ perspective much easier. Below explicit definitions of these metrics are presented.

4.3. Fairness

This metric is the most important performance criteria in resource allocation in the uplink. This metric defines how fair the resources are allocated among users and is calculated using Jain’s Fairness Index [5]. When the available resources are going to be allocated to N users, and i^{th} users throughput is T_i the fairness can be calculated using the formula below;

$$J = \frac{(\sum_{i=1}^N T_i)^2}{N \sum_{i=1}^N T_i^2} \quad (1)$$

The parameter J varies between 1 and 0. When the value of the parameter is close to 1 this indicates highly fairness of resource allocation.

4.4. Throughput

Throughput is one of the important parameters in the evaluation of the system performance. Throughput is basically defined as the amount of transmitted bits. But in this work we are focusing on average throughput for each user. Average throughput, is the amount of successfully transmitted bits to the eNodeB for the given simulation time and it is calculated as:

$$T_{avg} = \frac{(\sum_{i=1}^N BA_i)}{D_{sim}} \quad (2)$$

4.5. User Ratio

In order to increase the efficiency of allocating resources the LTE/ LTE-A standards, define ‘flexible bandwidth allocation’.

Also depending on the number of transmission antennas used the physical layer throughput for the downlink changes from 100 Mbps, 150 Mbps and 300 Mbps for 1, 2 and 4 antenna ports respectively. In this work we consider only the single antenna case. In terms of time without changing frame or slot, when bandwidth is changed, the number of resource blocks allocated to the user also changes. The bandwidth allocation varies between 1.4 MHz – 20 MHz. For example only 1.4 MHz can be allocated or different chunks can be merged (i.e. 1.4 MHz and 20 MHz) in order to achieve higher data rate. Therefore, not only the overall system bandwidth but also the allocated bandwidth to the users will change. Furthermore, comparing the performance of different scheduling algorithms, using different channel models makes it difficult to balance and observe the trade-off between fairness and throughput. Therefore, we propose a new performance metric so called ‘user ratio’ which is calculated as given below;

$$User\ Ratio = \frac{T_i}{S_{BW}} \quad (3)$$

After dividing each user’s throughput to the available bandwidth, the ratio gives us the fairness of the resources allocated among users and allows us to compare the throughput of each user individually under different algorithms.

4.6. Proposed Algorithm

As mentioned in the previous section a good resource allocation algorithm has to balance between the user needs (fairness and high user throughput) and the system requirements (high average throughput). It also has to be efficient and simple to execute. The two most popular algorithms, RR and B-CQI stand at the two extremes: RR maximizes fairness, B-CQI maximizes average system throughput. RR is very simple and fast to execute, while B-CQI is quite sophisticated and computationally demanding. That is why, in our work we try to balance their advantages and reduce their disadvantages as much as possible.

The algorithm proposed in this work is called Fair and Efficient Resource Allocation (FERA). Our main idea is based on keeping track of the current channel conditions while allocating resources in accordance with user expectations. When channel conditions are good (high SNR) all users will be allocated some resources so it is important to utilize the situation to increase average throughput. On the other hand, when the channel conditions are generally bad, the algorithm should try to increase fairness in order not to eliminate users with low SNR and let them starve for service. To keep track of this sensitive balance between fairness and throughput, in the previous section we introduced the user ratio parameter and use as a switch in our proposed algorithm.

The FERA scheduling algorithm incorporates the advantages of RR and AMT scheduling algorithms. Following numerous simulations and analysis of different system bandwidth configurations the right proportions were determined. Our algorithm operation is TTI based. The allocation of resources is dependent on the channel quality in terms of user SNR. When the channel conditions are bad, reflecting in low SNR values, (up to 15 dB determined empirically) running the complex AMT algorithm does not provide any gains. Instead, trying to provide

some resources to all users our proposed algorithm operates similarly to the RR ensuring increase in the fairness of resource distribution. However, for high SNR values, higher than 15 dB, FERA operates similarly to AMT with the goal to maximize average throughput. As can be seen in the pseudo code of the algorithm, presented in Figure 2 below, for each TTI, first the channel conditions are evaluated and then based on the current user ratio a decision is made for which algorithm branch to follow.

Algorithm: Fair and Effective Resource Allocation Scheduling Algorithm (FERA)

```

1: calculate SNR for each UE
2: calculate user ratio for each UE
3: if ( $UE_{ratio} > 0.4$  and  $SNR > 15$ ) then
4:   order UEs by increasing max SNR
5:    $UE_{max} \leftarrow RB_N$ 
6: else
7:   if  $\sim \text{mod}(RB_{grid}, nUEs)$  then
8:      $nUEs \leftarrow RB_{grid}$ 
9:   else
10:     $\text{round}(RB_{grid}/nUEs)$ 
11:     $nUE \leftarrow RB_N$ 
12:     $RB_N \leftarrow RB_N + 1$ 
13:   end if
14: end if

```

Figure 2: Pseudo Code for FERA Algorithm

5. Simulation Environment and Scenarios

In this work, MATLAB based “LTE-A Uplink Link Level Simulator” is used [17]. The simulator allows us to observe the results of different scenarios under different network related parameters, concerning UEs and eNodeB’s (eNodeB and UE numbers, channel modelling UE mobility, etc.) and scheduling algorithms.

5.1. Considered Simulation Scenarios

In this work, three scenarios are considered. All scenarios consist of a single eNodeB located in the center of the cell, and the users are randomly distributed in the cell. Their speeds are set to pedestrian (3 km/h). The antenna configuration is selected single-input-single-output (SISO), with transmitter and receiver antennas set to one. In each scenario the bandwidth of the system varies between 5 MHz and 20 MHz. The number of users varies from 10 to 30.

In the first scenario, Typical Urban (TU) channel model is considered and the input traffic model is full-buffer. The TU channel model represents the environment with high population density, towns and cities with high-rise buildings, which translates to the highest random multiple path loss. In this scenario the varying parameters are system bandwidth and the number of users. The system bandwidth is set to 5 MHz, 10 MHz, 15 MHz and 20 MHz; the number users - 10, 20 and 30. The performance metrics considered in this scenario are fairness, throughput and user ratio.

The Rural Area (RA) channel model and full-buffer input traffic model are considered in the second scenario. This channel model represents areas with low population and building density

having the minimum path loss e.g. farms, forests and agricultural lands etc.

In the third scenario the Pedestrian channel model (PedB), standardized by ITU and specified as, environment for indoor and outdoor users, is adopted, where the indoor coverage is provided by the outdoor transmitter. This channel model is dominated by Doppler frequency component for a certain number of paths and path delays. Two different delay spreads are considered: low delay spread (A) and medium delay spread (B). In this work, the pedestrian model with medium delay spread (PedB) is used.

5.2. Simulation Parameters

The simulation parameters used are summarized in Table 1 below. To increase the accuracy of the results averages over simulations were repeated 5 simulation runs are presented.

Table 1: Simulation Parameters

Parameters	Values
Simulation Time	500 TTI
TTI Duration	1 ms
System Bandwidth	5 MHz, 10 MHz, 15 MHz, 20 MHz
Number of Available RBs	25, 50, 75 and 100
Number of Users	10, 20 and 30
Channel Models	TU, RA, PedB
Scheduling Algorithms	RR, AMT, B-CQI, FERA (Proposed Algorithm)

6. Results and Discussion

The resource allocation algorithm that is located in the eNodeB aims at maximizing the efficiency and optimizing the use of resources by providing the necessary QoS and keeping the power consumption at a minimum level.

In Table 2, Table 3 and Table 4, the comparison of the fairness values for the well-known scheduling algorithms in LTE-A which are RR, B-CQI and AMT along with the FERA scheduling algorithm are presented. Simulations were conducted with three different channel models (TU, RA and PedB), system bandwidths of 5 MHz, 20 MHz and for 10 and 30 UE numbers. As theoretically expected the RR scheduling algorithm has the highest fairness values, regardless of the system bandwidth and the number of users. The AMT scheduling algorithm gives the second highest fairness values. On the other hand, the B-CQI scheduling algorithm gives the most varying and low fairness values. The reason for this is that the algorithm allocates resources to the users according to the CQI values so when the system bandwidth is fixed and the number users increases the users with bad conditions will increase as well. Therefore, the fairness in the allocation of resources is decreased. The proposed FERA algorithm on the other hand ensures high degree of fairness in the allocation of resources among the users. It should be noticed that in certain cases it even outperforms the AMT algorithm. For

example, the fairness for the B-CQI algorithm drops for the first scenario (TU channel model) with nearly 21% when the number of users is increased from 10 to 30 for 5 MHz system bandwidth, while for the same case the fairness for the FERA drops with only 2% and for the AMT algorithm with 3%.

Table 2: Fairness Values for the TU Channel Model

Bandwidth/ UE Number	Fairness			
	RR	B-CQI	AMT	FERA
5 MHz/10 UE	0.99	0.88	0.99	0.99
5 MHz/30 UE	0.99	0.70	0.97	0.98
20 MHz/10UE	0.99	0.85	0.99	0.99
20 MHz/30UE	0.99	0.69	0.99	0.99

Table 3: Fairness Values for the RA Channel Model

Bandwidth/ UE Number	Fairness			
	RR	B-CQI	AMT	FERA
5 MHz/10 UE	0.99	0.82	0.96	0.98
5 MHz/30 UE	0.99	0.70	0.87	0.93
20 MHz/10UE	0.99	0.82	0.96	0.98
20 MHz/30UE	0.99	0.69	0.95	0.97

Table 4: Fairness Values for the PedB Channel Model

Bandwidth/ UE Number	Fairness			
	RR	B-CQI	AMT	FERA
5 MHz/10 UE	0.99	0.90	0.98	0.99
5 MHz/30 UE	0.99	0.78	0.96	0.98
20 MHz/10UE	0.99	0.88	0.99	0.99
20 MHz/30UE	0.99	0.74	0.97	0.98

It is also interesting to note that there aren't large differences between the fairness values for the TU and the PedB channel model, while all algorithms show variations in the case of the RA channel model.

The second group of results is related to the user ratio. First we present the comparative results for the 4 algorithms investigated, followed by a more detailed analysis of the user ratio values for the proposed FERA algorithm. In all the figures the x-axis gives the changing SNR values, while the y-axis gives the user ratios, i.e the performance as experienced by the individual user, resulting from the allocation in the uplink provided under the specific resource allocation algorithm.

For the FERA algorithm, the mean and the variance of the user ratio is calculated and given using the box plots below the respective graphics. These two parameters allow us to meticulously differentiate the resource allocation when large number of users is considered. As can be seen in the results, the limits that these values change are indicative of the effect of both relationships: increasing the number of users for a given bandwidth and the specific channel model used (TU, RA, PedB).

The results for the first scenario (TU channel model) for 5 MHz system bandwidth with varying number of users are given in Figure 3 through Figure 6. The comparison with RR, AMT and

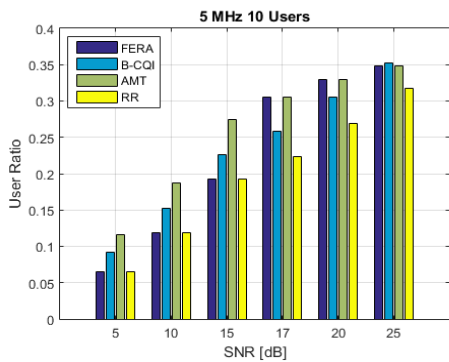


Figure 3: Scheduling Algorithms Comparison, TU Channel Model, 5 MHz system bandwidth / 10 UE

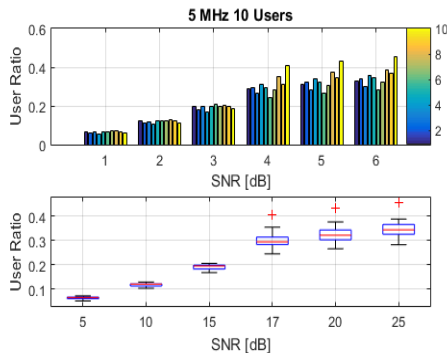


Figure 4: User Ratio, Mean and Variance for 10 UEs/Cell with TU Channel Model and 5 MHz available system bandwidth

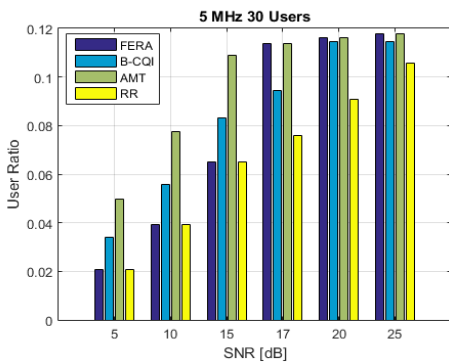


Figure 5: Scheduling Algorithms Comparison, TU Channel Model, 5 MHz system bandwidth / 30 UE

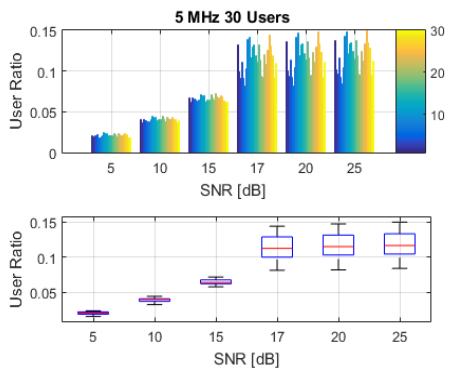


Figure 6: User Ratio, Mean and Variance for 30 UEs/Cell with TU Channel Model 5 MHz available system bandwidth

B-CQI algorithms, done based on the average user ratio, (As shown in Figure 3 and Figure 6) shows that the proposed algorithm adapts to the changing channel conditions and even

though it does not outperform for all cases provides a stable average performance with changing SNR. Furthermore, by examining the mean and variance of the user ratio for different values of the SNR we evaluate the effect of the channel on the performance that the individual user experiences. It can be seen that the variance is quite small, less than 0.01 for both 10 users and 30 users (see Figure 4 and Figure 6), which also confirms the fact that despite the changing SNR value the allocation algorithm adapts and provides the users with nearly the same performance. The proposed algorithm is both simple in terms of computational resources and quite effective. Thus, using the user ratio metric defined in section 4, allows us to more clearly trace the trade-off between system efficiency and user fairness.

Similar to the case with 5 MHz system bandwidth, simulations are carried out for 20 MHz system bandwidth and 10 and 30 users per cell respectively. Both the comparative results (see Figure 7 and Figure 10) and the in-depth user ratio analysis (see Figure 8 and Figure 10) show that FERA allocation algorithm ensures a stable average performance from the user's perspective.

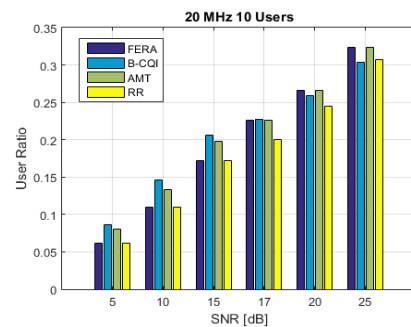


Figure 7: Scheduling Algorithms Comparison, TU Channel Model, 20 MHz system bandwidth / 10 UE

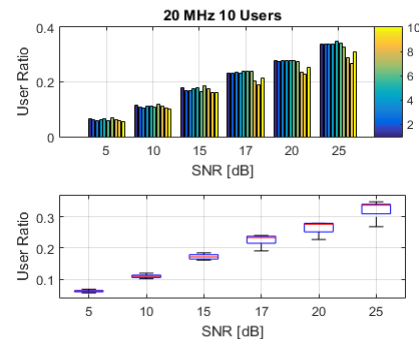


Figure 8: User Ratio, Mean and Variance for 10 UEs/Cell with TU Channel Model 20 MHz available system bandwidth

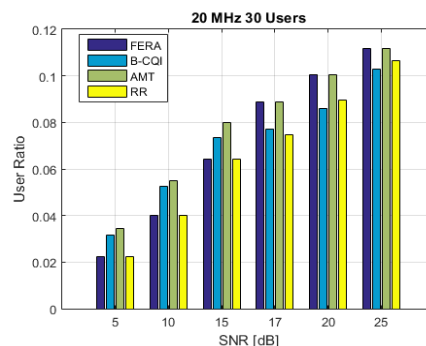


Figure 9: Scheduling Algorithms Comparison, TU Channel Model, 20 MHz system bandwidth / 30 UE

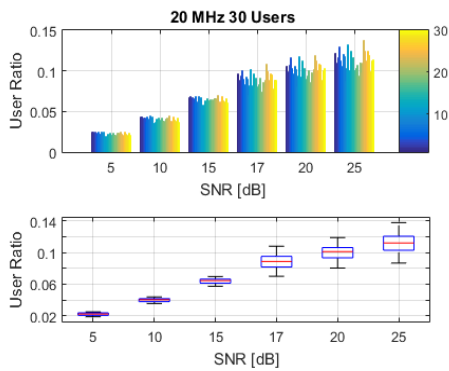


Figure 10: User Ratio, Mean and Variance for 30 UEs/Cell with TU Channel Model and 20 MHz available system bandwidth

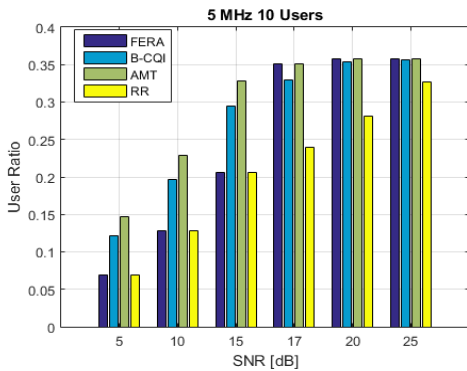


Figure 11: Scheduling Algorithms Comparison, RA Channel Model, 5 MHz system bandwidth / 10 UE

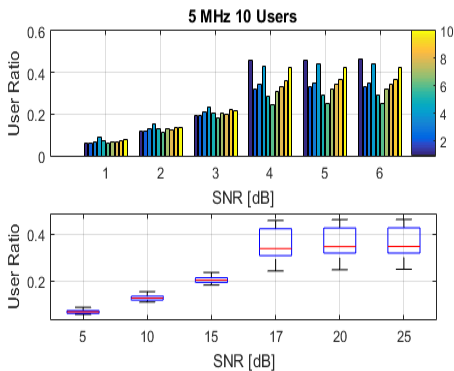


Figure 12: User Ratio, Mean and Variance for 10 UEs/Cell with RA Channel Model 5 MHz available system bandwidth

In Figure 11 through Figure 14 (for 5 MHz system bandwidth) and Figure 15 through Figure 18 (for 20 MHz system bandwidth) results from the second scenario are given, where the RA channel model is used. The simulations were conducted with system bandwidth of 5 MHz, 20 MHz with 10 UEs/cells and 30 UEs/cell respectively. The results show that, because of the bad channel conditions (low SNR values) the user ratios resulting from the FERA allocation algorithm are low, however the resource allocation among users is still fair. One important conclusion that can be made is that for bad channel conditions the variation in the user ratio is nearly negligible, while for high SNR values it is larger. (see Figure 12 and Figure 14). Yet, when the system bandwidth is increased i.e. 20 MHz instead of 5 MHz system bandwidth these variations are less expressed. On the other hand, the results show that in conditions when the SNR values are low, FERA scheduling algorithm maintains constant relation between

fairness and throughput among the users. At high SNR values it is observed that the algorithms adapt to changing conditions.

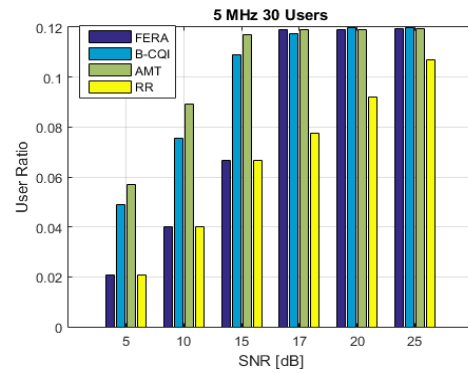


Figure 13: Scheduling Algorithms Comparison, RA Channel Model, 5 MHz system bandwidth / 30 UE

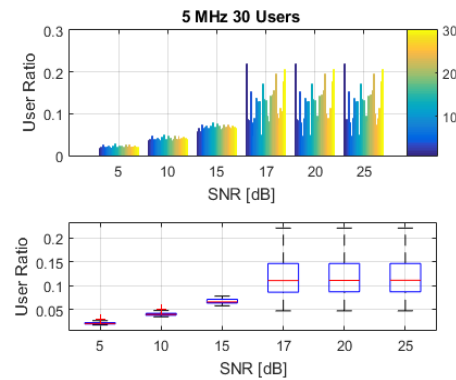


Figure 14: User Ratio, Mean and Variance for 30 UEs/Cell with RA Channel Model 5 MHz available system bandwidth

Results from the third scenario, the PedB channel model, are given in the Figure 19 through Figure 22 for the 5 MHz system bandwidth and Figure 23 through Figure 26 for the 20 MHz system bandwidth below. For low SNR values the user ratio is low because the users channel conditions are bad. But nonetheless the resource allocation among users is fair. It is observed that the user ratio drops when the system bandwidth is fixed and the number of users increase, and this drop is higher for high SNR values. The general trend that has been observed for the other scenarios is also observed here. The FERA algorithm ensures average stable performance from the user point of view.

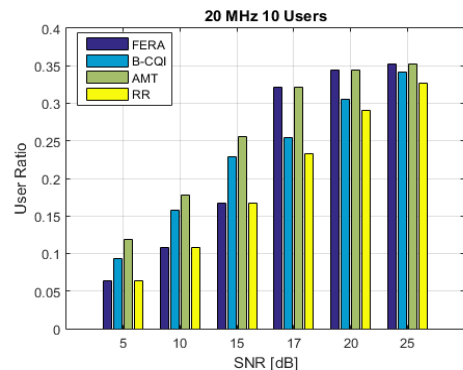


Figure 15: Scheduling Algorithms Comparison, RA Channel Model, 20 MHz system bandwidth / 10 UE

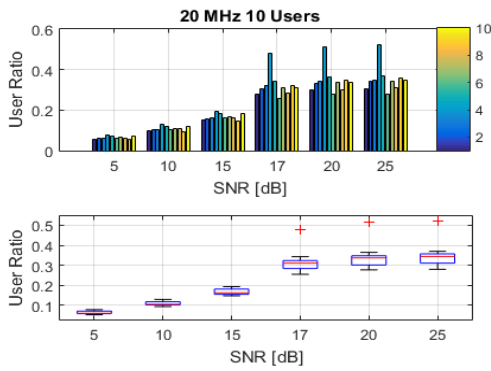


Figure 16: User Ratio, Mean and Variance for 10 UEs/Cell with RA Channel Model 20 MHz available system bandwidth

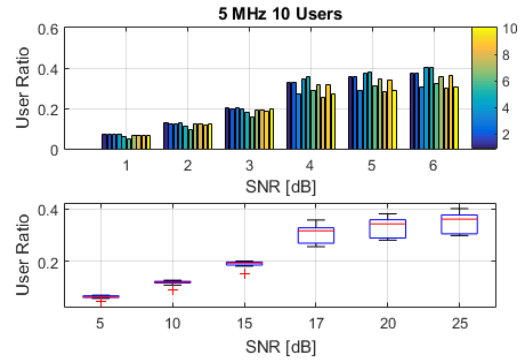


Figure 20: User Ratio, Mean and Variance for 10 UEs/Cell with PedB Channel Model 5 MHz available system bandwidth

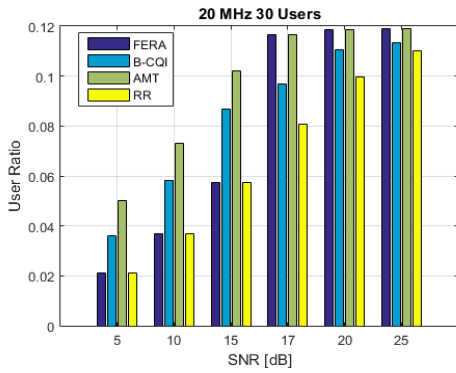


Figure 17: Scheduling Algorithms Comparison, RA Channel Model, 20 MHz system bandwidth / 30 UE

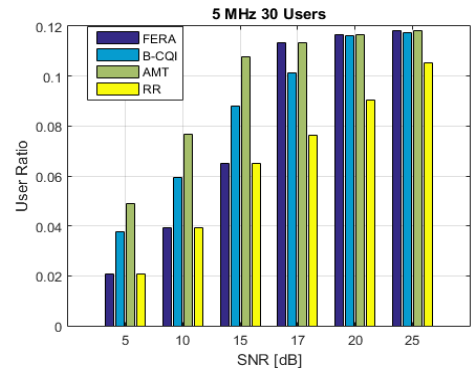


Figure 21: Scheduling Algorithms Comparison, PedB Channel Model, 5 MHz system bandwidth / 30 UE

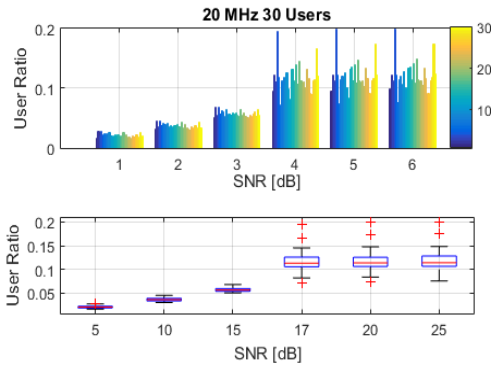


Figure 18: User Ratio, Mean and Variance for 30 UEs/Cell with RA Channel Model 20 MHz available system bandwidth

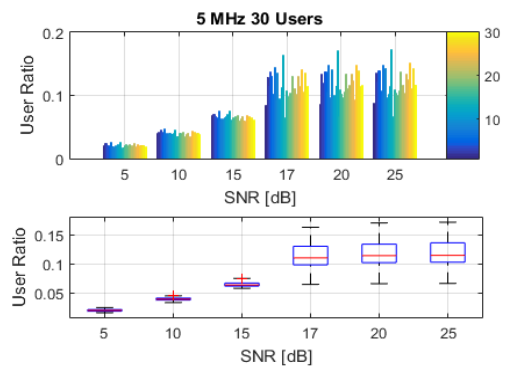


Figure 22: User Ratio, Mean and Variance for 30 UEs/Cell with PedB Channel Model 5 MHz available system bandwidth

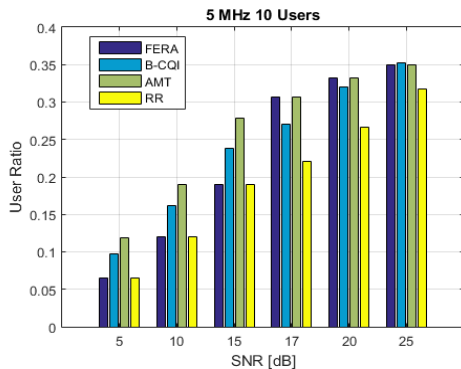


Figure 19: Scheduling Algorithms Comparison, PedB Channel Model, 5 MHz system bandwidth / 10 UE

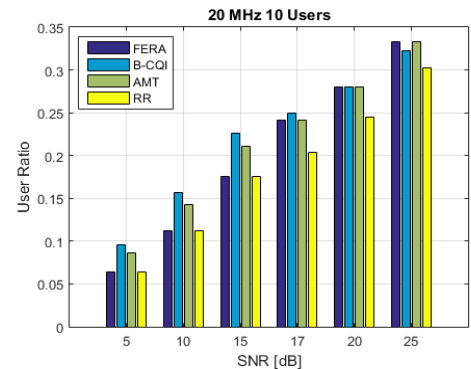


Figure 23: Scheduling Algorithms Comparison, PedB Channel Model, 20 MHz system bandwidth / 10 UE

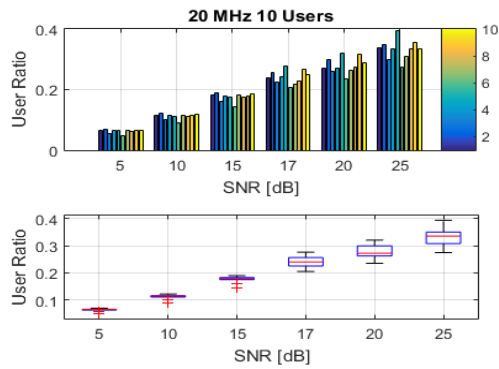


Figure 24: User Ratio, Mean and Variance for 10 UEs/Cell with PedB Channel Model 20 MHz available system bandwidth

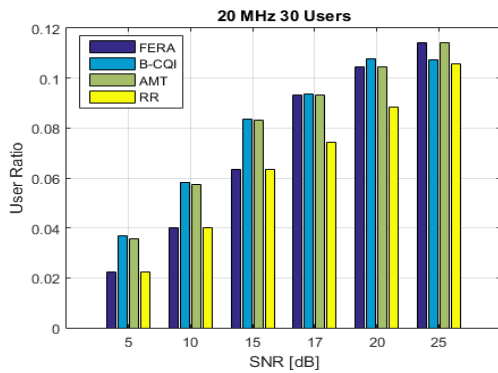


Figure 25: Scheduling Algorithms Comparison, PedB Channel Model, 20 MHz system bandwidth / 30 UE

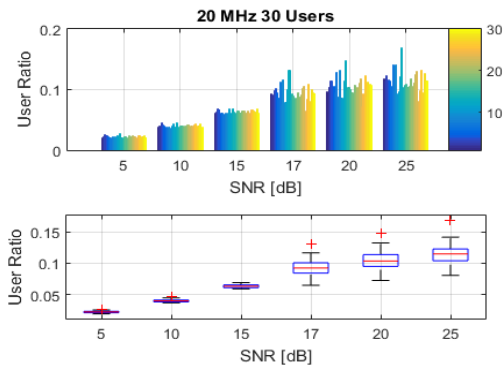


Figure 26: User Ratio, Mean and Variance for 30 UEs/Cell with PedB Channel Model 20 MHz available system bandwidth

The simulation results presented above examine the performance of the proposed FERA algorithm both in comparison with the other 3 algorithms (RR, AMT and B-CQI) and also in terms of mean and variance of the user ratio. The results show that, the variance values for all scenarios are very low, meaning that the performance experienced by the user for different channel conditions is quite stable due to the resource allocation done according to the proposed FERA algorithm.

When the simulations with TU and PedB channel models are compared, as the bandwidth increase, the simulations with PedB channel model give better results. The reason is that the TU channel model represents densely populated areas where the path loss is highest. The second important observation that can be

made is that the channel model affects the variance and the mean of the user ratio. This means that the performance experienced by the user in the uplink depends on the environment and can be partially compensated by a proper resource allocation algorithm.

Conflict of Interest

The authors declare no conflict of interest.

Conclusion

In this paper we have focused on examining different resource allocation algorithms for the uplink in LTE systems. The most popular algorithms, Round Robin, Best-CQI and Approximate Maximum Throughput are simulated, and their performance is evaluated in terms of throughput and fairness and compared with the performance of the proposed novel scheduling algorithm FERA. Furthermore, a new metric is introduced, the “user ratio”, which allows us to keep track of the trade-off between increased throughput and fairness from the point of view of the user. Three different scenarios, involving different channel models (Typical Urban, Rural area and Pedestrian) are investigated. The goal of the newly proposed algorithm FERA is to be simple and efficient while providing comparable results with respect to earlier presented algorithms. Its evaluation proves that its performance in terms of throughput and fairness is good and stable under different channel conditions. Compared to the other algorithms is it less computationally demanding, simpler but at the same time sufficiently efficient and adaptable.

References

- [1] H. Mousavi, Iraj S. Amiri, M.A. Mostafavi, and C.Y. Choon, “LTE physical layer: performance analysis and evaluation”, *Appl. Comput. Inform.*, 1-11, 2017. <https://doi.org/10.1016/j.aci.2017.09.008> [Online]
- [2] 3GPP TR 36.819, 2010, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects (Release 11)
- [3] N. Abu-Ali, A. M. Taha, M. Salah, and H. Hassanein, “Uplink scheduling in LTE and LTE-Advanced: tutorial, survey and evaluation framework”, *IEEE Comms. Surv. & Tuts.*, **16**(3), 1239-1265, 2014. <https://doi.org/10.1049/ic:19980218>
- [4] Ö. Yildiz and R. Sokullu, “A Novel Mobility Aware Downlink Scheduling Algorithm for LTE-A Networks”, in 9th International Conference on Ubiquitous and Future Networks, Milan Italy, 2017. <https://doi.org/10.1109/TELFOR.2017.8249288>
- [5] A. Kanagasabai and A. Nayak, “Opportunistic Dual Metric Scheduling for LTE Uplink”, in International Conference on Communication Workshop, London England, 2015. <https://doi.org/10.1109/ICCW.2015.7247382>
- [6] R. D. Trivedi, and M. C. Patel, “Comparison of different scheduling algorithm for LTE”, *Int. Jour. of Em. Tech. and Adv. Eng.*, **4**(5), 334-339, 2014.
- [7] M. Asvial, G. Dewandaru, and A. N. Rachman, “Modification of round robin and best CQI scheduling method for 3GPP LTE downlink”, *Int. Jour. of Tech.*, **6**(2), 130-138, 2015. <https://doi.org/10.14716/ijtech.v6i2.964>
- [8] A. Marincic, and D. Simunic, “Performance Evaluation of Different Scheduling Algorithms in LTE Systems”, in 39th International Convention on Information and Communication Technology, Electronics and Microelectronics, Opatija Croatia, 2016. <https://doi.org/10.1109/MIPRO.2016.7522211>
- [9] R. E. Ahmed, and H. M. AlMuhallabi, “Throughput-Fairness Tradeoff in LTE Uplink Scheduling Algorithms”, International Conference on Industrial Informatics and Computer Systems, Sharjah United Arab Emirates, 2016. <http://doi.org/10.1109/ICCSII.2016.7462415>
- [10] S. Schwarz, C. Mehlführer and M. Rupp, “Low Complexity Approximate Maximum Throughput Scheduling for LTE”, in Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers, Pacific Grove USA, 2010. <https://doi.org/10.1109/ACSSC.2010.5757800>
- [11] L Fatma, M. Hend and K. Hend, “Opportunistic Dual Metric scheduling based on Quality of service and Power Control for LTE Uplink Network” in 13th

- International Wireless Communications and Mobile Computing Conference, Valencia Spain, 2017. <https://doi.org/10.1109/IWCMC.2017.7986487>
- [12] C. Kong, Y. Chen and I. Peng, "Referential Bounds Analysis of Uplink Radio Resource Scheduling in LTE Network", 11th Consumer Communications and Networking Conference, Las Vegas USA, 2014. <https://doi.org/10.1109/CCNC.2014.6866620>
- [13] D. C. Dimitrova, H. van den Berg, R. Litjens and G. Heijenk "Scheduling Strategies for LTE Uplink with Flow Behavior Analysis" in 4th Ercim Workshop on E-Mobility, Lulea Sweden, 2010.
- [14] H. Safa, W. El-Hajj, and K. Tohme, "A QoS-Aware Uplink Scheduling Paradigm for LTE Networks", in International Conference on Advanced Information Networking and Applications, Barcelona Spain, 2013. <https://doi.org/10.1109/AINA.2013.38>
- [15] A. Mukhopadhyay, G. Das, V. Reddy, "A Fair Uplink Scheduling Algorithm to Achieve Higher MAC Layer Throughput in LTE", in International Conference on Communications, London England, 2015. <https://doi.org/10.1109/ICC.2015.7248803>
- [16] D. Lekomtcev, E. Kasem and R. Marsalek, "Matlab-based Simulator of Cooperative Spectrum Sensing in Real Channel Conditions" in 25th International Conference Radioelektronika, Pardubice Czech Republic, 2015. <https://doi.org/10.1109/RADIOELEK.2015.7129040>
- [17] [Online]. Available: <http://nt.tuwien.ac.at/ltesimulator>.

Computational Techniques to Recover Missing Gene Expression Data

Negin Fraidouni*, Gergely Zaruba

Electrical Engineering & Computer Science Department, Wichita State University, Wichita, KS. 67260, USA

ARTICLE INFO

Article history:

Received: 27 June, 2018

Accepted: 29 October, 2018

Online: 10 November, 2018

Keywords:

Gene Expression Prediction

Pearson Correlation Coefficient

Cosine Similarity

Robust Principal Component

Analysis

Alternating Direction Method of

Multiplier

ABSTRACT

Almost every cells in human's body contain the same number of genes so what makes them different is which genes are expressed at any time. Measuring gene expression can be done by measuring the amount of mRNA molecules. However, it is a very expensive and time consuming task. Using computational methods can help biologists to perform gene expression measurements more efficiently by providing prediction techniques based on partial measurements. In this paper we describe how we can recover a gene expression dataset by employing Euclidean distance, Pearson correlation coefficient, Cosine similarity and Robust PCA. To do this, we can assume that the gene expression data is a matrix that has missing values. In that case the rows of the matrix are different genes and columns are different subjects. In order to find missing values, we assume that the data matrix is low rank. We then used different correlation metrics to find similar genes. In another approach, we employed RPCA method to differentiate the underlying low rank matrix from the sparse noise. We used existing implementations of state-of-the-art algorithms to compare their accuracy. We describe that RPCA approach outperforms the other approaches with reaching improvement factors beyond 4.8 in mean squared error.

1. Introduction

This paper is an extension of works originally presented in ICCABS 2017 (International Conference on Computational Advances in Bio and Medical Sciences) [1] and CSCI 2017 (International Conference on Computational Science and Computational Intelligence) [2].

Almost every cell in an organism's body contain the same genetic information and series of genes, what makes cells different is which genes are expressed at any time. Gene expression is what makes a blood cell different from a liver cell and a normal healthy cell from an abnormal one (like a cancer cell) [3]. Gene expression process has two steps, transcription and translation. Transcription means a particular part of DNA is encoded into messenger RNA (mRNA) and in translation, mRNA is decoded to build a protein that contains a specific series of amino acids. We can measure the gene expression level by measuring the amount of mRNA inside the cell. Each step of the process is regulated by control points that determine the presence and the amount of proteins in any specific cell [4]. Usually a group of genes work accordingly to manage every simple or complex process that control the structure and actions of the cells. This means that group of genes must work

together in order to control structure and actions of cells [5]. Knowing this, we can conclude that the gene expression levels should be highly correlated so if the data has missing values, we might be able to predict them based on the correlation between genes.

Recently scientists have the opportunity to find the association between genes and diseases using some methods. Examples of this methods are RNA sequencing, northern blotting, western blotting, DNA microarray, fluorescent in situ hybridization and reporter gene. However the costs of measuring gene expression levels are extremely high and also the complete process needs a huge amount of time [6] which makes it difficult to measure and access this information. Also gene expression data usually suffers from missing values. This can happen due to some reasons like failures in hybridization, noise in data and also data corruption. Missing values in gene expression data can negatively affect gene disease studies [7]. Since due to the huge amount of time and money needed for repeating the measurements, an alternative way is to use computational methods which can be employed to predict the missing values and recover the dataset. So there is a high demand for novel techniques to find the missing values.

The first step to make our model is to store gene expression data in matrices where each row is a different gene, each column

*Negin Fraidouni, Department of Computer Science, The University of Texas at Arlington, Arlington, TX. 76010, USA. Email: neginfraidouni@gmail.com

is a different disease and the entries of the matrix are the gene expression values (mRNA measurements) of corresponding rows and columns [8]. Based on this model we can assume that people with similar diseases show similar expression patterns so the gene expression data matrix must be highly overdetermined and significantly low rank [9].

1.1. Recommendation system

Recommendation systems rely on information filtering in order to deal with data overload by filtering necessary information which is significantly less than total data of users' preferences, interests or observations about an item. Recommender systems have the ability to recommend a new movie to a specific user based on his/her previous preferences [10]. Recently different designs for recommendation systems have been proposed which are based on one of these methods: collaborative filtering method [11], content-based filtering method [12] and hybrid filtering method [13].

1.2. Collaborative Filtering Method

Collaborative filtering method recommends item to users by recognizing users with similar tastes. It combines other ratings in order to recommend new items to each specific user. Collaborative filtering techniques are categorized into two groups: model-based technique and memory-based technique. The main goal for model based method is making a model and extract the necessary part of the data matrix so there is no need to use the entire dataset in order to make predictions [14]. Memory based technique uses previously collected data in order to predict the missing ratings and they use the entire user-item database. The common memory based method is based on nearest neighbors and uses a distance measure metric to find the neighbors [15]. This is also called neighborhood based approach which similar users are grouped together based on their interests [16].

Netflix is an example of recommendation systems that can benefit from collaborative filtering technique. For the Netflix example, the proposed model contains a $m * n$ matrix. Each row of the matrix corresponds to a different user and each column corresponds to a different movie and the entries of the matrix are the ratings users gave to the movies. The data matrix is very sparse because most users usually tend to rate a very small fraction of the movies, the matrix is very sparse. So the goal is to find the hidden pattern and predict the missing values in order to make a recommendation to users for the movies that they have not watched yet.

1.3. Low Rank Matrix Completion

Matrix completion (MC) involves recovering an incomplete matrix where only a small fraction of its entries are known which is significantly smaller than the total size of the matrix. Low rank MC problem can be seen in different practical contexts such as image processing [17], machine learning [18] and bioinformatics [19]. To solve this problem, we should find the lowest rank matrix which is consistent with the known values of the incomplete matrix. We can write:

$$\begin{aligned} &\text{minimize rank } (Y) \\ &\text{such that } R_\omega(Y) = R_\omega(X) \end{aligned} \tag{1}$$

Here X is the incomplete matrix that we want to reconstruct, ω shows the known values such that $(a,b) \in \omega$ if $X_{a,b}$ is known. R_ω is the orthogonal projection matrix where:

$$|R_\omega(X)|_{a,b} = \begin{cases} X, & (a,b) \in \omega \\ 0, & (a,b) \notin \omega \end{cases} \tag{2}$$

Because the rank minimization problem is NP-hard, this problem can be remodeled as minimizing trace norm or nuclear norm. Nuclear norm is the sum of singular values of the given matrix [20]. The reason is that a rank r matrix with has exactly r singular values which are greater than zero. The nuclear norm of matrix Z is defined as:

$$\|Z\|_* = \sum_{a=1}^r \sigma_a \tag{3}$$

Where:

σ_a is the a^{th} singular value (nonzero) of matrix Z and r is the rank of matrix Z .

So we can rewrite the problem as:

$$\begin{aligned} &\text{minimize } (\|Y\|_*) \\ &\text{such that } R_\omega(Y) = R_\omega(X) \end{aligned} \tag{4}$$

The advantage of using nuclear norm over minimizing the rank is that its optimum point can be calculated efficiently and it is convex.

1.4. Robust PCA method (RPCA)

One problem with gene expression datasets is the presence of noise in expression measurements. This happens because of some reasons like different degrees of uniformity, small spots, process errors and also inconsistency in hybridization.

For the aim of showing the most variability of the data for a noise free dataset, we can easily perform PCA using SVD (singular value decomposition). In the presence of noise we can use RPCA in order to reconstruct a low rank matrix and find the sparse noise. Assume that our data matrix E is decomposed as:

$$E = Y + S \tag{5}$$

Where Y is the underlying low rank matrix and S is a sparse matrix capturing noise. Because the number of unknowns to infer for Y and S is considerably higher than known values in E , this problem is overdetermined. So we need to use tractable convex optimization as denoted by:

$$\begin{aligned} &\text{minimize } \|Y\|_* + \lambda \|S\|_1 \\ &\text{subject to } Y + S = E \end{aligned} \tag{6}$$

Where $\|S\|_1 = \sum_{i,j} |S_{i,j}|$ is the ℓ_1 -norm of S and λ is a parameter. This should work even in the situations when the rank of Y is not low rank (when rank is equal to the dimension of the matrix). For RPCA method to work efficiently, we need to know the location of the non-zero entries in matrix S . Problem (6) can

be solved at a cost not so much higher than classical PCA [21]. One of the methods that can be employed here is the Alternating Direction Method of Multipliers (ADMM) that we will summarize in the next section.

1.5. The ADMM method

The ADMM method is a powerful method because it mixes two methods of multipliers and dual ascent. The algorithm solves the problems in the form:

$$\begin{aligned} \min \quad & f(a) + g(b) \\ \text{such that} \quad & Xa + Yb = z \end{aligned} \tag{7}$$

Where f and g are both convex. The optimal value for the problem above is defined as:

$$p^* = \inf\{f(a) + g(b) \mid Xa + Yb = z\} \tag{8}$$

Here the augmented Lagrangian is:

$$\begin{aligned} L_\rho(a, b, m) = & f(a) + g(b) + m^T (Xa + Yb - z) + \\ & \frac{\rho}{2} \|Xa + Yb - z\|_F^2 \end{aligned} \tag{9}$$

Where m is the Lagrangian multiplier and $\rho > 0$ is a parameter. ADMM method consists of the multiple iterations as denoted below:

$$\begin{aligned} a^{k+1} &= \arg_a \min L_\rho(a^k, b^k, m^k) \\ b^{k+1} &= \arg_b \min L_\rho(a^{k+1}, b^k, m^k) \\ m^{k+1} &= m^k + \rho(Xa^{k+1} + Yb^{k+1} - z) \end{aligned} \tag{10}$$

The algorithm consists of multiple steps:

1. An a-minimizing step
2. A b-minimizing step and
3. A variable update.

In the last step (variable update step) the step size is equal to the m (the augmented Lagrangian parameter).

2. Methods

2.1. Correlation based Matrix Completion method

The main goal of the correlation based matrix completion method (CMC) is finding correlation between genes (neighbors). To predict a value for a missing entries, we need to consider all other subjects' expression values. If a gene is more similar to the one with a missing value, its expression value has more impact on the predicted value. For finding correlation between genes, we will use Pearson correlation coefficient (PCC), Euclidean distance (ED) and Cosine similarity (CS).

Pearson Correlation Coefficient

PCC is a common measure of linear dependency between two variables. The PCC can take any value from 1- (means negative

association) to 1 (means positive association) with 0 indicating orthogonality. The PCC can be calculated by:

$$PCC = \frac{\sum_{k=1}^m (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^m (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^m (y_k - \bar{y})^2}} \tag{11}$$

Or:

$$PCC = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{12}$$

Where X and Y are two datasets.

Euclidean Distance

Euclidean distance (ED) is a metric that if x and y have zero distance, then $x = y$ holds. ED between two points (x and y) can be calculated by:

$$ED = |x - y| = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \tag{13}$$

Cosine Similarity

The Cosine Similarity (CS) is a metric of the cosine of the angle between two vectors. This is a measurement of orientation instead of magnitude. Like PCC, CS can take any value from 1- (means negative association) to 1 (means positive association) with 0 indicating orthogonality. The CS value can be calculated as shown below:

$$\cos\theta = \frac{\bar{x} \cdot \bar{y}}{\|\bar{x}\| \cdot \|\bar{y}\|} = \frac{\sum_{k=1}^m x_k y_k}{\sqrt{\sum_{k=1}^m x_k^2} \sqrt{\sum_{k=1}^m y_k^2}} \tag{14}$$

And the dot product:

$$\bar{x} \cdot \bar{y} = \|\bar{x}\| \cdot \|\bar{y}\| \cdot \cos\theta \tag{15}$$

CMC Approach

Here we explain how CMC approach works in order to find missing values in partially known matrices. Let us assume that we have a complete (means all the entries are known) low rank matrix Y . Now if we randomly remove some of the values, the problem becomes how to find missing values in a way that there was not a huge difference between the original values and the predictions. We will use PCC, ED and CS in order to reconstruct matrix X and the reconstructed matrices when using each of the aforementioned similarity metrics are Y_P , Y_E and Y_C respectively.

This is how the CMC method works:

- First a PCC, an ED and a CS value should be calculated for each pair of the genes.
- When there is a missing value, we need to calculate the mean of all known entries in that column weighted by how correlated they are (based on their PCC, ED and CS values).
- Finally we will measure the accuracy of each method by calculating the error between the reconstructed matrices and the original matrix.

For finding a missing value at $Y(m,n)$ we will calculate:

$$Y_P(m,n) = \frac{\sum_{r=1}^K PCC'(m,n,r) * y_{r,n}''}{\sum_{r=1}^N x(r,n)} \quad (16)$$

$$Y_E(m,n) = \frac{\sum_{r=1}^K ED'(m,n,r) * y_{r,n}''}{\sum_{r=1}^K x(r,n)} \quad (17)$$

$$Y_C(m,n) = \frac{\sum_{r=1}^K CS'(m,n,r) * y_{r,n}''}{\sum_{r=1}^K x(r,n)} \quad (18)$$

Where $r! = m$ is matrix Y 's row, and:

$$PCC'_{r,n} = \begin{cases} PCC(m,r) & (r,n) \in \omega \\ 0 & (r,n) \notin \omega \end{cases} \quad (19)$$

$$ED'_{r,n} = \begin{cases} ED(m,r) & (r,n) \in \omega \\ 0 & (r,n) \notin \omega \end{cases} \quad (20)$$

$$CS'_{r,n} = \begin{cases} CS(m,r) & (r,n) \in \omega \\ 0 & (r,n) \notin \omega \end{cases} \quad (21)$$

$$y''_{r,n} = \begin{cases} y'_{r,n} & (r,n) \in \omega \\ 0 & (r,n) \notin \omega \end{cases} \quad (22)$$

$$x(r,n) = \begin{cases} 1 & (r,n) \in \omega \\ 0 & (r,n) \notin \omega \end{cases} \quad (23)$$

Algorithm 1 shows the pseudo code for correlation based matrix completion approach.

Algorithm 1. Correlation based matrix completion approach

```

Input: Y, ω
For row m of Y:
  For row n of Y:
    If m != n :
```

```

  Find PCC (m,n)
  Find ED(m,n)
  Find CS (m,n)
```

```

For row m of Y:
  For row r of Y:
    If (m,r) in ω:
      Y_p(m,r) = Y(m,r)
      Y_e(m,r) = Y(m,r)
      Y_c(m,r) = Y(m,r)
    Else if (m,r) not in ω:
      Y_p(m,r) = 0
      Y_e(m,r) = 0
      Y_c(m,r) = 0
      x = 0
    for row p of Y:
      if (p,n) in ω:
        Y_p(m,r) += Y_p(r,n) * PCC(m,p)
        Y_e(m,r) += Y_e(r,n) * ED(m,p)
        Y_c(m,r) += Y_c(r,n) * CS(m,p)
      x++
    Y_p(m,r) /= x
    Y_e(m,r) /= x
    Y_c(m,r) /= x
output: Y_P, Y_E, Y_C
```

2.2. Convex Optimization Formulation of RPCA method

The ADMM formulation of our RPCA model (6) is defined by:

$$L_\rho(Y,S,m) = \|Y\| + \lambda \|R_\omega(S)\| + m^T (R_\omega(E - Y - S)) + \frac{\rho}{2} \|R_\omega(E - Y - S)\|_F^2 \quad (24)$$

In each iteration we repeat:

Updating matrix Y

We update Y by:

$$Y^k = \min_Y \|Y\| + \frac{\rho}{2} \|R_\omega(E - S^{k-1} - Y^{k-1})\|_F^2 + m^T (R_\omega(E - S^{k-1} - Y^{k-1})) \quad (25)$$

Which can be rewritten as:

$$\min_Y \|Y\| + \frac{\rho}{2} \|P_\Omega(Y^k + S^k - E) - \frac{m}{\rho}\|_F^2 \quad (26)$$

For solving above problem, we can use a soft thresholding operation from [22]. So the problem would become:

$$Y^k = \text{shrink}(A^{k-1}, \rho^{-1}) \quad (27)$$

$$A^{k-1} = (E - S^{k-1} + \frac{m}{\rho^{k-1}})$$

Where ρ^{-1} is the step size which decreases the singular values of matrix A and the shrink is a soft-thresholding operator and can be defined as:

$$\begin{aligned} shrink(M, \tau) &:= \sum_{k=1}^r u_k \max(\sigma_k - \tau, 0) v_k^T \\ M &= \sum_{k=1}^r (u_k \sigma_k v_k^T) \end{aligned} \tag{28}$$

Where σ_k is the singular values and u_k is the left and v_k is the right singular vectors of matrix M.

Updating matrix S:

After updating Y, we can update S through:

$$\begin{aligned} S^k &= \min_S \lambda \|R_\omega(S^{k-1})\| + m^T (R_\omega(E - S^{k-1} - Y^k)) \\ &+ \frac{\rho}{2} \|R_\omega(E - S^{k-1} - Y^k)\|_F^2 \end{aligned} \tag{29}$$

Which can be rewritten as:

$$\min_S \lambda \|R_\omega(S^{k-1})\| + \frac{\rho}{2} \|R_\omega(Y^k - E + S^{k-1}) - \frac{m}{\rho}\|_F^2 \tag{30}$$

To solve the above problem, we can use a shrinkage operator:

$$\begin{cases} S_{ij} = H_{\frac{\lambda}{\rho}}(E - Y^k + \frac{m}{\rho}) & (i, j) \in \omega \\ S_{ij} = 0, & (i, j) \notin \omega \end{cases} \tag{31}$$

Where $H_{\frac{\lambda}{\rho}}$ is the shrinkage operator discussed in [23] and can be calculated by:

$$H_\sigma(S_{ij}) = \begin{cases} S_{ij} - \sigma, & S_{ij} > \sigma \\ S_{ij} + \sigma, & S_{ij} < -\sigma \\ 0 & \text{Otherwise} \end{cases} \tag{32}$$

We can assume that entries in matrix S that represent missing values are equal to zero.

Updating m:

After updating Y and S, we can update m by:

$$m^k = m^{k-1} + \rho(E - Y^k - S^k) \tag{33}$$

Algorithm 2 shows the pseudo code for solving RPCA problem using ADMM.

Algorithm 2. Solving RPCA problem using ADMM

Input: E, ρ , λ , ε
While $\|E - Y^k - S^k\|_F > \varepsilon$:

Updating matrix Y:

$Y^k = \arg_Y \min L_\rho(Y^{k-1}, S^{k-1}, m^{k-1})$
 $Y^k = shrink((E - S^{k-1} + \frac{m}{\rho}), \rho^{-1})$
 $(U, S, V) = SVD(E - S^{k-1} + \frac{m}{\rho})$
 For singular values σ in S:
 If $\sigma < \frac{1}{\rho}$:
 $\sigma = 0$
 $Y^k = U S V^T$

Updating matrix S:

$S^{k+1} = \arg_S \min L_\rho(Y^{k+1}, S^k, m^k)$
 for row p of S:
 for column r of S:
 if (p,r) in ω :
 $S_{pr} = H_{\frac{\lambda}{\rho}}(E - Y^k + \frac{m}{\rho})$
 else:
 $S_{pr} = 0$

Updating m:

$m^k = m^{k-1} + \rho(E - S^k - Y^k)$

output: Y^k and S^k

3. Competitive methods

3.1. K-nearest Neighbors method

K-nearest neighbors (KNN) is one of the most essential classification algorithms in machine learning. It can be widely used in real-life scenarios since it does not make any assumption about the distribution of the data. The model representation for KNN is the entire dataset and it can make predictions using the training data set directly. When there is a missing value, prediction can be made by searching through the dataset for the K most similar neighbors and the result is the weighted average of those neighbors [24]. To determine which K neighbors are the most similar ones, a distance measure should be used and Euclidean distance is the most popular one for real-valued variables.

3.2. Nuclear Norm Minimization

There are various numerical methods available to solve(4). The important problem is that because of the high dimensionality aspect of biological data, many numerical methods fail to solve the problem efficiently. Kapur et al. [25] used a method called soft thresholding operator which can scale well on large datasets. So the problem would become:

$$\begin{aligned} \min \tau \|Y\|_* + \frac{1}{2} \|Y\|_F \\ \text{such that } R_\omega(Y) = R_\omega(X) \end{aligned} \tag{34}$$

Where $\|Z\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |z_{i,j}|^2}$ is the Frobenius norm and τ is the thresholding parameter and it should be greater than 0. We can reconstruct the expression matrix iteratively so the kth iteration would be:

$$\begin{aligned} Y^k &= shrink(M^{k-1}, \tau) \\ M^k &= M^{k-1} + \delta_k R_\omega(X - Y^k) \end{aligned} \tag{35}$$

Where shrink is the soft thresholding operator [22]. The parameter δ_k is the step size and the parameter τ minimizes the rank by decreasing the singular values. The shrink operator is defined by:

$$\text{shrink}(M, \tau) := \sum_{i=1}^r \max(\sigma_i - \tau, 0) u_i v_i^T$$

$$M = \sum_{i=1}^r \sigma_i u_i v_i^T \tag{36}$$

Where u_i is the left and v_i is the right singular vectors of data matrix M . In each iteration the SVD of matrix M is calculated and those singular values that are smaller than τ parameter, will be set to zero. The new matrix M will be reconstructed. Algorithm 3 shows the pseudo code for this method.

Algorithm 3. Nuclear Norm Minimization Problem

```

Input: Y, ω, ε
δ = 1.2 * (mn) / |ω|
τ = 5 * (mn)0.5
Shrink(Y, τ)
  (U,S,V) = SVD (Y)
  For singular values σ in S:
    If σ < τ:
      σ = 0
  M = U S VT

Minimize(Y, ω)
  For row a of Y:
    For column b of Y:
      If (a,b) in ω:
        Rω (Y) = Y
      Else:
        Rω (Y) = 0
  M0 = 0
  k = 1
  while ||Rω (Yk - X)||F / ||Rω (X)||F < ε:
    Yk = shrink (Yk-1, τ)
    Mk = Mk-1 + Rω (X - Yk)
    k++
output: Xk
    
```

3.3. Singular Value Thresholding Algorithm (SVT)

This approach considers using a Robust PCA approach in order to reconstruct a low rank matrix from noisy measurements.

$$\min_{A,E} \lambda \|E\|_1 + \|A\|_* + 2\tau^{-1} \|E + A\|_F^2$$

such that: $D = E + A$ (37)

Where:

- D is the noisy dataset.
- A is the low rank matrix.
- E is the noise and it assumed that it only affect a fraction of the data (E is sparse).
- τ is a scalar and $\tau > 0$.

We can apply the Lagrangian multiplier Y in order to replace the equality constraint:

$$L(E, A, Y) = \lambda \|E\|_1 + \|A\|_* + 2\tau^{-1} \|E + A\|_F^2 + \frac{1}{\tau} \langle Y, D - A - E \rangle \tag{38}$$

Then in each iteration A , E and Y will be updated by minimizing (38) with respect to A , E and Y . Algorithm 4. Shows the pseudo code for SVT approach where σ is the step size.

Algorithm 4. The SVT Algorithm

```

Input: τ, D, λ
While not converged:
  (U,S,V) = SVD(Yk)
  Ak = arg minA τ ||D||_* + 1/2 ||D - Ek-1}||F
  Ek = arg minE τ ||D||_1 + 1/2 ||D - Ak}||F
  Yk = Yk-1 + σk (D - Ak - Ek)
End while
output: A = Ak, E = Ek
    
```

3.4. Exact Augmented Lagrangian Multiplier (ELAM)

ELAM method was proposed in [26] and can be used for solving Robust PCA problem. To solve the problem can apply the Lagrangian multiplier as denoted below:

$$X = (A, E)$$

$$f(X) = \|A\|_* + \lambda \|E\|_1 \tag{39}$$

$$h(X) = D - A - E$$

And the Lagrangian function is:

$$L(A, E, Y) = \|A\|_* + \lambda \|E\|_1 + \frac{\mu}{2} \|D - A - E\|_F^2 + Y^T (D - A - E) \tag{40}$$

Algorithm 5. Shows the pseudo code for ELAM method.

Algorithm 5. Exact Augmented Lagrangian Multiplier (ELAM)

```

Input: matrix D, λ
While not converged:
  (Ak+1, Ek+1) = arg minA,E L(A,E,Y)
  While not converged:
    (U,S,V) = SVD(D - Ek + Y/μ)
    Ak+1 = arg minA 1/μ ||D||_* + 1/2 ||D - Ek}||F
    Ek+1 = arg minE 1/μ ||D||_1 + 1/2 ||D - Ak+1 + Y/μ||F
  End while
  Yk+1 = Yk + μk (D - Ak+1 - Ek+1)
  k++
End while
output: A = Ak+1, E = Ek+1
    
```

3.5. Inexact Augmented Lagrangian Multiplier (ILAM)

ILAM method was proposed in a study by Lin et al [27]. The RPCA problem is closely connected to MC problem so the MC can be formulated as:

$$\begin{aligned} & \text{minimize } \|A\|_* \\ & \text{such that } E + A = D, R_\omega(E) = 0 \end{aligned} \quad (41)$$

$R_\omega(E) = 0$ means that E is zero at indices where the value is known and the augmented Lagrangian is:

$$\begin{aligned} L(E, A, Y) = & \|A\|_* + \frac{\mu}{2} \|D - E - A\|_F^2 \\ & + Y^T (D - E - A) \end{aligned} \quad (42)$$

So the ILAM approach can be used for MC problem. The pseudo code for ILAM method is described below.

Algorithm 6. Exact Augmented Lagrangian Multiplier (ELAM)

```

Input: matrix D, ω, λ
While not converged:
    (Ak, Ek) = arg minA, E L(E, A, Y)
    (U, S, V) = SVD(D - Ek +  $\frac{Y}{\mu}$ )
    Ak = arg minA  $\frac{1}{\mu} \|D\|_* + \frac{1}{2} \|D - E^{k-1} - A\|_F^2$ 
    Ek = arg minE  $\frac{\lambda}{\mu} \|D\|_1 + \frac{1}{2} \|D - A^k + \frac{Y}{\mu}\|_F^2$ 
    Rω(E) = 0
    Yk = μk(D - Ek - Ak) + Yk-1
    k++
End while
output: E = Ek, A = Ak

```

4. Evaluation

Here we measure the accuracy of different approaches as they apply to biomedical data MC. All of the following experiments were performed using Python 3.6 and Matlab 2016 on an Intel Core i7 PC running Windows 10 with 16GB main memory.

4.1. Datasets

We downloaded and used 4 gene expression datasets from NCBI (National Center for Biotechnology Information) for our experiments. We used the following gene expression datasets:

Lung Cancer Study

Title of this study is: "Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival". Scientists found that smoking tobacco is the reason for the most of lung cancer cases, but the exact details of this process is still unknown. In this study Landi et al. used 135 tissue samples of adenocarcinoma and non-involved lung tissue from 3 groups (current, former and never smokers). They found out that

www.astesj.com

expression of some genes is significantly different in smokers and non-smokers. The lung cancer dataset has 22283 rows- and 107 columns [28].

Dementia Study

Title of this study is: "Variations in the progranulin gene affect global gene expression in frontotemporal lobar degeneration". The symptoms of frontotemporal lobar degeneration is progressive decline in language and function. Despite the excessive research on the reason for this disease, its mechanisms remain unknown. Plotkin et al. isolated postmortem brain samples from normal controls, patients with mutations in progranulin gene and patients without mutations in progranulin gene. The dementia dataset has 22277 rows and 56 columns [29].

Autism Study

Title of this study is: "Autism and increased paternal age related changes in global levels of gene expression regulation". Autism is a neurodevelopmental disorder and it is the results of transcription factor mutations that can change the gene expression regulation. In this study Alter et al. analyzed gene expression values of 82 subjects with autism and 64 controls. The results showed that autism and increased paternal age can change the gene expression regulation. The Autism dataset has 54613 rows and 146 columns [30].

Bladder Cancer Study

Title of this study is: "Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer". In this study Riester et al. analyzed data of patients with bladder cancer (n = 93) and measured the gene expression. The bladder cancer dataset has 54675 rows and 93 columns [31].

4.2. Calculating Error

We used two metrics to determine how accurate the MC algorithm is. So we start with a known matrix Y , remove a random portion of it (i.e., simulating missing entries), and then try to reconstruct the matrix Y' .

Relative Error

Relative error can be used to describe accuracy; specifically, how accurate a measurement is compared to the true value. We use the relative error (RE) which can be calculated as below:

$$\text{Relative Error} = \frac{\|Y - Y'\|_F}{\|Y\|_F} \quad (43)$$

Where Y is the original matrix and Y' is the reconstructed matrix.

Mean Square Error

We will also use mean squared error (MSE) to measure the accuracy of the different approaches. We can find MSE by calculating the mean of the squares of the deviations, which is the

Y and Y' is:

$$\text{Mean Squared Error} = \frac{\|Y - Y'\|_F}{n_1 * n_2} \quad (44)$$

5. Results

We used the complete datasets as starting points for all experiments and removed a random set of values from the data matrices. The resulting incomplete matrices were then used as inputs to the algorithms to predict the unknowns (missing values). In order to measure the accuracy of the different methods, we used two metrics to compare the predicted and the original matrices. We will evaluate the performance of the methods when the degree of missing values changes. To do this, for each dataset we made nine incomplete matrices with 10% to 90% missing values. For each matrix with varied proportion of missing values, we employed the algorithm 10 times so in figures 1, 2 and 3, each data-point shows an average of ten different experiments with randomly removed values from the original matrix.

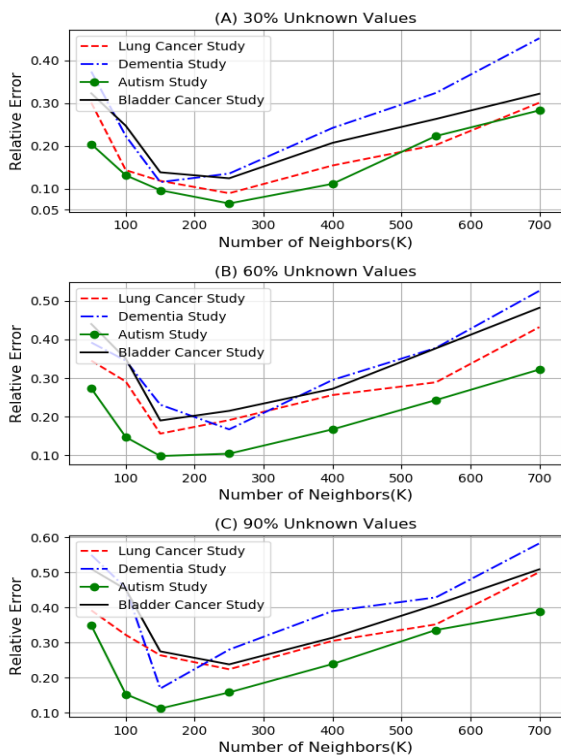


Figure 1. Comparison of relative errors for different values of k for KNN method.

For the aim of finding the best k for KNN method, we varied the value of k from the list 50, 100, 150, 250, 400, 550, 700 and calculated the relative error in cases where 30%, 60% and 90% of the values are unknown. As Figure 1 shows, when the value is around 150-250, the relative error is the least so we selected 200 as the number of neighbors. The results of the comparison between KNN algorithm and PCC- CMC method are summarized

in table 1. KNN method is extremely slow when the size of the dataset is large.

Table 1. Comparison of relative error averages for 4 datasets of KNN method and PCC-CMC

	KNN (k = 200)	PCC-CMC
30% Unknowns	0.103	0.047
60% Unknowns	0.169	0.069
90% Unknowns	0.225	0.102

We compared the performance of nuclear norm minimization to PCC, ED and CS based correlation approach on 4 NCBI-GEO datasets (Dementia, Autism, Lung cancer and Bladder cancer) and results are displayed in figure 2. The horizontal axis of all graphs (A - H) represents the ratio of the missing entries. The vertical axis of graphs A, C, E and G represents the relative error (Eq. 43) and the vertical axis in graphs B, D, F and H shows mean squared error (Eq. 44). The performance of PCC-CMC approach is shown by the dotted red line while the black, the dotted blue and the green lines depict the performances of the ED-CMC, CS-CMC and nuclear norm minimization approaches respectively. As the figure 2. Shows, for all four datasets, the PCC-CMC approach consistently beats the nuclear norm minimization approach. The nuclear norm minimization (green line) represents an increasingly growing error but the error of the PCC-CMC approach (red line) shows a decreasing acceleration. The CS-CMC (blue line) also represents improvements when compared to nuclear norm minimization but the ED-CMC approach does not show any improvements. The relative error and mean squared error of PCC-CMC grew very much slower than that of the nuclear norm minimization approach in cases of an increase in the ratio of missing entries. We can explain the different results of PCC-CMC and CS-CMC by this hypothesis that PCC might be better in catching the genes correlation compared to CS.

Based on our results, The PCC-CMC approach has higher accuracy compared to other approaches. In the case of 90% missing values for relative error, in the best case, PCC-CMC outperforms CS-CMC, ED-CMC, and Nuclear norm minimization by a factor of 1.4, 2 and 2.2 respectively. In the worst case PCC-CMC outperforms CS-CMC, ED-CMC, and Nuclear norm minimization by a factor of 1.2, 1.6 and 1.7 respectively. When looking at MSE, PCC-CMC outperforms the other three approaches by as much as a factor of 1.7, 2.3 and 2.4 respectively (for the same order as previously) and in the worst case we get an improvement factor of 1.1, 1.3 and 1.4 respectively.

We then compared the performance of the featured ADMM approach to SVT, ELAM, ILAM and PCC-CMC approach for our 4 datasets and the results are presented in figure 3. The horizontal axis in all graphs (A - H) represents the ratio of the missing entries. The vertical axis in graphs A, C, E and G shows the relative error and the vertical axis in graphs B, D, F and H shows mean squared error. The performance of the featured ADMM approach is shown by the dotted red line while the black, the dotted blue, the green and the gray lines represent the performances of the SVT, ELAM, ILAM and PCC-CMC approaches respectively. As Figure 3. Shows, for all four datasets, the ADMM approach beats the other four approaches. The black,

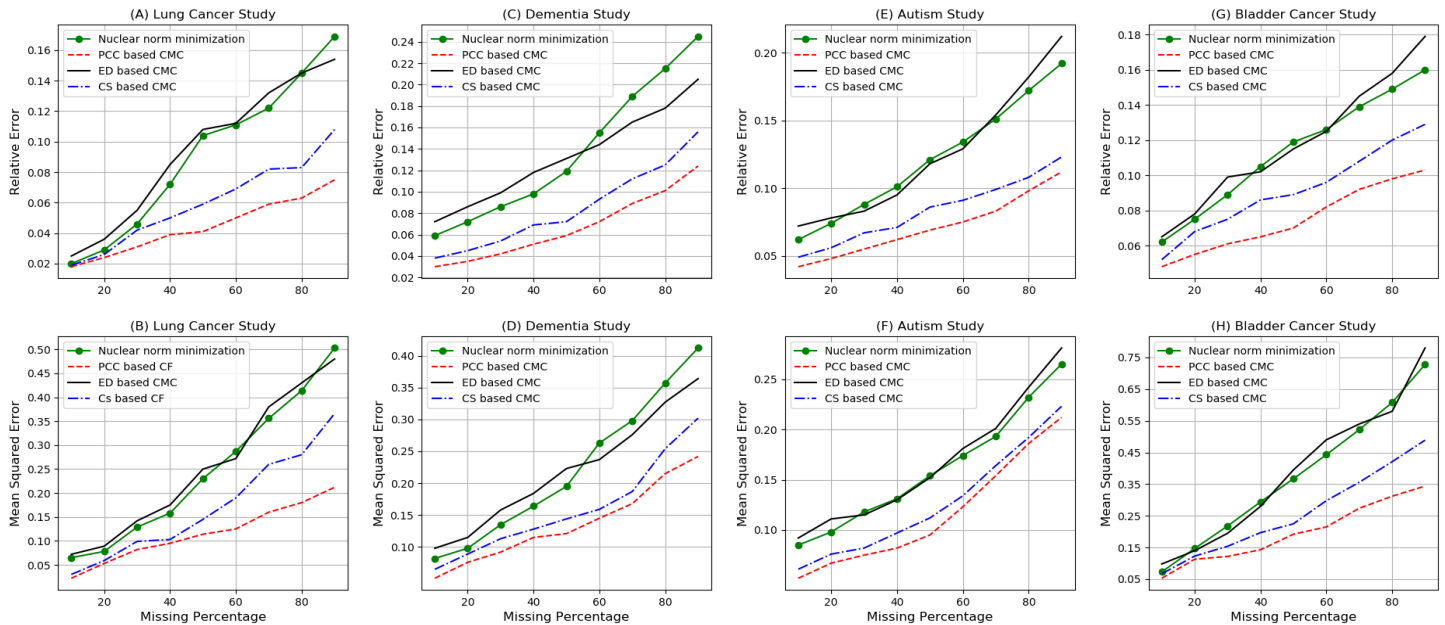


Figure 2. Comparison of the performance of 4 different methods on NCBI-GEO dataset

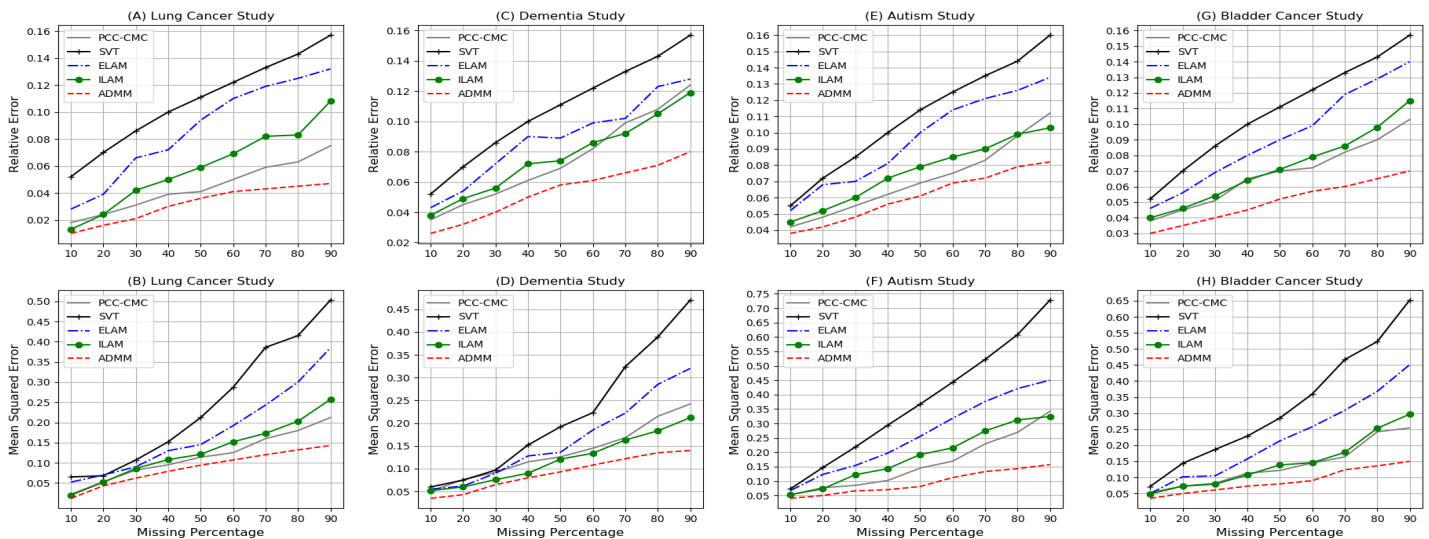


Figure 3. Comparison of 5 different approaches on 4 NCBI-GEO datasets.

blue and green lines show increasingly growing errors but the errors of the red line show a decreasing acceleration. The mean squared error of the ADMM approach grows much slower than that of the other four in cases of an increase in the ratio of missing entries.

Based on our results, The ADMM approach has higher accuracy especially in the cases where the matrix has more missing values. In the case of 90% missing values for relative error, in the best case, ADMM outperforms PCC-CMC, ILAM, ELAM, and SVT by a factor of 1.85, 3, 3.3, and 4 respectively. In the worst case ADMM outperforms PCC-CMC, ILAM, ELAM,

and SVT by a factor of 1.4, 1.3, 1.7 and 2 respectively. When looking at MSE, ADMM outperforms the other four approaches by as much as a factor of 2.2, 2.3, 3 and 4.8 respectively (for the same order as previously) and in the worst case we get an improvement factor of 1.6, 1.6, 2.3 and 3.4 respectively.

6. Conclusion

In this paper we employed three similarity metrics (Pearson Correlation Coefficient, Euclidean distance and Cosine Similarity) and Robust Principal Component Analysis (RPCA) on gene expression datasets. In section 2, we described the

correlation based MC approach, the RPCA approach and also we briefly explained the Alternating Direction Method of Multipliers (ADMM) algorithm. We used 4 different gene expression datasets from NCBI and in the first step, we randomly removed a fraction of the entries (from 10% - 90%). So for each dataset we had 9 incomplete matrices that we aim to recover and predict the missing values using one the aforementioned approaches. When we measured the accuracy of the three correlation based approaches, K-nearest neighbors and a recent nuclear-norm minimization based approach, we found out the PCC-CMC approach outperforms the other methods.

In another experiment we evaluate the performance of ADMM approach. To do this, we described three well known algorithms that can be used when recovering low rank matrices and we compared the performances of them. We found that ADMM approach outperforms the other approaches.

This paper can provide an inspiration for developing new approaches especially in gene expression studies and also has implications to recommender systems. There is a high demand for new efficient and fast methods to reduce the huge amount of time and resources that is often needed for gene expression studies. Using such computational methods can help biologists find missing values in partially known gene expression datasets and also can help identify promising directions for studies based on partial measurements in gene expression experiments.

References

- [1] N. Fraidouni and G. Zaruba, "A correlation based matrix completion approach to gene expression prediction," in *7th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, Orlando, FL, 2017.
- [2] N. Fraidouni and G. Zaruba, "A Robust Principal Component Analysis via Alternating Direction Method of Multipliers to Gene-Expression Prediction," in *Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, 2017.
- [3] R. Hammamieh, N. Chakraborty, A. Gautam and S. Muhie, "Whole-genome DNA methylation status associated with clinical PTSD measures of OIF/OEF veterans," *Transl Psychiatry*, PMID: 28696412, 2017.
- [4] Y. Bromberg, "Chapter 15: Disease gene prioritization," *PLoS Computational Biology*, 2013.
- [5] A. Wong, W. H. Au and K. Chen, "Discovering high-order patterns of gene expression levels," *Journal of Computational Biology*, pp. 625-637, 2008.
- [6] S. Welsh and S. Kay, "Reporter gene expression for monitoring gene transfer," *Current Opinions in Biotechnology*, pp. 617-622, 1997.
- [7] A. W. Liew, N. Law and H. Yan, "Missing value imputation for gene expression data: computational techniques to recover missing data from available information," *Briefings in Bioinformatics*, vol. 12, no. 5, pp. 495-513, 2011.
- [8] V. Gligorijevic and N. Przulj, "Computational Methods for Integration of Biological Data," *Springer International Publishing*, pp. 137-178, 2016.
- [9] X. Feng and X. He, "Inference on low rank data matrices with applications to microarray data," *The Annals of Applied Statistics*, pp. 217-243, 2010.
- [10] F. O. Isinkaye, Y. O. Folajimi and B. A. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261-273, 2015.
- [11] A. M. Acilar and A. Arslan, "A collaborative filtering method based on Artificial Immune Network," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8324-8332, 2009.
- [12] L. S. Chen, F. H. Hsu, M. C. Chen and Y. C. Hsu, "Developing recommender systems with the consideration of product profitability for sellers," *International Journal of Geographical Information Science*, vol. 187, no. 4, pp. 1032-1048, 2008.
- [13] M. Jalali, N. Mustafa, M. Sulaiman and A. Mamay, "WEBPUM: a web-based recommendation system to predict user future movement," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6201-6212, 2010.
- [14] M. Ekstrand, J. T. Reidl and J. Konstan, "Collaborative Filtering Recommender Systems," *Foundations and Trends in Human-Computer Interaction*, vol. 4, pp. 81-173, 2011.
- [15] Y. El Madani El Alami, E. H. Nfaoui and O. El Beqqali, "Improving Neighborhood-Based Collaborative Filtering by A Heuristic Approach and An Adjusted Similarity Measure," in *Proceedings of the International Conference on Big Data, Cloud and Applications*, Tetuan, Morocco, 2015.
- [16] F. Alqadad, C. Reddy and J. Hu, "Biclustering neighborhood-based collaborative filtering method for top-n recommender systems," *Springer-Verlag London*, 2014.
- [17] X. Zhou, C. Yang, H. Zhao and W. Yu, "Low-Rank Modeling and Its Applications in Image Analysis," *ACM Computing Surveys*, vol. 47, no. 2, 2014.
- [18] J. Gillard and K. Usevich, "Structured low-rank matrix completion for forecasting in time series analysis," *Elsevier*, 2018.
- [19] E. C. Lai, P. Tomancak, R. W. Williams and G. M. Rubin, "Computational identification of Drosophila MicroRNA genes," *Genome Biology*, vol. 4, 2003.
- [20] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Applied and Computational Mathematics*, 2008.
- [21] J. Wright, Y. Peng and Y. Ma, "Robust principal component analysis: exact recovery of corrupted low rank matrices by convex optimization," in *NIPS*, 2009.
- [22] J. F. Cai, E. J. Candes and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal of Optimization*, vol. 20, no. 4, pp. 1956-1982, 2010.
- [23] I. Daubechies, M. Defrise and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, 2008.
- [24] S. Taneja, C. Gupta, K. Goyal and D. Gureja, "An enhanced K-nearest neighbor algorithm using information gain and clustering," in *Fourth International Conference on Advanced Computing & Communication Technologies*, 2014.
- [25] A. Kapur, K. Marwah and G. Alterovitz, "Gene expression prediction using low-rank matrix completion," *BMC Bioinformatics*, pp. 1634-1654, 2016.
- [26] Z. Lin, M. Chen and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *Mathematical Programming*, 2010.
- [27] Z. Lin, M. Chen and Y. Ma, "Linearized Alternating Direction Method with Adaptive Penalty for Low Rank Representation," in *Conference on Neural Information Processing Systems (NIPS)*, 2011.
- [28] M. T. Landi, T. Dracheva, M. Rotunno and J. D. Figueroa, "Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival," *PLoS One*, PMID: 18297132, 2008.
- [29] A. S. Chen-Plotkin, F. Geser, J. B. Plotkin and C. M. Clark, "Variations in the progranulin gene affect global gene expression in frontotemporal lobar degeneration," *Human Molecular Genetics*, vol. 17, no. 10, p. PMID: 18223198, 2008, PMID: 18223198.
- [30] M. D. Alter, R. Kharkar, K. E. Ramsey and D. W. Craig, "Autism and increased paternal age related changes in global levels of gene expression regulation," *Plos One*, vol. 6, no. 2, p. PMID: 21379579, 2011.
- [31] M. Riestler, J. M. Taylor, A. Feifer and T. Koppie, "Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer," *Clinical Cancer Research*, PMID: 22228636, 2012.

Closed Approach of a Decoder Mobile for the 406 Mhz Distress Beacon

Billel Ali Srihen^{1,*}, Jean-Paul Yonnet², Malek Benslama¹

¹Laboratory of Electromagnetism and Telecommunications, University of Brothers Mentouri Constantine, 25000 Constantine, Algeria.

²Laboratory of Electrical Engineering, University Joseph Fourier Grenoble, 38000 Grenoble, France.

ARTICLE INFO

Article history:

Received: 16 August, 2018

Accepted: 28 October, 2018

Online: 10 November, 2018

Keywords:

Frame Decoder

COSPAS Sarsat

Decoder Mobile

GPS coordinates

406 MHz Distress Beacon

ABSTRACT

This article presents the design and realization of decoder mobile for distress beacon 406 MHz. The use of the sparse Fourier transforms in the processing of distress beacon signals. The characteristics of these decoders are mobility, able to decode all the frequencies of the beacons. It locates with great precision the location of the emitting emergency beacon, which is equipped with GPS, indicates the actual distance from which the alerts emanate. This compressed detection can give us a more detailed form of signals, so new results are given in the detection, the performance of the results is presented through experience.

1. Introduction

Canada, France, the Soviet Union and the United States are the first countries to sponsor the COSPAS Sarsat system at its inception which is a permanent international humanitarian satellite search and rescue system, detecting and locating beacons of emergency installed in ships, aircraft, and individuals. The program was subsequently joined by many other countries today (total forty countries and two organizations in 2016) [1] [2]. Regarding distress in the three cases cited, the finding the position of the disaster, and the necessary time of rescue and effectiveness depend on the reliability of the communication between the LUTs and the satellites [3] [4]. In order to provide semi-realistic alerts and location information from GEOSAR, the second-generation 406 MHz beacon is used to allow the location data to be encoded in the transmitted 406 MHz message [5]. The importance of this research is to reduce the time spent in research during the rescue process through the use of decoding screens, which in turn contain information inside the frame (GPS coordinates and the distance between the mobile decoder and markup), which enables us to identify the location and the implementation of the rescue process quickly.

2. Cospas Sarsat Concept

Figure1 presents a basic structure of the COSPAS Sarsat system [1] [2]. In the most disasters cases, the emergency beacons are activated when human lives are threatened. These alerts are received by the satellites and are retransmitted to the terrestrial

stations distributed throughout the world "Local User Terminals" (LUT). Alerts are routed to the country's Mission Control Center (MCC), which operates the satellite terrestrial stations from which LUT is located. Routing alerts include the location of the calculated beacon to the LUT received by one of the orbiting low-Earth (LEO) satellites.

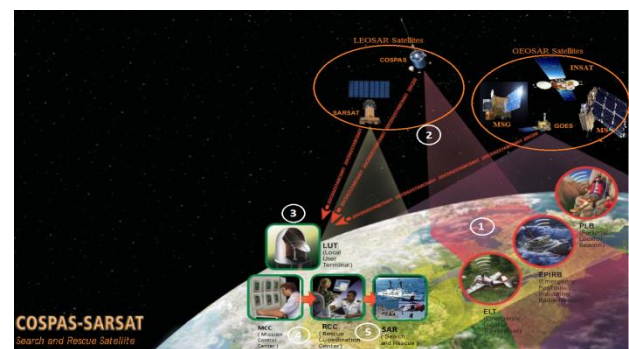


Figure 1. COSPAS Sarsat system [1]

The rescue operation is a carefully coordinated phase. Initially, the alerts are verified (based on the Doppler effect) after the confirmation the information is sent according to the country of the registration or to any rescue coordination center (RCC), which is equipped with emergency search and rescue correspondence [1] [2]. SAR satellites can also receive low distress signals from beacons.

After 2009, all beacons transmit on 406 MHz [6]. The universe is so complicated; the human being is always thirsty for discoveries enabling him to discover it, which exposes him to a danger, which

*Corresponding Author: Billel Ali Srihen, E-mail: bil_rst@yahoo.fr

has led through research and practice, the creation and realization of three types of emergency beacons currently in service [1] [2]:

- a. Personal Locator Beacons (PLBs) for individuals,
- b. Emergency Position Indicating Radio Beacons (EPIRBs) for maritime applications, and
- c. Emergency Locator Transmitters (ELTs) for aviation applications.

3. Frequency beacons cospas-sarsat 406 mhz

In January 2010, the 406.040 MHz frequency began to be used. Where the principles frequencies are 406.049 MHz and 406.052 MHz [6].

Where, the band 4 0 6.0 – 4 0 6.1 MHz has been set aside by the International Telecommunication Union (ITU) for the low power satellite emergency position Currently, Cospas Sarsat distress beacons transmit to 406.025, 406.028, 406.037, and 406.040 MHz.

4. Specification and Design

4.1. Specification

Specification of the second Generation 406 MHz SAR beacon and specification of ELT is defined in COSPAS-SARSAT technical document [5] [7].

4.2. Functional block diagram of decoder

Figure 2 shows the Mechanism functioning of a 406 MHz decoder. The main objective of the decoder is to identify the system, which that retransmits, determine when the reception took place, and then retransmit it again. The information to retransmit the weft shall be received as well as the temporal information. All the information's are displayed on one single screen with 4 lines of 20 characters. To do all this, the information is intensified.

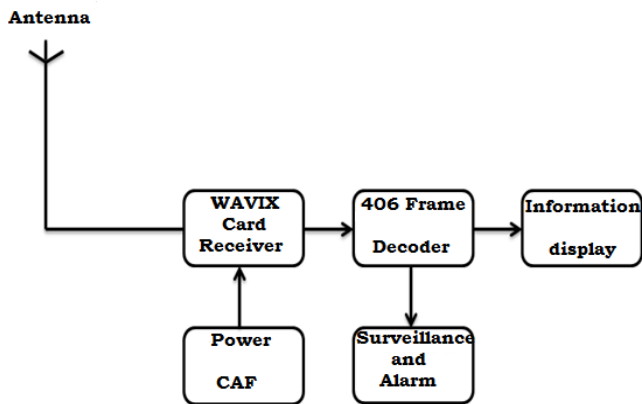


Figure 2. Functional schema of a 406 MHz decoder

5. Implementation of the Decoder

Figure 3 and 4 presents the Frame Decoder and complete system respectively. These two figures show: on the left side the jack 2.5 mm for the GPS connection, one on-off switch in the middle and the power jack which also serves as the battery charge point, and the jack 3.5 mm to connect the output receiver on the frequency of the beacon.



Figure 3. Frame decoder with display homepage.

On the right side, LED is used to monitor the operation and two push buttons allow navigating in memory (Figure 9). Since the system permanently records new message data, it is preferable to disconnect the entry when you want to see the pages in memory.

The complete mobile decoder after improvement of the old decoder. The system “receiver decoder” is now operational to be with the permanent listening of the beacons of distress is shown in Figure 4 [10]. Its sensitivity is very good. As soon as a beacon passes in emission, whatever its frequency, the receiver is fixed automatically on the real frequency of the beacon and the decoder displays the contained information in the screen. The sound of the buzzer is then made hear during one second.

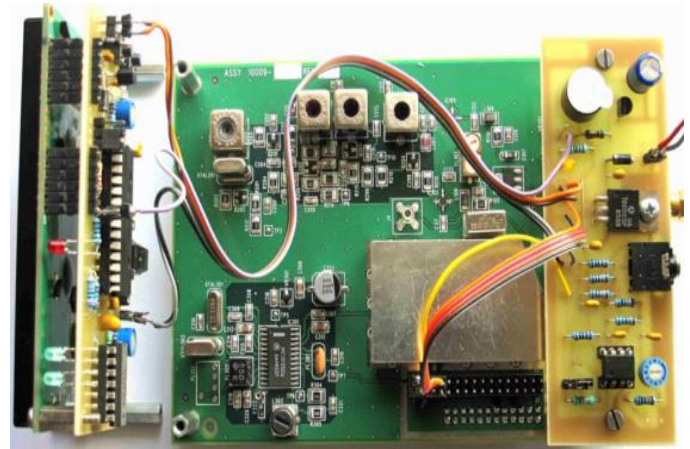


Figure 4. The complete system for receiving and listening to 406 MHz beacons

6. Experiment Results

6.1. Performance of the ELT Beacon

The 406.040 MHz ELT beacon signal is experiment results. The frequency meter and Spectrum Analyzer indicates the frequency is 406.040 MHz [10], it has the same spectrum of an EPIRB Beacon [11].

6.2. Position of the Beacon

Figure 5 corresponds to GPS position transmitted by the beacon. The frame is containing the position displayed on the fourth line. Where, all information is present except the time of receipt.



Figure 5. The GPS position transmitted by the beacon

The experience of the GPS receiver. Figure 6 shows Longitude Est. 02.0955⁰ and North Latitude 45.4700⁰ shows position in Google Earth. This indicates the point indicated in the figure this experiment was carried out.

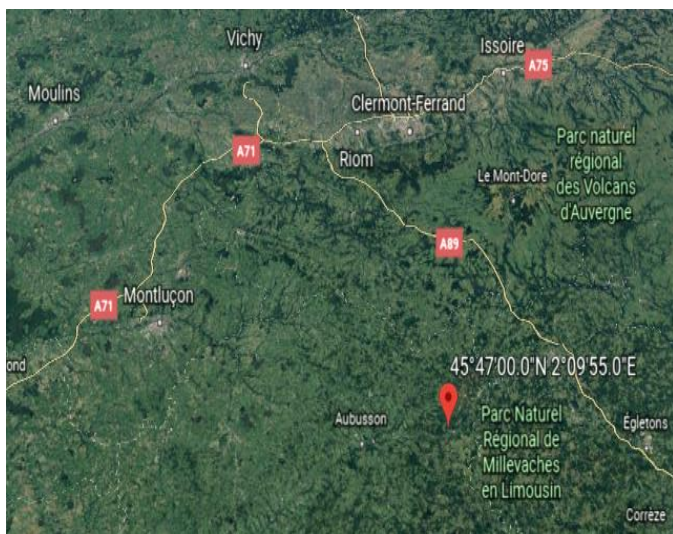


Figure 6. Position Plot in Google Earth

6.3. Functioning of frame decoder

Output of the receiver listening to the beacon

The signals emitted by the beacon are a series of peaks, alternately positive and negative. These peaks are 1.25 ms or 2.5 ms. they correspond to the phase jumps of the PSK modulation as shown in Figure 7 and Figure 8 . In fact these peaks correspond to the rising brows and the descending brows of the crenels that control the PSK modulator.

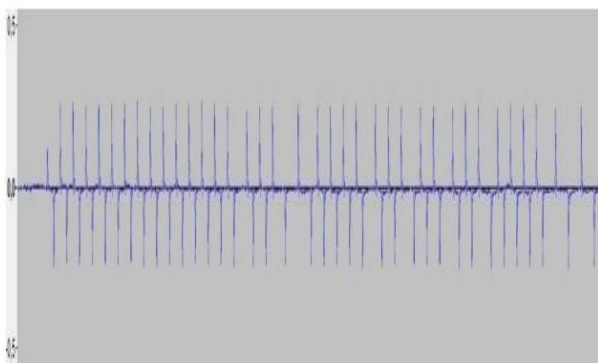


Figure 7. The frames emitted by the beacon

Figure 8 shows the signals at the output emitted by the Beacon; the frame information is more scope by battlements but by peaks. With this type of output by peaks, the frame generator performs an excellent simulation of the chain "Beacon 406 + receiver"

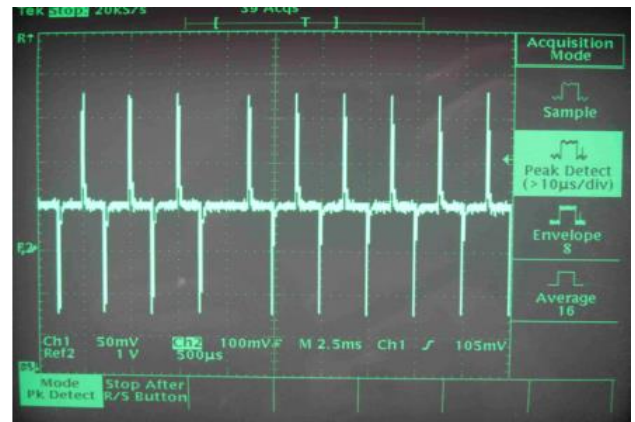


Figure 8. The signals at the output in the form of alternating peaks

Display of the distance between mobile decoder and the beacon

The calculation and display of the distance are fully automatic. As soon as the decoder knows its position and that of the beacon, it displays the distance and heading in place of the identification of the beacon on the second part of the third line. Figure 9, and 10 shows the distance display: it is 10:09 Local time, and heard beacon is a 49.9 Km north west.



Figure 9. Display of distance and heading

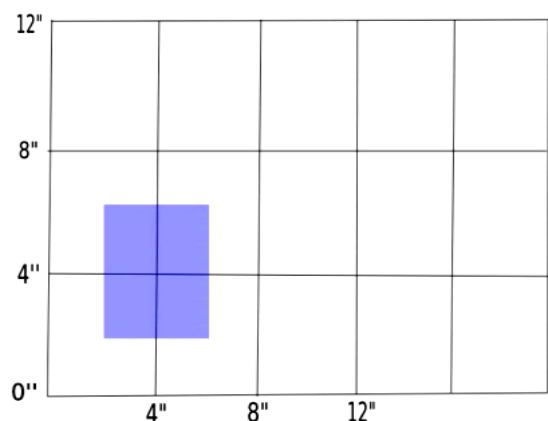


Figure 10. The position of the beacon is discretized at no 4 arc seconds, which leads to a position of uncertainty 123 m North and 87 m East

7. Conclusion

This work concerns the disaster beacon, even at great distances, and when the signals are weak, it is quite easy to recognize the signal of the 406 MHz beacon in the noise.

By coupling the receiver with a 406 MHz fram decoder, a beacon receiving station is obtained, which can operate as a permanent listener, and it will automatically decode the transmitted frame, so that our receiver-decoder is operational to be permanently listened to. The reception of a signal on any frequency, the receiver automatically stabilizes on the frequency of the beacon and our decoder displays the information contained in the frame (GPS coordinates and distance between Mobile decoder and beacon), which will allow us to quickly locate the place, and facilitate the intervention of rescues, consequently, reduce the search time. According the new development in monitoring beacon as showing by [13], we can expect more interesting results in the using of disaster beacon.

Conflict of Interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was done in the framework of a partnership between the University of Brothers Mentouri Constantine Algeria and the University Joseph Fourier Grenoble 1 France. Under the direction of CNRS Research Director emeritus Jean-Paul Yonnet.

References

- [1] <http://www.cospas-sarsat.org>, 30.05.2018.
- [2] <http://www.sarsat.noaa.gov>, 30.05.2018.
- [3] C. W. Scales and R. Swanson, "Air and Sea Rescue via Satellite Systems", *IEEE Spectrum* (ISSN 0018 – 9235), vol 21, pp. 48–52, March 1984. <https://doi.org/10.1109/MSPEC.1984.6370206>
- [4] I. W. Taylor and M. O. Vigneault, "A neural network application to search and rescue satellite aided tracking (SARSAT), In Proceedings of the Symposium", Workshop on Applications of Experts Systems in DND, pp. 189-201, Royal Military Coll. Of Canada, 1992.
- [5] J. V. King "New Developments in the COSPAS-SARSAT Satellite System for Search and Rescue", 55th International Astronautical Congress 2004 <https://doi.org/10.2514/6.IAC-04-M.4.07>
- [6] Specification for COSPAS – SARSAT 406 MHz Distress Beacons, C/S T.001 Issue 3 – Revision14, October 2013
- [7] COSPAS-SARSAT 406 MHz Frequency Management Plan, C/S T.012 Issue 1 – Revision 9, October 2013.
- [8] COSPAS-SARSAT 406 MHz Distress Beacon Type Approval Standard, C/S T.007 Issue 4 –Revision8, October 2013.
- [9] COSPAS-SARSAT Guidelines on 406 MHz Beacon Coding, Registration and Type Approval, C/S G.005, Issue 2 - Revision 5, October 2010.
- [10] B ALI SRIHEN, J.P Yonnet, and M BENSLAMA, "Design and realization an ELT beacon and decoders of frames 406 MHz". International Conference (IEEE) on Engineering & MIS (ICEMIS), 8-10 May 2017, Monastir, Tunisia <https://doi.org/10.1109/ICEMIS.2017.8272967>
- [11] I. Joo, J.H. Lee, Y.M. Lee, C.S. Sin, S.U Lee, and J Kim, "Development and Performance Analysis of The Second Generation 406 MHz EPIRB " International Conference (IEEE) 4th Advanced Satellite Mobile Systems, 26-28 Aug. 2008, Bologna, Italy <https://doi.org/10.1109/ASMS.2008.64>
- [12] S. STATLER "Beacon technologies" Apress 2016; p 21; p 82; San Diego; California, USA

An Improved Cross-Connection Abatement Algorithm with RSSI Using In-Band Magnetic Field Control in Densely Located LC Wireless Charger Environments

Nam Yoon Kim¹, Jinsung Cho², and Chang-Woo Kim^{3,*}

¹*Dept. of Marine Technology and Commissioned Officer, Dae Duk University, 34111, Republic of Korea*

²*Dept. of Computer Science and Engineering, Kyung Hee University, 17104, Republic of Korea*

³*Dept. of Electronic Engineering, Kyung Hee University, 17104, Republic of Korea*

ARTICLE INFO

Article history:

Received: 08 August, 2018

Accepted: 29 October, 2018

Online: 10 November, 2018

Keywords:

Wireless Power Transfer

Cross-connection error

Magnetic field control

ABSTRACT

In densely located loosely coupled wireless charger environments, cross-connection errors can occur when wireless chargers operate at the same time within the same wireless communication range. In this work, an effective algorithm is proposed to prevent cross-connection error. The algorithm based on the cross-connection abatement technique with the received signal strength indicator of out-band BLE communication has improved using an in-band magnetic field signal controlled by a pulse-width-modulation-like waveform. The experimental results obtained from a proper test set in a RF shield room verify that the wireless charging system using the proposed algorithm provides wireless charging services without cross-connection errors.

1. Introduction

At present, loosely-coupled (LC) wireless chargers are employed to support the simultaneous charging of a multiplicity of devices, such as laptops, tablets, and mobile phones in multiple wireless charging pads (power transmitting units: PTUs) and charging devices (power receiving units: PRUs) [1-3]. In densely located LC wireless charger environments, cross-connection errors can occur when PTUs and PRUs operate simultaneously in the same wireless communication range. The cross-connection errors are a kind of the control error between the main PTUs and neighboring PRUs due to the interconnection of the PTU/PRU communication. Figure 1 shows the concept of the cross-connection errors. When two (or more) PTUs with their own PRU are disposed closely together and operate at the same time, a control error may occur as one PTU's communication connects to the other PRU. If the PTU recognizes the battery information of the other PRU, over-power or low-power transmission can occur which may damage the PTU or prevent charging. In other words, the cross-connection error may cause the over-power or low-power transfer that may damage the PTU and PRU. In [4], we proposed an effective algorithm that focuses on the measured value of the received signal strength indicator (RSSI) to avoid cross-connection error. The RSSI, however, is heavily affected by nearby objects

and by environments regardless of distance [5]; *i.e.*, an RSSI signal tends to fluctuate due to many external factors, such as obstacles, multipath fading, interference diffraction, and absorption. Moreover, when the distances between PTUs are small, it is difficult for a PTU to distinguish its own PRU from the others via only comparison of the RSSI values because the RSSI values become similar. Therefore, the algorithm using only the RSSI signals is not able to prevent communication connection errors in very dense environments.

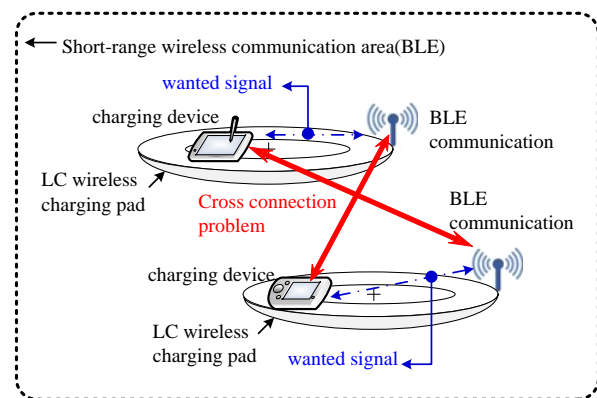


Figure 1. Conceptual diagram of the cross-connection errors between the charging devices and the wireless charging pads within the short-range wireless communication area.

*Chang-Woo Kim, 1732 Deogyong-daero, Giheung-gu, Yongin, Gyeonggi-do, Republic of Korea, cwkim@khu.ac.kr

In this work, we propose an effective algorithm to solve the RSSI problems mentioned above. The proposed cross-connection abatement (CCA) algorithm is based on the CCA technique with the RSSI of out-band (2.4-GHz BLE) communication improved using the controlled magnetic field signal of in-band (6.78 MHz) power transmission. The CCA technique uses the out-band RSSI and in-band magnetic field signals progressively to enable the wireless chargers (PTUs) to be used in all dense environment without cross-connection errors. The PTU first finds the PRUs that are close to each other using the RSSI, then finds its own PRUs using the magnetic field signal controlled by the output power of the power amplifier. The PTU uses a waveform-controlled magnetic field signal, such as a pulse width modulation (PWM) signal, as a second signal to sense its PRU. To validate the performance of the proposed CCA algorithm, a test set was built in a dense charger environments.

2. System Description and Experimental Results

A test set was fabricated in a dense, loosely coupled (LC) measuring environment in a radio frequency shielded chamber with a 6.38 (length) m × 5.48 (width) m × 3.75 (height) m dimension [4]. The test set was composed of 10 wireless charging pads, 10 smartphones, a 16-V power on-off control switch, and a personal computer. The wireless charging pad with an area of 20 cm × 30 cm act as a PTU and the smartphone with a built-in wireless charging module acts as a PRU. The wireless charging pad that operates at 6.78 MHz with a 2.4-GHz Bluetooth low energy used to establish between PTUs and PRUs communication. The graphical user interface (GUI) monitors Bluetooth low energy RSSI values of all PRUs.

Figure 2 shows block diagrams of a PTU and PRU system used in the test set in a dense wireless-charger environment (two chargers within a 10-cm distance) and their output voltage waveforms as applied to the proposed CCA algorithm. The PRU is composed of a rectifier, a DC/DC converter, and a monitoring circuit block. The rectifier converts received AC power signal to DC voltage. The DC/DC converter converts a received low voltage to a high voltage and the monitoring block checks the charging conditions. The PTU is composed of a power amplifier, a buck converter, a wireless power on/off control switch, and a microprocessor. The power amplifier produces a 6.78-MHz RF power and the buck converter controls the output power, which produces a PWM-like waveform. For the PTU and PRU communication, a 2.4-GHz BLE was used. In [6], we proposed an effective CCA technique. The working sequence of the proposed technique is as follows.

To recognize the presence of a PRU in the PTU charging area, the wake-up power signal is transmitted over a regular cycle to sense the load. Once the advertisement signal is received, before the charging device is registered, an RSSI threshold level is identified to ensure that the PRU exists in the charging area. The RSSI threshold is measured as the position of the PRU in the PTU's charging area changes. The RSSI threshold level can be selected to be 3 dB (for noise margin) less than the minimum value of the highest RSSI level (for example, -12 dBm in [4]) among the received RSSI of PRUs. Once the threshold level is determined, PRUs with a value lower than that threshold is considered to have the cross-connection error. When this happens, the PTU turns off

the wireless power for 100ms and returns to the load detection phase for a normal reconnection. The PRU with the RSSI threshold level registers its ID to initiate the wireless charging service. Then, the PTU monitors the load detection and advertisement signal of the PRUs in order to detect PRUs added or removed in real time during the wireless charging process.

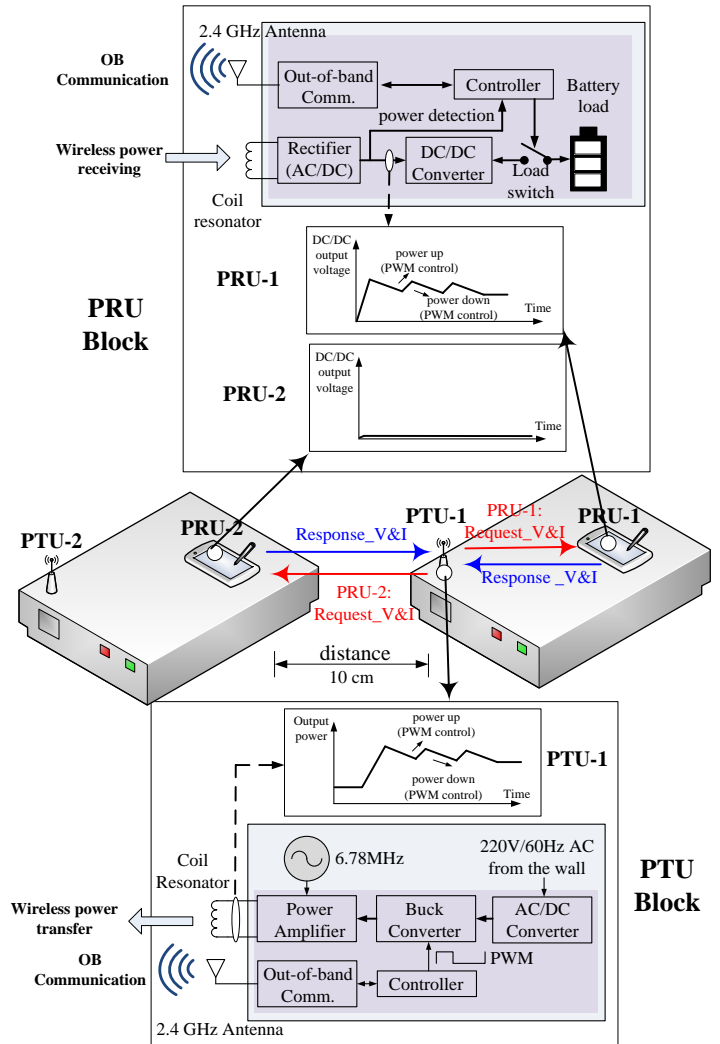
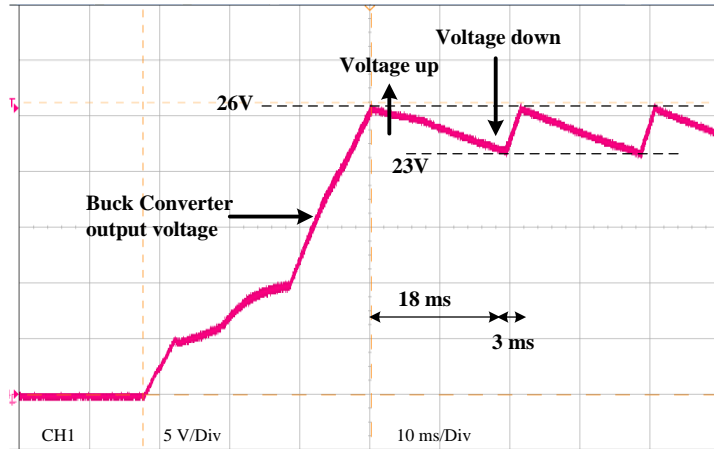


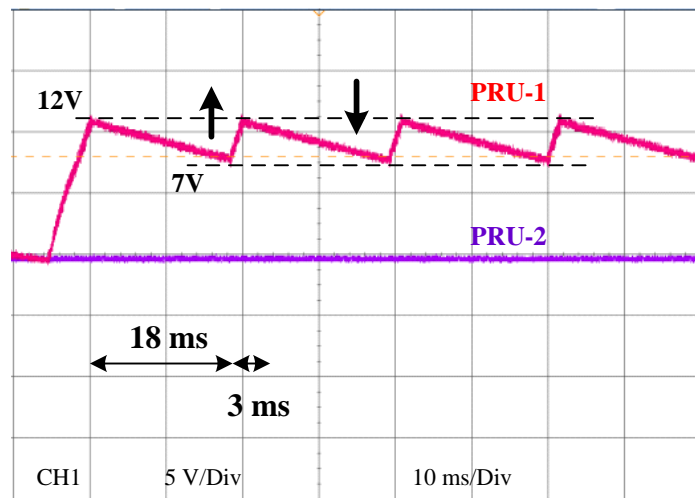
Figure 2. Block diagrams of a PTU and PRU system used in the test set and schematic output voltage plots measured from PTU-1 and PRU-1, 2. PRU-1 belongs to the PTU-1, whereas PRU-2 is cross connected.

As the distance between two wireless charging pads (PTUs) decreases (less than 10 cm as shown in Figure 2), it becomes difficult for the PTU to distinguish its own PRU from the other one merely by comparing the RSSI values. This is because the received RSSI values of two PRUs eventually become similar to each other at the threshold level. For PRUs with RSSI values below the threshold, the communication connection is not made because the PTU identifies them on different charging pads. However, if the RSSI values of PRUs may be greater than or equal to the threshold, these PRUs should be revalidated for the cross-connection status. In that case, the PTU uses the magnetic field control of the transmit power signal to reaffirm the cross-connection status. A change in the magnetic field strength on the resonator of the PTU induced by varying the input current supplied from the power amplifier results in the change of the power information received by PRUs. The

PTU sends a magnetic field signal controlled by a waveform, such as a PWM-like waveform, which increases or decreases the output power of the power amplifier. This signal is then compared with the power change information received by the PRUs. If the power variation pattern is different from the PWM-like variation pattern, the cross-connection is confirmed.



(a)



(b)

Figure 3. Measured voltage waveforms of the PTU and the PRUs shown in Fig. 2. (a) Output voltage waveform of the power amplifier in PTU-1 (inset in Fig. 2). (b) Received voltage waveform at PRU-1 and PRU-2 (inset in Fig. 2).

Figure 3 shows the voltage waveforms measured at PTU-1, PRU-1, and PRU-2. Figure 3 (a) shows the output voltage measured at the Buck converter of PTU-1 which induced the magnetic field signal controlled by a waveform like PWM. Figure 3 (b) shows those at DC/DC converters of PRU-1 and PRU-2. In Figure 3 (a), the output voltage of the power amplifier varies from 23 V to 26 V with a 21-ms period. The rise time is 3 ms and the fall time is 18 ms. As shown in Figure 3 (b), the voltage change from 7V to 12V is repeatedly measured in PRU-1, which has a 21-ms period with a 3-ms rise time and 18-ms fall time. On the other hands the PRU-2 does not exhibit the voltage change. The measured results indicate that PRU-1 is a self-charging device of PTU-1, while PRU-2 belongs to another pad.

After PTU-1 identifies that PRU-2 is not its own PRU, PTU-1 disconnects communication with PRU-2 and connects to PRU-1 again to start the charging process. As a result, it is possible to

accurately detect the occurrence of the cross-connection error by comparing the change in the intensity of the magnetic field of the PTU and the power change pattern received in the PRU.

Table 1 shows a comparison of the effectiveness of cross-connection abatement algorithms on the distance between PTUs. When the distance between PTUs is relatively long (longer than 15 cm), both techniques are very effective. As the distance is short (shorter than 10 cm), the CCA algorithm with RSSI and modulated magnetic field signal is still effective, while the CCA algorithm with only RSSI signal is not effective because the PTU does not recognize its own PRU.

Table 1. The comparison of effectiveness of cross-connection abatement algorithms on the distance between PTUs.

	Long distance between PTUs	Short distance between PTUs	Ref
CCA algorithm with only RSSI signal	effective	not effective	[4]
CCA algorithm with RSSI & modulated magnetic signals	effective	effective	This work

3. Improved Algorithm Description

Figure 4 shows the process flow of the proposed CCA algorithm with the RSSI and the controlled magnetic field signals. As shown in the figure, the left line (PTU) shows the command frames and the power levels (dark areas) of the PTU, and the center line (cross-connected PRU) and right line (own PRU) show the response and report frames, respectively, according to the request of the PTU. This process flow operates in four states: a load detection state, cross-connection prevention state, registration state, and charging state. The basic operation sequence of the CCA algorithm is as follows.

In the load detection state, the wake-up power supplies 5 V to the PA for 500 ms (t_{BEACON}) for every 3 sec ($t_{\text{BEACON_PERIOD}}$) cycle. The wake-up power is the minimum power level necessary to activate the micro control unit (MCU) of the PRUs to establish the communication. When a PRU is detected, the MCU of the PTU increases the voltage supply at the PA to 7 V to supply sufficient power for the PRU communication and the stable operation of the MCU. The woken-up PRUs proceed to the next stage, namely, the cross-connection prevention state. After the load detection between the PTU and PRUs, the PTU transmits a “request RSSI” frame to the PRUs in order to verify the state of cross connection. Upon receipt of the “request RSSI” frame, the woken-up PRUs immediately transmit their RSSI values to the PTU. By comparing the RSSI values with the threshold value (-15 dBm in [4]), the PTU first recognizes those PRUs as the possible PRUs that have greater RSSI values than the threshold value, before proceeding to accurately judge them as its own PRUs. In the case of PRUs that have lower RSSI values than the threshold value, no communication connection is made because the PTU recognizes them as belonging to other charging pads (PTUs). The PRUs that have higher RSSI values than the threshold value should be rechecked for the state of cross connection because RSSI values can become greater than the threshold value when the distance between the charging pads decreases. The PTU rechecks the state

of cross-connection using the magnetic field control of the transmitting power signal. The change in the magnetic field strength on a resonator of the PTU, induced by changing the input current supplied from a PA, changes the power (voltage and current, V&I) received by the PRUs. The power (V&I) change information received by the PRUs is reported through the “report V&I” frames whenever the “request V&I” frames are transmitted during this state.

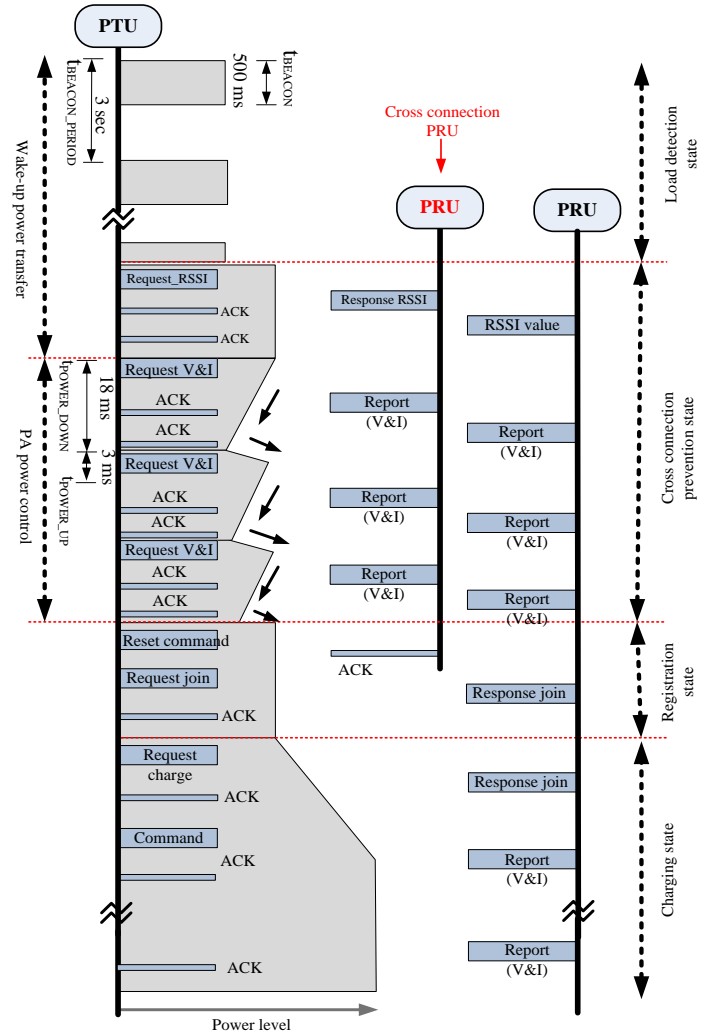


Figure 4. Sequence diagram of the proposed algorithm for cross-connection abatement.

The PTU causes an increase or decrease in the output power of the power amplifier and compares it with the power change information received by the PRUs. If these two power variation patterns are different, then a cross connection has occurred. In the algorithm, the magnetic field control is implemented for 18 ms (t_{POWER_UP}) for a power increase at the PA and for 3 ms (t_{POWER_DOWN}) for a power reduction at an interval of 21 ms. The control is carried out three times in a sequence. To detect the power change information (V&I) received by the PRUs based on the magnetic field control, the PTU transmits the relevant “request V&I” frames. Upon the receipt of the “request V&I” frames, the PRUs transmit the measured V&I information through the “report V&I” frames. The communication between the PTU and the PRUs occurs during the timespan of 21 ms of the power change. The three repetitions of the power-level change at the PTU and the power change information received by the PRUs are compared, and if the

variation patterns turn out to be different or there are no changes in the received power, the PRUs start detecting places where cross-connection errors have occurred.

Consequently, the PTU transmits a reset command to the PRUs (such as the PRU-2) where cross-connections have occurred to disconnect the communication. The PRUs receive the reset command and switch to a standby state in order to receive the “request” frames from other PTUs. The PRUs (such as the PRU-1) that are verified to be its own proceed to the registration state.

After detecting the PRUs that are to be charged on its own pad in a cross-connection prevention state, the PTU sends a subscription request to its network and receives a session ID and the charging permit. The PTU then transmits the “request join” frames to its own PRUs. Consequently, the PRUs transmit the battery capacity information, which is necessary for charging, via the “response join” frames. By comparing the power capacity that the PTU can transmit and the battery capacity of the registered PRUs, the PTU decides whether the charging should be allowed. If the power transfer capacity exceeds, charging is not allowed, and the PRUs revert to the standby state. This information is notified to the PRUs through the acknowledgement (ACK) frames. After this registration state, the PTU proceeds to the charging state and starts a charging process. During the charging state, the PTU transmits the “charge start” commands to relevant PRUs to switch them to the charge state for charging the batteries. The PRUs switch on the load attached to the battery path and receive the power. The PTU monitors whether the power is delivered properly to the PRUs through the “report V&I” frames. Based on the power information received from PRUs, the PA adjusts its output power. The PTU sends back the “report” frames several times so that the power can be delivered stably to the PRUs. When the charging power is delivered to the PRUs, the voltage and current received by the battery are continuously monitored so that the required power can be controlled according to the charge mode. If the monitored voltage of the battery reaches a buffer voltage, the PTU judges that the charge has been completed. Thereupon, the PTU issues the “charge finish” commands to PRUs and switches the load connected to the battery to the off state, along with the transmission of an ACK command before displaying a buffer to the users through a user interface.

4. Conclusions

The proposed CCA algorithm has shown that cross-connection errors between wireless charging pads and charging devices in BLE wireless coverage can be prevented using simple RSSI threshold level detection and magnetic field control for power signal transmission. This approach can provide a reliable loosely coupled wireless charging service without an abrupt increasing in transmit power at the start of charging due to the gradual power-up pattern of the controlled magnetic field.

Acknowledgment

This work was supported by the Basic Science Research program through the National Research Foundation Korea funded by the Ministry of Education (NRF-2015R1D1A1A02061041) and by the Ministry of Science and ICT Korea under the ITRC (information technology research center) support program (IITP-2016-R2718-16-0012) supervised by the IITP (national IT industry promotion agency).

Reference

- [1] J.J. Casanova, Z.N. Low, and J. Lin, "A loosely coupled planar wireless power system for multiple receivers," *IEEE Trans. Ind. Electron.*, vol. 56, no. 8, pp. 3060-3068, Aug. 2009. <https://doi.org/10.1109/TIE.2009.2023633>
- [2] R. Tseng, B. von Novak, S. Shevde, and A.K. Grajski, "Introduction to the alliance for wireless power loosely coupled wireless power transfer system specification, Version 1.0," in 2013 IEEE WPT, May 2013, pp. 79–83. <https://doi.org/10.1109/WPT.2013.6556887>
- [3] I. Yasar, L. Shi, K. Bai, X. Rong, Y. Liu, and X. Wang "Mobile phone mid-range wireless charger development via coupled magnetic resonance," in 2016 IEEE ITEC, June 2016, pp.1–8. <https://doi.org/10.1109/ITEC.2016.7520217>
- [4] N.Y. Kim, S.-W. Yoon, and C.-W. Kim, "Cross-connection abatement in dense loosely coupled wireless charging pad environments," *Electronics Lett.*, vol. 50, no. 6, pp. 461-462, Mar. 2014. <https://doi.org/10.1049/el.2013.3754>
- [5] B.-G. Lee and W.-Y. Chung, "Multi-target three-dimensional indoor navigation on a PDA in a wireless sensor network," *IEEE Sensors Journal*, vol. 11, no. 3, pp. 799–808 Mar. 2011. <https://doi.org/10.1109/JSEN.2010.2076802>
- [6] N.Y. Kim, J. Cho, and C.-W. Kim, "An effective technique for preventing cross-connection errors in dense loosely-coupled wireless charging pad environments," in 2017 ICUFN, July 2017, pp623-626. <https://doi.org/10.1109/ICUFN.2017.7993866>

3D Reconstruction of Monuments from Drone Photographs Based on The Spatial Reconstruction of The Photogrammetric Method

Andras Molnar*

John von Neumann faculty of Informatic, Obuda University, 1034 Budapest, Hungary

ARTICLE INFO

Article history:

Received: 25 September, 2018

Accepted: 15 October, 2018

Online: 10 November, 2018

Keywords:

Drone

Aerial Photography

3d Surface Model

Photogrammetry,

Monument Protection

Virtual Museum

ABSTRACT

Due to their efficient flight control systems and their camera of high quality modern drones can fly precisely and take aerial photos of high resolution. Although these multi-rotor devices are not able to fly long distances yet, they are very efficient instruments for taking aerial photographs of their proximate environment. As they are able to float and rotate along their vertical axis to a discretionary direction, they can be used for monitoring and taking pictures of a building from several directions. If pictures are taken of a building from all directions with significant overlap (of at least 50%) the 3D reconstruction of the building in a photogrammetric way becomes possible. The reconstructed models do not only contain the spatial forms but also the visual information based on the snapshots. As a result of the entire reconstruction a virtual object is gained in the virtual space that can freely be accessed and visible from all directions by enlarging or reducing the size. A virtual collection can be established for monuments, historical buildings or other spectacular objects worth recording. The objects of the collection (buildings) are lifelike and can be perceived and studied by anyone. The 3D models, of course, cannot substitute the photographs of high resolution but they can complement them in the collection.

1. Introduction

The fundamentals of photogrammetry have been in existence since 1858 [1]. It is interesting to note, however, that surveys based on photographs were first carried out for buildings. Surface models in aerial photography were given birth only later. The procedure of photogrammetry basically calling for a lot of calculations has been made widely accessible to the users due to computers of high performance and GPGPU cards that make calculations significantly faster. A kind of revolution is also noticeable with regards to aerial photography due to the spread of the cheaper and cheaper remote controlled flying objects of higher standard that carry cameras. These two technological developments made aerial photography and processing photos by computers widespread.

The aerial photographs taken by unmanned aerial vehicles (UAV) can substitute for the traditional aerial photography by airplanes. Following the processing of pictures the resolution of the photo fitted together can be 3-3.5 cm/pixel (10-14 Mpixel native resolution) even by using standard cameras. While processing overlapping photographs [2] (overlapping of 60% is required alongside all edges) the 3D relief map of the area can also

be made in addition to the most frequently used orthophotos. According to experience in measuring it can be reduced even below 2-3 cm by using multi-rotor devices when taking an up-close picture (3-5 m) of a smaller object (building, monument etc.) specially. The results of 3D surveys can be used in several areas of the recording systems including archiving the temporary state of certain formations, establishing virtual museum for buildings, monuments and relics of architecture, monitoring and recording the processes of real estate investments, making records of public utilities, registering the output of surface mines, surveying and recording the recultivation stage of abandoned surface mines etc [3,4,5].

The paper is focused on the possibilities of making 3D pictures of different architectural monuments and their presentation in a virtual exhibition by presenting real surveys [6].

2. Photogrammetry

Photogrammetry is almost as old as photography. In 1858 when photochemical picture recording became well-known [7] Albrecht Meydenbauer a young architect resulting from his fortunate accident worked out a procedure where measuring and surveying buildings became possible on the basis of photos taken

*Andras Molnar, , Email: molnar@uni-obuda.hu

of them. The term photogrammetry is also linked to Meydenbauer. At first the term was published in an anonymous professional article of *Wochenblatt des Architektenvereins zu Berlin* in 1867 [8]. It was only later disclosed that the author of the article was Meydenbauer. The procedure is based on elementary geometrics by making use of the linear spread of light and relating the photo to the object being photographed by taking the camera and its optical parameters into consideration. Imaging itself is not too complicated but in the case of complex objects, calculations on its corner points are extremely time-consuming. It is worth noting that originally Meydenbauer worked out photogrammetry for surveying buildings but later on this method became widespread by processing aerial and space photographs. By using this method planar and orthophotos can be taken which have real significance in photography. Basically, orthophotos mean the aerial pictures taken in photography on which factual measures (distance, area) can be made. In many cases these methods help make maps designed by traditional methods more precise. The photogrammetry procedure for the 3D model construction of buildings has become more popular with the spread of computers of powerful calculating performance. The reason lies in the fact that fitting a lot of parts of photographs together manually is demanding but the use of traditional computers was very restricted (in the case of few photos of small resolution) for automation. In general, good quality textured 3D models require the analysis and processing more than ten Gigabytes data.

3. Making 3D models on photogrammetric basis

The necessary prerequisite of making photogrammetric 3D models [9] is to have several photographs of the same relief from slightly different positions. This can happen when such aerial pictures are taken in practice where overlapping between frames is ensured with all the imaged pixels appearing at least in two pictures (when special boundary conditions are met). However, in practice pixels must appear in more than two photos. If the spatial position of recording the image is known together with the direction vector of the optical axis of the camera, the line running through the projection plane and defining the pixel concerned can be described. If this pixel can also be found in another picture, then another line also runs through the projection plane but its point of intersection is not identical with the one of the projection plane of the previous pixel disregarding some special cases. It is made possible by the fact that the same pixel observed from different perspectives is seemingly placed elsewhere. At the same time, the two projection intersecting lines define a point outside the projection plane that corresponds with the spatial position of the pixel in question.

Two basic requirements of defining projection beams are knowing the exact position of recording the image and the optical axis of the camera together with the pixel pairs (the same pixels) of the overlapping pictures. The position of recording the image can derive from the sensor system built in the camera or attached to it. Basically, this sensor system means 3D coordinates of summa rising GPS (Global Positioning System) and IMU (Inertial Measurement Unit) data and a unit for orientation. However, these data can also be defined by the visual information of the images recorded. With regard to the fact that due to the reasons listed above the pictures are taken with significant necessary overlapping the optical axis of the camera and its relative spatial position can

also be defined based on their content analysis. Provided that the central pixel of the image from a geometrical aspect is free from distortion in each picture and it is regarded as the reference point between the pictures, special images in a different position can also be found in several pictures. The distance of these pixels from the reference point does not change in real although it differs from picture to picture. This makes the definition of such a transformation possible which gives the spatial position of recording the image. The method can be well automated but calls for a visual content accessible where the special pixels described above can be defined properly. Some examples include the corner points of buildings and intersections of roads or other objects discernibly separated from the background etc. The method cannot be used in homogeneous pictures such as ruffle and water surfaces free from structured reflected images. From a practical point of view when examining surface objects in most cases the well-arranged and contrastive structure serve with enough reference points, fortunately.

The similar point pairs ensure connection between the images. Finding these point pairs can also be automated similarly to finding the position. The basis of searching for these point pairs is to find the so-called corner points in the pictures that stand apart from the other pixels. They are typically pixels on the border of strong intensity changes such as corners of buildings, borders of forests, highways etc. Certain characteristics can be assigned to these points such as intensity gradient and/or the intensity spread of adjacent pixels. Several corner point detecting algorithms can be used for that purpose including the most frequently used Harris-algorithm [10,11]. If these characteristics are invariant to any magnification or rotation, there is a possibility for finding more corner points with similar characteristics in the overlapping point pair. Similarly to detecting corner points, several algorithms are also available for defining such point pairs including SSD (Sum of Squared Difference) algorithm. The corner points founded and arranged in pairs must be further refined by filtering algorithms as at this stage of processing there are several faulty point pairs in the system. One of the best known filtering methods is RANSAC (RANdom SAMple Consensus).

By using the algorithms above a spatial point cloud can be created from the overlapping images that correspond with the discrete set of points of the surface of the surveyed area. The contiguous surface model is created by connecting the points together. Of course, while creating the surface model several algorithms can also be applied. One of the best known is called the Delaunay triangulation. As a result, a surface model covered by contiguous triangles can be derived. Special morphological filters containing the presumptions typical of the relief conditions can also be applied. These filters refine the strikingly high elements of the point cloud.

If the position of recording the images was defined on the basis of the visual content and not GPS the surface model created is not capable of making quantitative measurements yet. To this end, geo-referencing the data set created is required. The condition of geo-referencing is that the original area should have geodesic reference points or reference points exactly measured and identified in the pictures prior to recording. A less exact method is to define the coordinates of the well-identified objects in the picture by using other databases such as Google Earth. A 3D object

placed in the world and measurable is created as a result of georeferencing.

There are customized systems for processing overlapping images that ensure all the services listed above and in half-automated way they create the 3D model of the surveyed area. Such systems include Agisoft PhotoScan [12].

4. The practical aspects of surveying buildings

The method applied when surveying areas for a general orthophoto is not suitable for surveying buildings. However, there is a surface model from the pictures of the same flight line which includes the extensions of the building by large in real but some details are never entered into the system. In the case of buildings there are several parts which are covered when taking an aerial photo.

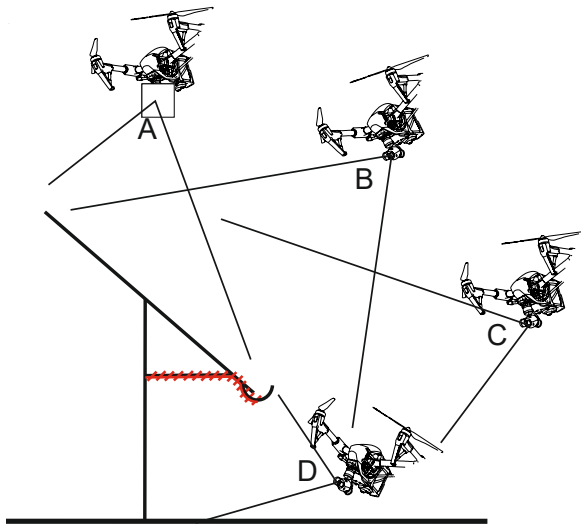


Figure 1: The visibility of the surface from different camera positions.

Figure 1 presents the visibility of surfaces when taking pictures of a typical building. It can generally be stated that taking photos from positions 'A', 'B' and 'C' does not mean a problem. At the same time, however, it can be seen that from these positions the bottom of the eaves marked by a stripe can never be seen so a real surface of this area cannot be taken in 3D reconstruction. Camera position 'D' is necessary to make pictures of the nonvisible parts. Unfortunately, in several cases this position is risky as the multirotor device carrying the camera stays very close to both the building and the ground. In the case of high buildings position 'D' is relatively easy to carry out as the minimum height of the multirotor above the area is higher, too. If it is feasible that the flying device should not go lower than 1-2 metres, pictures from position 'D' can also be taken. In the case of lower buildings taking pictures from position 'D' is also feasible if the drone is taken by hand and setting its camera to a proper direction photos of the building can be taken.

In the case of buildings the lateral and, in many cases, upward camera position is of high importance. That is why the fixed-wing aircrafts are not suitable for this task. Multirotor devices are the most efficient camera carriers at the current level of technological development when surveying buildings. Although helicopter-like devices are able to perform a similar movement to that of multirotor devices but their mechanical design is much more

complex. They are more prone to damage so their transportation is more complicated. The only disadvantage of multirotor devices is that they are unable to auto-rotate. It means when the power batteries are out these devices cannot fly. As flight is very close to buildings when surveying only the most developed multirotor devices with the safest features should be used. As a result of most recent developments the control panels of the 8-rotor-multirotors are able to stand in for an engine while properly controlling the other working ones. Using such instruments can significantly improve the safety of surveys. As the primary position data of multirotor flying objects are processed on the basis of on-board GPS good reception is essential. Unfortunately, it is the low flight and the covering of the building that can negatively influence GPS reception. Weak GPS data result in the multirotor's apparent swift and imbalance, which has to be compensated by the operator. Although this position keeping flaw is not dangerous on its own, it can really be a cause for concern due to the proximity of the building or any other objects. During flights close to objects it is not practical to apply automated flight lines as its precision is not satisfactory. In many cases blocked live images of the camera must also be taken into consideration as higher buildings can shadow the frequency of the video signal carrier. For security reasons it is practical to plan flights during which the operator can see the multirotor. In practice, it means that the flying object must be accompanied round and round the object on the ground.

The automated point pair search in the pictures makes standard photo taking from all directions unnecessary. Attention must only be paid that every part of the given building should be recorded in as many pictures as possible so a lot of overlapping pictures should be taken.

Practical experience says that pictures of 3-4 m are necessary for good quality reconstruction by means of a 12 megapixel camera free from distortion.

Figure 2 illustrates the process of photographing a small chapel. The positions of making the single pictures can be seen above the building in the picture together with the optical axis of the camera at the moment of taking pictures. It can also be seen that in the survey the pictures that should have been taken upward next to the building from position 'D' of Figure 1 are missing. Despite the missing photographs 3D reconstruction could be carried out although at some parts of the building it is of bad quality (Figure 3).

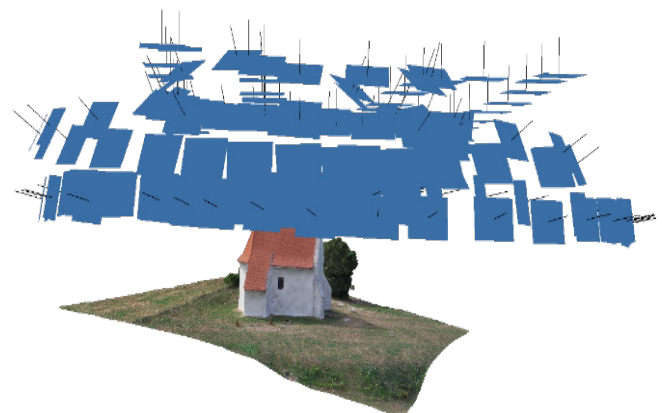


Figure 2: The spatial position and orientation of pictures taken when surveying a chapel.

The order of making photos is not fixed, which means that pictures can be taken discretionary even randomly. The only criterion is that all parts of the object must be photographed and each part should appear in several pictures. In the case of a bigger building this requirement can be met if pictures are taken systematically such as taking pictures around the building (Figure 2). The practical implementation of going round the building can sometimes be defined by the nature of the building in many cases. The tower in Figure 4 is a narrow but high building. By choosing the more favourable and simpler way when photographing, pictures are taken 'up and down' alongside the four sides. It is worth noting that while taking pictures of the tower (Figure 4) the photos taken from position 'A' of Figure 1, i.e. above the top of the tower, are missing. However, 3D construction was successful as the pictures taken of the entire roof structure from positions 'B-C' of Figure 1 have the necessary overlapping.



Figure 3: Part of the reconstructed chapel under the eaves full of serious flaws.

In an extreme case pictures randomly taken are also suitable for 3D reconstruction provided they meet the requirements outlined above. At the same time, in the case of random photographs part of an object may sometimes be missing or not apparent in several pictures so that is why systematic photography is recommended.



Figure 4: The process of taking pictures of a high tower.

Figure 5 presents the 3D model of such a building where parts non-visible from above are significantly represented. As a result of applying position 'D' of Figure 1 the vaulted joist of the passage

of the building is also visible (Figure 5 (a)). Of course, this part of photography calls for utmost care as flights should be carried out extremely low and close to the building. In cases if the suspension of the camera system of the robot aircraft makes it possible, pictures can be taken by the manually rotating the multicopter (as if it was a handycam). It is important that the pictures taken 'manually' should be taken by a similar camera system and image fittings must be flawless.

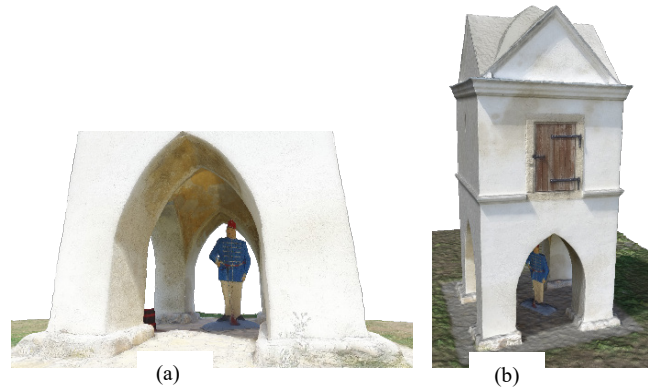


Figure 5: Display of parts of the historical tower non-visible from above.

The building to be photographed cannot always be freely accessible. If the walls of the building are at least laterally visible it is possible to make a correction of limitedly removing the trees or other covering object while processing.



Figure 6: The tree next to the wall of the building impedes photography.

The tree on Figure 6 does not make it possible for the multicopter to make a picture of the wall of the building freely. At the same time, however, the tree does not over the wall laterally so part of the wall covered from the front can be constructed from the necessary number of lateral photos. If no correction is made in processing, the 3D model will contain the tree and the wall, as well. By making use of the fact that the entire wall can be reconstructed the point cloud making up the tree can be removed during the 3D edition. It is important that such corrections be made on the point cloud serving as the basis of the model and not on the model ready. The 3D model of Figure 7 was made after the digital removal of the tree. The shadow of the tree is visible on the model but the tree itself no longer covers the building, so it is accessible.



Figure 7: The result of digitally removing the tree in the 3D model.

The spatial point cloud generated while processing is connected with elementary triangles by the programme. These triangles make up the surface, which is the 3D model of the area already surveyed. Of course, the number of polygons or any other means to detail the surface can be stated here. The ready-made 3D model can be displayed from a discretionary point. The model contains only surface, i.e. the polygons covering the surface have no thickness. In a further step of processing texture can be fitted on the 3D surface. The texture is made by the programme by means of the photo analysed in the previous steps. If we take the chance of correcting the 3D point cloud then the automated use of the pictures serving as the basis of the texture is practical to be manually modified. In this concrete case removing the tree itself is not enough. To have the proper texture the parts depicting the tree must be excluded from the photos of the given area. In this case the program prepares the texture of the part concerned based on the pictures of the wall if they exist.



Figure 8: Ruins of a church from the Árpád era following retouching the neighboring trees.¹

¹Retrieved from <https://sketchfab.com/models/8dda3b73b7a14c2cb606b15ef75408c1>

²Retrieved from <https://sketchfab.com/models/6ee2ae74e48846c6a26e2884a51ac9c5>
www.astesj.com

The 3D surface model (Figure 8) is created by exporting the ready-made 3D model and one or more image files containing the texture (it depends on the size and the details of the model itself). All these files are necessary to display the 3D objects presented. Display itself is possible by using several programmes such as the freely accessible and downloadable MeshLab programme.



Figure 9: The mausoleum of the Luppa family in Pomáz.²

5. Further 3D models

Figure 9 presents the 3D reconstructed photo of the mausoleum of the Luppa family from Pomáz. The model of the mausoleum was made on the basis of 160 photos of 12 megapixels by using Agisoft Photoscan software. The pictures were taken by a DJI Inspire 1 drone within one take-off of 15 minutes of flight. By walking round it in the virtual space the demolished state of the mausoleum is visible. From an aesthetic point of view several graffiti ruin the sight of the monument, which are also displayed on the reconstructed photo. It is interesting to note that the graffiti mentioned above are advantageous from the point of view of 3D reconstruction as it is easier to find contrasting point pairs when fitting the parts of the picture together than in the case of homogenous whitewashed walls.



Figure 10: A deserted farmhouse near Nádudvar.³

Figure 10 presents the 3D reconstructed picture of a deserted farmhouse. The farmhouse is approximately 8,2 m high, 25,8 m long and 10,8 m width. The model of the house is based on 160 photos of 12 megapixels by using Agisoft Photoscan software. The pictures were taken by a DJI Inspire 1 drone within one take-off of 15 minutes of flight. When these photos were made, it was gloomy so the contrast of the reconstructed picture is a bit beyond the ones made in sunny weather but it proved to be advantageous for

³Retrieved from <https://sketchfab.com/models/5de27e3e8ad442d7bb4122fafcd268b6>

reconstruction as there were no shadows and lighting proved to be constant when going round the entire building. The model reconstructed is detailed enough to reflect the state of the house. The mortar coming off and the bricks used for the walls staying bare are visible.



Figure 11: A renovated windmill in Kulcs.4

Figure 11 presents the 3D reconstructed photo of a renovated windmill. The model of the windmill is based on 160 photos of 12 megapixels by using Agisoft Photoscan software. The pictures were taken by a DJI Inspire 1 drone within one take-off of 15 minutes of flight. When taking pictures, utmost care must be paid to the thin wooden structure of the mill as reconstructing thin parts is usually difficult.



Figure 12: Papd chapel.5

Figure 12 presents the 3D reconstructed picture of a dilapidated chapel. The chapel is approximately 15,7 m high, 13,3 m long and 7,3 m width. The model of the chapel is based on 285 photos of 12 megapixels by using Agisoft Photoscan software. The pictures were taken by a DJI Inspire 1 drone within two take-offs of nearly 30 minutes of flight. When the pictures were taken it was typically sunny but the sun was shaded by clouds for a short time. Due to this fact, both sunny and cloudy pictures were taken of the building, which proved to be useful in 3D reconstruction. As a result, the ready-made model is contrastive but free from the

distortions of heavy shadowing. Reconstructing the cross at the top of the church requires special care and reconstruction of proper quality needs up-close photos of the cross and a small part of the dome. To this end, pictures of the top of the tower were made while the multicopter went round the cross at a distance of 4 metres. Pictures were once taken from the bottom and from the top of the cross. Despite the up-close photos the manual filtering in the 3D point cloud was necessary during the reconstruction of the cross due to which the points created as faults by the programme were removed.

6. Results

Based on the experience of several experiments the requirements of successful 3D reconstruction can be outlined.

Regarding environmental conditions diffuse light is the best for reconstruction, which corresponds with slightly cloudy weather. Suitable pictures in direct sunlight can only be made if the object moderately reflects light of a non-homogenous, i.e. well-structured surface. The 3D reconstruction of the shadowy side of objects with homogenous white surface such as several churches and chapels could only be carried out in significantly weaker quality (roughness of surface) in direct sunlight.

Regarding the cameras applied the parameter of the objectives is of great significance in addition to high resolution (of at least 12 Megapixels). Most pictures were taken by an objective of 20 mm working distance of, 12.4 Megapixels with 6.17x4.55 mm CCD sensor. In many cases very proximate flights were necessary to provide enough details. It can be proved by practice that experiments yield better results with an objective of 34 mm working distance of 16.0 Megapixels with 17.3x13.0 mm CCD sensor. In case of the latter one there is no need for dangerously approaching the object to have the necessary details.

Follow-up filtering proved to be necessary for several objects. In all cases filtering meant removing excess or disturbing points of the 3D point cloud. In some cases parts of the pictures deleted had to be removed, as well. However, it is important to note that the preliminary image improvement of the pictures used for the reconstruction significantly damaged the final result of the reconstruction. That is why exclusively unprocessed pictures could be used for processing.

The 3D models of the paper and further models of drone pictures are accessible at <https://sketchfab.com/fuhur>.

References

- [1] J. Albrecht, Albrecht Meydenbauer - Pioneer of Photogrammetric Documentation of the Cultural Heritage, Proceedings 18th International Symposium CIPA 2001, pp. 19-25, Potsdam (2001).
- [2] M. Marghany, M. R. B. M. Tahar, M. Hashim, "3D stereo reconstruction using sum square of difference matching algorithm", Scientific Research and Essays, Vol. 6(30), 6404-6423 (2011) [doi:10.5897/SRE11.1661].
- [3] Ildiko Horvath, Anna Sudar, „Factors Contributing to the Enhanced Performance of the MaxWhere 3D VR Platform in the Distribution of Digital Information”, Acta Polytechnica Hungarica Vol. 15, No. 3, 2018 [doi: 10.12700/APH.15.3.2018.3.9].
- [4] Borbála Berki, 2D Advertising in 3D Virtual Spaces”, Acta Polytechnica Hungarica Vol. 15, No. 3, 2018doi: 10.12700/APH.15.3.2018.3.10].

⁴Retrieved from <https://sketchfab.com/models/ed87fd62fc1b4e1a90293192355c8a65>
www.astesj.com

⁵Retrieved from <https://sketchfab.com/models/5ea66604b8a2462ebc499887c9fc2f08>

- [5] Şasi, A , Yakar, M., “Photogrammetric Modelling of Hasbey Dar’ülhuffaz (Maşjid) Using An Unmanned Aerial Vehicle”, *International Journal of Engineering and Geosciences*, 3 (1), 6-11 (2018) DOI: 10.26833/ijeg.328919
- [6] Doğan, Y, Yakar, M., “Gis and Three-Dimensional Modeling for Cultural Heritages”. *International Journal of Engineering and Geosciences*, 3 (2), 50-55 (2018). DOI: 10.26833/ijeg.378257
- [7] H. Gernsheim, "The 150th Anniversary of Photography," in *History of Photography*, Vol. I, No. 1, (1977).
- [8] A. Grimm, "Der Ursprung des Wortes Photogrammetrie", in *Internationales Archiv für Photogrammetrie*, Ackermann, F. et al., Ed., Vol. XXIII, Teil B10, Komm. V, VI, VII Nachtrag, pp. 323-330, Hamburg (1980).
- [9] N. A. Matthews, *Aerial and Close-Range Photogrammetric Technology: Providing Resource Documentation, Interpretation, and Preservation*. Technical Note 428. U.S. Department of the Interior, Bureau of Land Management, National Operations Center, 42 pp., Denver, Colorado (2008).
- [10] C. Harris, M. Stephens, "A combined corner and edge detector", in *4th Alvey Vision Conference*, Proc. AVC88, 147-151 (1988) [doi:10.5244/C.2.23].
- [11] A. Willis, Y. Sui, An Algebraic Model for fast Corner Detection, in *12th International Conference on Computer Vision*, Proc. ICCV.2009, 2296-2302 (2009) [doi:10.1109/ICCV.2009.5459443].
- [12] Ulvi, A , Toprak, A., “Investigation of Three-Dimensional Modelling Availability Taken Photograph of The Unmanned Aerial Vehicle; Sample Of Kanlidivane Church”, *International Journal of Engineering and Geosciences*, 1 (1), 1-7 (2016) DOI: 10.26833/ijeg.285216

Student Performance Evaluation Using Data Mining Techniques for Engineering Education

Veena Deshmukh^{*1}, Srinivas Mangalwede¹, Dandina Hulikunta Rao²

¹Research Centre Gogte Institute of Technology Belgaum, India,

²Cambridge Institute of Technology Bengaluru, India

ARTICLE INFO

Article history:

Received: 13 September, 2018

Accepted: 01 November, 2018

Online: 10 November, 2018

Keywords :

Mamdani Fuzzy Inference System

Scoring Rubrics Tool

Bloom's Taxonomy

ABSTRACT

In this research work, we are implementing a student performance evaluation model using Mamdani Fuzzy Inference System (FIS) and Neuro Fuzzy system and comparing the results with classical averaging method for Network Analysis (NA) course studied by third semester Electronics and Communication Engineering students. This work explains the designing of scoring rubrics using Bloom's levels as the criteria of assessment for NA course. Also at initial stages of learning how students' strengths and weaknesses can be identified using rubrics and develop critical thinking skills. The five inputs identify, understand, apply, analyze and design/create are five levels of learning as per Bloom's Taxonomy. Fuzzy rules are applied and the evaluated results are expressed in both crisp and linguistic variables and compared with classical aggregate scores.

1. Background and Motivation

Performance evaluation is always a challenge for a tutor after a learner undergoes a learning process or training program. Earlier the final grade for any course would mean the averaging of internal assessment scores and semester end examination scores. Later different activities like quiz, mini projects, seminars etc. were introduced in the course learning process to inculcate the higher order or critical thinking skills. In this process to the averaging method fails to highlight on the learners' critical thinking abilities. Also take a case of two students with same average marks but scores of both are like this, first student 65,75, 85 and second student 85,75,65. Here the average of the two students is same but first student is continuously improving the performance where as the second student is not. But a simple averaged grading method at the end of the course may not through light on all these issues of understanding, improvement or attainment of higher order thinking skills etc. [1]. The grades of evaluation of learners should in an ideal situation reflect their understanding of the course learnt, application of it in solving challenges of real life situations or solving similar issues by modifying the existing solutions or developing a new solution. Instead, continuous assessment of learner on daily/weekly basis is performed for given tasks and if a rubric with Bloom's criteria for evaluation on one side of the table and scores on the other side for

predefined criteria is designed to evaluate the critical thinking skills for each of the tasks assigned then the rubrics scores clearly indicate the strengths and weakness of learner, levels of critical thinking skills attained, understanding of the course on daily/weekly basis [2,3]. The scores so obtained are then subjected to suitable soft computing evaluation technique where each level attained by learner is weighted. The final grade so obtained should give details of the rubrics scores highlighting these thinking abilities of a learner for the course learnt, the activities performed during the training period. Whenever the rubrics scores for daily/weekly (formative) assessments are discussed with the learner the learner understands the natural thinking abilities and they can be strengthened and if the learner is willing to work upon weak areas for improvement, then he/she can take guidance from the tutor and improve upon them. The rubrics scores for every task assigned can give early indications of strengths and weakness' to both learner and tutor. Such formative assessments provide scope for improvement in final grades and skills of learner and even tutor can plan new learner centric teaching methodologies.

The scoring rubrics [2,3,4] can be designed for each activity stating clearly the evaluation criteria/objectives to attain the standards set for the said activity. They can be used for formative and summative assessments. The grading is unbiased, more objective and the degree of attainment is easily understood by

^{*}Veena Deshmukh, KLSGIT Belagavi Karnataka, India, Contact No: +919916508826 & vbdeshmukh@git.edu

both learner and tutor. The rubrics can be modified during learning phase if required to improve standard of learning. The NA course demands identifying different circuit elements, understanding and applying Kirchhoff's Laws and finally calculating the various currents, voltages and designing equivalent networks. The different levels set to evaluate critical thinking skills are identify, understand, apply, calculate and create. The Bloom's Taxonomy levels are set as standards for evaluation.

Bloom's Taxonomy was designed in 1965 by Dr. Benjamin Bloom [5,6] to improve learning standards in higher education. As per this, three domains of learning are described. Namely Cognitive - Knowledge, Affective- Attitude and Psychomotor-Skills. The cognitive domain involves the development of intellectual skills and knowledge. This is further divided into six orderly categories. These are commonly referred as Bloom's levels and are in the increasing order of complexity and thinking abilities. Now it is referred globally to evaluate critical thinking skills, for defining course objectives, designing cognitive level course activities and tests, setting question papers at different levels etc.

In this work we have tried to overcome the drawbacks of averaging method by evaluating learner performance using scoring rubrics tool designed for Network Analysis course. The rubric is designed keeping the critical requirements of the said course like identifying different electrical elements of a network, analyzing the network for its loop and branch currents, node and branch voltages, related power calculations etc. Later, making use of this knowledge in creating or building new simple solutions for real life scenario. Also while designing rubrics tool to evaluate the learners' for critical thinking skills, Bloom's levels or higher order thinking skills are set as criteria/bench mark for evaluation. After designing the rubrics, it is discussed with the learners so that they understand the different criteria for evaluation. The score/grade card mentions the rubrics scores obtained indicating different levels attained by each learner for different tasks completed along with average/weighted average grade of performance (it may be CGPA -cumulative grade point average or percentage of scores). The soft computing techniques are investigated for suitability to evaluate the performance [7,8]. The Neuro-fuzzy (NF), Fuzzy inference system (FIS) are tested and performances are compared with classical averaging method for student performance evaluation. .

Professor L. A. Zadeh invented the concept of Fuzzy Logic in 1964. In 1974, Mamdani developed first fuzzy logic controller which is used in predicting results when data is imprecise, vague or some data is missing. Fuzzy controllers are widely used in forecasting weather, stock market, product market, health monitoring systems, aviation systems, temperature and pressure controllers in manufacturing industries etc. It is rule based and reliable.

2. Literature Review

A good number of researches have been reported since 1995 on student performance evaluation using soft computing techniques. Authors have proposed a fuzzy logic based model for performance evaluation of Network Analysis course [1,2] . The early work by Biswas used fuzzy sets for evaluation of students answer scripts by

matching the answer scripts in 1995 [9]. But when large number of papers are to be evaluated this becomes tedious. Intelligent expert systems were implemented not using the complicated matching operation of answer scripts, instead [10] a more generalized method using degree of satisfaction and extended fuzzy grade sheets. A cricketer performance evaluation model was presented to predict international rank of a cricketer and also the effect of each input parameter like ranking of the teams being played, current ranking of the player, run rate etc are evaluated and rank of cricketer is predicted. Also the effect of each parameter on performance is discussed [11]. The drawbacks of conventional method of evaluation performed in universities considering higher weightage to attendance is discussed in [12] also a new method using fuzzy logic considering student attendance as one of the parameters along with internal and external marks as input parameters is implemented. A personalized student evaluation method is presented in comparison to the back propagation and conventional statistical methods in [13]. Every student is unique and fuzzy systems can make decisions and evaluate student performance along with students learning progress Such a unique model is presented in [14]. In [15] the authors opine that the performance of the students depends upon their previous performance, medium of instruction and type of board affiliated to using WEKA tool and compare the results of various classifiers. There are several factors which affect the performance of students, [16] authors use the combination of Genetic Algorithm and Artificial Neural Networks to predict the performance and also to find the factors which influence the performance of students. In another paper authors predict performance considering gender, location of house, family income, medium of instruction along with previous semester grades, attendance [17]. The application of fuzzy inference system in predicting the traffic flows by considering traffic related parameters is discussed in [18]. Neuro fuzzy systems in which the different attributes are used to predict performance in crisp values. It also provides an alternate solution when data available is vague and imprecise and classify the students' performance into different categories [19,20]. Student performance evaluation and prediction can be done using Sugeno-type ANFIS architecture. The membership functions generated by ANFIS using the training data predict the student performance (21). Prediction and evaluation can also be done using different types of data mining algorithms like C5, CART, ANN, SVM (22). Ensemble is one of the data mining algorithm used to classify recorded heart sound in (23). Classification using KNN is studied using certain distance algorithms such as Cosine, Correlation, Euclidean and City blocks in (24). By considering students' previous performance, motivation level, family background future performance is predicted in to different classes as Very good, Good and Poor (25). The data mining techniques are used to obtain the hidden knowledge about the student performance (26). The performance of different data mining algorithms are compared by considering their 'No of True positive' values (27).

3. Methodology

In this work we design the rubrics for Network Analysis course as given in Table 2. for third semester Electronics and Communication Engineering students (ECE). This course deals with study of different electrical circuit analysis and simplification techniques like Mesh and Node analysis, Network Theorems for

DC and AC circuits etc. This requires the basic knowledge of electrical elements, Ohm’s law and Kirchoff’s laws.

The steps involved are as follows:

- Designing of Rubric - the rubric is designed after consultation with course experts and students.
- Content delivery and Designing of question paper - the course content is delivered and question paper is designed as per Bloom;s taxonomy.
- Data collection - The test is conducted and scores of 41 students in rubric profile are collected.
- Evaluation using FIS and NF models - The data collected is investigated using the two models mentioned as shown in Figure 1.

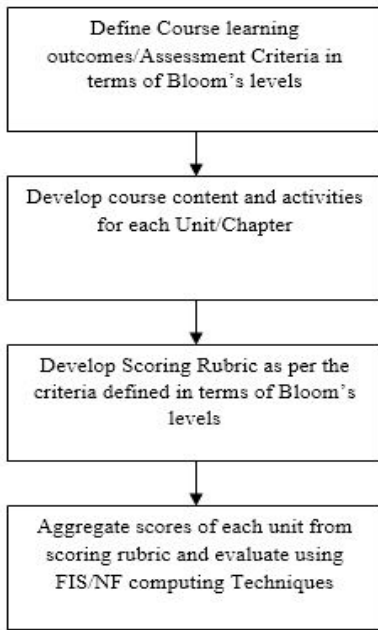


Figure 1. Designing of Scoring Rubric.

3.1. Fuzzy inference system (FIS)

The system which maps inputs to the outputs using fuzzy set theory is known as the Fuzzy Inference System. This can be either Mamdani or Sugeno. This system involves 3 steps, fuzzification of inputs using membership functions, formation of rule base using IF-THEN rules, defuzzification using output membership functions to get crisp values of outputs as shown in Figure 3. The membership functions can be chosen depending upon the requirement from the set of triangular, trapezoidal for linear variations and Gaussian, sigmoidal, zigmoidal etc.for nonlinear variations. We have used trapezoidal membership functions for both inputs and outputs as shown in Figure 2. Different weights are assigned to lower and higher order skills so that drawback of classical aggregation of scores is overcome. For example the scores above 8 are considered excellent for lower order skills like Identify whereas scores above 7 are considered excellent for hogher order skills like Apple and Create as shown in Table 2. The system can be made robust and flexible with the help of rules (13).

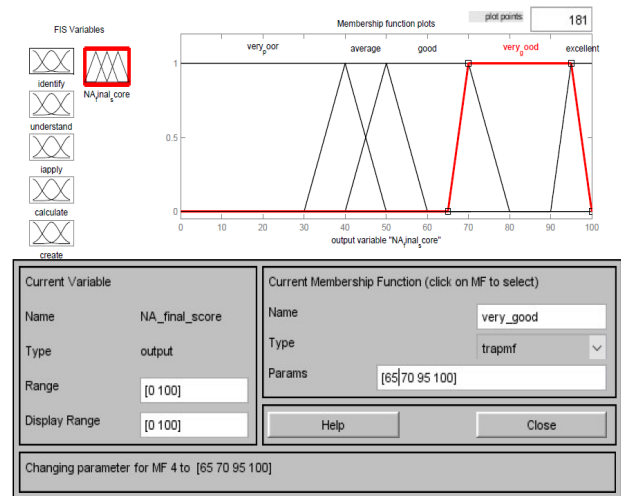


Figure 2. Membership functions for output.

Table 1. Input Variables

Input Variable	Evaluation Criteria
F1	Identify
F2	Understand
F3	Apply
F4	Calculate
F5	Create

3.2. Neuro-fuzzy (NF)

The hybrid of fuzzy inference system and neural networks is Neuro- Fuzzy (NF) system as shown in Figure 4. It combines the advantages of ANN and FIS in terms of ability to learn and think respectively making any system intelligent and think like human beings. This hybrid system basically is a neural network and is trained to generate IF-THEN fuzzy rules and membership functions of fuzzy systems. It is possible to incorporate common sense, intelligence and knowledge into the structure of neuro-fuzzy systems. The neuro-fuzzy system consists of 5 layers. The first one is the input and last one is the output layer and remaining three are the hidden layers responsible for generating membership functions, calculations and normalization (10).

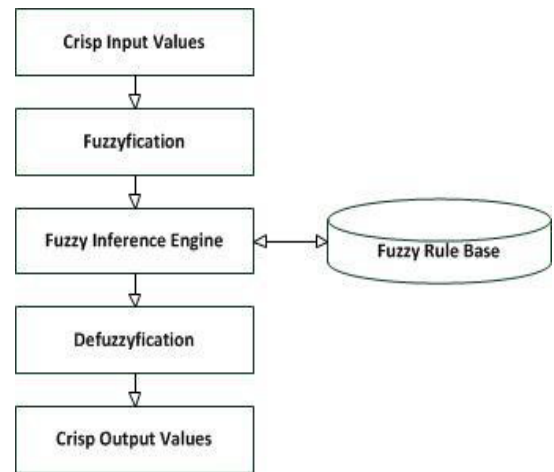


Figure 3. Basic flow of FIS

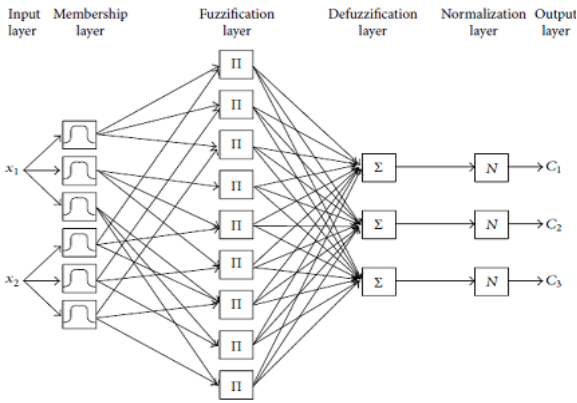


Figure 4: Basic block diagram of NFZ

4. Results and Discussion

We have studied the rule based classifiers like Fuzzy (2), Neuro fuzzy systems which give 100% accuracy for classification. The crisp values for results are compared with classical averaging method. The initial time taken for these two systems is more if the data set is large for forming rule base in case of FIS or training the system in case of NF. Neuro fuzzy system and Sugeno fuzzy system are used with input variables as identity, understand, apply, calculate, create as shown in Figure 4. and Figure 5. The output variables indicate the performance of student's in linguistic variables and are expressed as Poor<5, Average>5, Good (6.0-6.9), Very good (7.0-8.90) and Excellent (>9.0) and are given in Table 4 Final results in Crisp values and linguistic variable are shown in Figure 6. The comparison line graph of FIS, NF and classical averaging are shown in Figure 7. Comparison of different types of classifier results are as shown in Table 5.

Table 2. Final Scores in Linguistic Variables

Final score of NA(linguistic variable)	very poor	average	good	very good	excellent
Crisp value	<5.0	5.0-5.90	6.0-6.90	7.0-8.90	>9.0

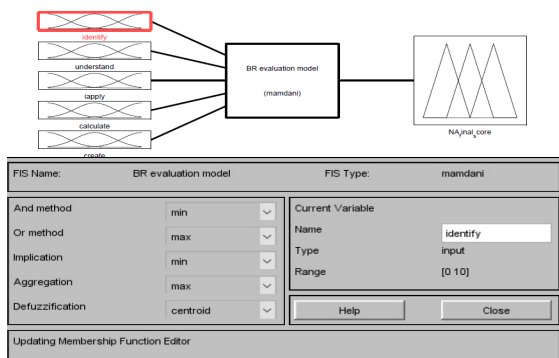


Figure 5. Fuzzy Inference System

5. Conclusions and future work

In this work we have compared the performance of Fuzzy Inference system and Neuro Fuzzy system with classical average

scoring method. All these classifiers can be used to evaluate the student performance depending upon the accuracy of classification and time taken to classify. Depending upon the students attribute their performances are evaluated and classified into 5 classes as very poor, average, good, very good and excellent. We have compared the performance based on factors such as training time, accuracy of the classifier performance. Fuzzy inference and Neuro fuzzy systems give 100% classification accuracy and the results are comparable to classical averaging method. But if data is large then the training period is more and formation of rules, selection of membership functions lead to complexity. Depending upon the data size, accuracy required and training time constraint one can choose any of the above classifiers for performance evaluation. We would like to evaluate student performance using other soft computing techniques like Support Vector Machine, K- nearest Neighbour, Ensemble and Discriminant analysis in future.

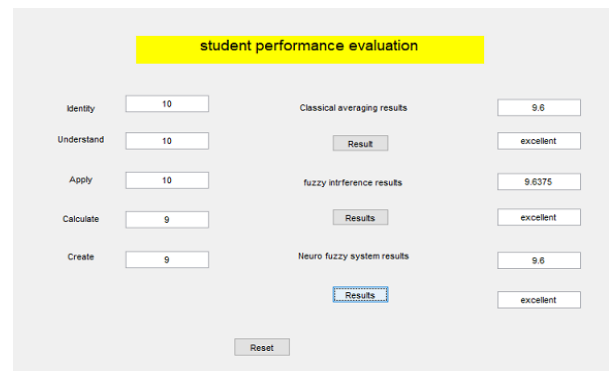


Figure 6. Result Comparison of different type of rule based classification

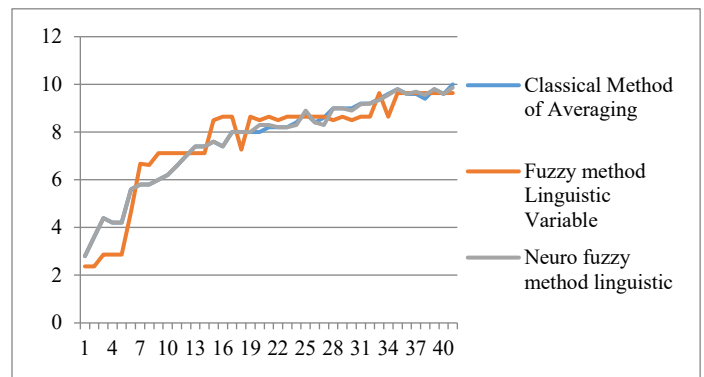


Figure 7. Observation Comparison Between the classical method, fuzzy and Neuro fuzzy system

Table 3. Different scores assigned for fuzzification of input variables

Input variables	unsatisfactory	satisfactory	good	very good	excellent
Identify	<5	5.0-5.9	6.0-6.9	7-7.9	>8.0
Understand	<5	5.0-5.9	6.0-6.49	6.5-7.49	>7.5
Apply	<5	5.0-5.49	5.5-5.9	6-6.9	>7.0
Calculate	<5	5.0-5.9	6.0-6.49	6.5-7.49	>7.5
Create	<5	5.0-5.9	6.0-6.49	6.5-6.9	>7.0

Table 4. Rubrics for Network Analysis Course

CLO or Criteria	Unsatisfactory(1-2)	Satisfactory(3-5)	Good(6-7)	Very Good(8-9)	Excellent(10)
Identify	Only Identifies and describes few circuit elements	Identifies and describes all the circuit elements	Identifies and describes the elements and demonstrates a limited understanding of the problem and solution	Identifies and describes the elements and demonstrates complete understanding of the problem and partial solution	Identifies and describes the elements and demonstrates complete understanding of the problem and solution
Understand/Select	Does not understand the problem including nodes and meshes	Does understand the problem including nodes and meshes	Selects and expresses nodes or meshes	Selects and expresses nodes or meshes in terms of equations	Selects and expresses nodes or meshes in terms of equations with complete understanding
Apply(KVL/KCL)	Does not apply mathematical/scientific principles/ laws	Applies laws with errors	Applies the mathematical/scientific principles/ laws with some errors and calculation mistakes	Applies the correct mathematical/scientific principles/ laws and all calculations are precise and not always appropriate	Applies the correct mathematical/scientific principles/ laws and all calculations are precise and appropriate
Calculate/ Compare	Fails to break down ideas/equations into simpler parts and analyze	Breaks down few ideas/equations into simpler parts	Breaks down ideas/equations into simpler parts, fails to analyze and compare/examine correctly	Breaks down ideas/equations into simpler parts analyzes and compares/examines sometimes correctly	Breaks down ideas/equations into simpler parts analyzes and compares/examines correctly
Summarize/Explain /Design	Does not acquire the knowledge in designing/developing electrical circuits	Acquires limited knowledge in only designing electrical circuits	Acquires limited knowledge in designing/developing electrical circuits	Acquires knowledge in only designing or developing electrical circuits	Acquires knowledge in designing and developing electrical circuits

Table 5.Student Data set.

Sl. No.	Identify	Understand/ Select	Apply	Calculate/ Compare	Create/ explain	Classical Method of Averaging	Fuzzy method results	Neuro fuzzy method results
1.	4	4	2	2	2	2.8	2.36944	2.80000
2.	4	4	3	3	4	3.6	2.36944	3.59999
3.	5	5	4	4	4	4.4	2.86609	4.39999
4.	5	5	3	4	3	4.2	2.86609	4.19997
5.	5	5	4	3	4	4.2	2.86609	4.19998
6.	7	7	5	4	4	5.6	4.64815	5.59999
7.	7	7	7	5	3	5.8	6.67544	5.79997
8.	8	8	5	6	2	5.8	6.62000	5.79999
9.	8	7	8	3	4	6.0	7.11905	5.99998
10.	6	7	6	6	6	6.2	7.11905	6.19995
11.	7	7	6	6	7	6.6	7.11905	6.59992
12.	7	7	7	7	7	7.0	7.11905	7.00001
13.	7	8	7	8	7	7.4	7.11905	7.40000
14.	8	7	8	7	7	7.4	7.11905	7.39999
15.	10	10	6	6	6	7.6	8.50000	7.59994
16.	10	10	6	6	5	7.4	8.64643	7.40000
17.	8	9	8	9	6	8.0	8.64643	7.99995
18.	8	8	8	8	8	8.0	7.26757	8.00000
19.	8	9	9	7	7	8.0	8.64643	8.00007
20.	10	9	10	7	8	8.0	8.50000	8.30003
21.	9	9	9	7	7	8.2	8.64643	8.29991
22.	10	10	7	7	7	8.2	8.50000	8.19828
23.	10	10	9	6	6	8.2	8.64643	8.20001
24.	9	9	9	8	7	8.4	8.64643	8.29991
25.	10	10	10	7	7	8.8	8.64643	8.89706
26.	10	10	10	6	6	8.4	8.64643	8.39996
27.	10	9	10	7	7	8.6	8.64643	8.30003
28.	9	10	9	7	8	9.0	8.50000	8.99999
29.	10	9	9	8	9	9.0	8.64643	8.99997
30.	10	10	10	7	8	9.0	8.50000	8.89706
31.	10	10	9	10	7	9.2	8.64643	9.17178
32.	10	9	9	10	8	9.2	8.64643	9.19997
33.	10	10	8	9	10	9.4	9.63750	9.34977
34.	10	10	10	8	10	9.6	8.64643	9.56348
35.	10	10	10	9	10	9.8	9.63750	9.80072

36.	10	10	9	10	9	9.6	9.63750	9.60004
37.	10	10	9	9	10	9.6	9.63750	9.69827
38.	10	10	9	9	9	9.4	9.63750	9.55849
39.	10	10	10	10	9	9.8	9.63750	9.80007
40.	10	10	9	10	9	9.6	9.63750	9.60004
41.	10	10	10	10	10	10	9.63750	9.86492

References

[1] V. B. Deshmukh, S. R. Mangalwede and Rao, D. H. "Student performance evaluation model based on scoring rubric tool for network analysis subject using fuzzy logic." In International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECOT), 2017, pp. 1-5. IEEE, 2017.

[2] V. B. Deshmukh, S. R. Mangalwede and Rao, D. H. "Student performance evaluation model based on Bloom's Taxonomy using fuzzy logic." In International Conference on Electrical, Electronics, Communication, Computer, Mechanical and Computing (EECCMC). IEEE, 2018.

[3] Srikant Jachderla, "A short note on Revised Bloom's taxonomy" Seer Academi and Global University 2014.

[4] Kai Meng Tay, Chwen Jen Chen and Kim Khoon Lee, "Application Of Fuzzy Inference System To Criterion-Referenced Assessment With A Case Study", Proceedings of the 2nd International Conference of Teaching and Learning (ICTL 2009) INTI University College, Malaysia

[5] Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). "Taxonomy of educational objectives: The classification of educational goals", Handbook 1: Cognitive domain.

[6] David R. Krathwohl. "Theory Into Practice", Volume 41, Autumn 2002 The Ohio State University.

[7] Brenda Sugrue, October, 2002 "Problems with Bloom's Taxonomy".

[8] Zadeh, L.A. 1965, "Fuzzy sets and systems", In J. Fox, editor, System Theory. New York: Polytechnic Press, pp. 29-39.

[9] Biswas, "An application of fuzzy sets in students' evaluation" 1995, Fuzzy sets and systems pp.187-194.

[10] Shyi-Ming Chen, Chia-Hoang Lee, "New methods for students' evaluation using fuzzy sets", Fuzzy sets and Systems 1999, pp.209-218.

[11] Singh, G., Bhatia, N., and Singh, S. "Fuzzy Logic Based Cricket Player Performance Evaluator". IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications" 2011, pp. 11-16.

[12] Meenakshi N., Pankaj N., "Application of Fuzzy Logic for Evaluation of Academic Performance of Students of Computer Application Course", IJRASET 2015. Volume 3 Issue X, October 2015 ISSN: 2321-9653 .

[13] Nidhi Arora, Jatinder Kumar R. Saini. "A Fuzzy Probabilistic Neural Network for Student's Academic Performance Prediction", International Journal of Innovative Research in Science, Engineering and Technology, 2013, Vol. 2. 4425-4432.

[14] Z.Jeremic, J.Jovonovic, D.Gasevic, "Student modeling and assessment in intelligent tutoring software patterns", Expert systems with Applications, Elsevier, 2012 pp.210-212 .

[15] R. Sumitha, E.S. Vinothkumar, "Prediction of Students Outcome Using Data Mining Techniques", International Journal of Scientific Engineering and Applied Sciences , Volume-2, Issue-6, June 2016 ISSN: 2395-3470.

[16] Omar Augusto Echegaray-Calderon, Dennis Barrios-Aranibar, "Optimal selection of factors using Genetic Algorithms and Neural Networks for the prediction of students' academic performance", 978-1-4673-8418-6/15/\$31.00 ©2015 IEEE.

[17] Gurmit Kaur, Williamjit Singh, " Prediction of Student Performance Using WEKA Tool", An International Journal of Engineering Sciences, 2016, Vol 17, ISSN: 2330-0332 (online).

[18] Deshpande, Minal, and Preeti R. Bajaj. "Short term traffic flow prediction based on neuro-fuzzy hybrid sytem." In ICT in Business Industry & Government (ICTBIG), International Conference on, pp. 1-3. IEEE, 2016.

[19] Isa, K., Mohamad, S. and Tukiran, Z., 2008, August. Development of INPLANS: An Analysis on Students' Performance using Neuro-Fuzzy. In Information Technology, ITSIM 2008. International Symposium on (Vol. 3, pp. 1-7). IEEE. 2008

[20] Sevarac, Zoran. "Neuro fuzzy reasoner for student modeling." In Advanced Learning Technologies, 2006. Sixth International Conference on, pp. 740-744. IEEE, 2006.

[21] Taylan, Osman, and Bahattin Karagözoğlu. "An adaptive neuro-fuzzy model for prediction of student's academic performance." Computers & Industrial Engineering 57, no. 3 (2009): 732-741

[22] Agaoglu, Mustafa. "Predicting instructor performance using data mining techniques in higher education." IEEE Access 4 (2016): 2379-2387.

[23] Homsí, Masun Nabhan, Natasha Medina, Miguel Hernandez, Natacha Quintero, Gilberto Perpiñan, Andrea Quintana, and Philip Warrick. "Automatic heart sound recording classification using a nested set of ensemble algorithms." In Computing in Cardiology Conference (CinC), 2016, pp. 817-820. IEEE, 2016.

[24] Kataria, Aman, and M. D. Singh. "A review of data classification using k-nearest neighbour algorithm." International Journal of Emerging Technology and Advanced Engineering 3, no. 6 (2013): 354-360

[25] Do, Quang Hung, and Jeng-Fung Chen. "A neuro-fuzzy approach in the classification of students' academic performance." Computational intelligence and neuroscience 2013 (2013): 6.

[26] Amra, Ihsan A. Abu, and Ashraf YA Maghari. "Students performance prediction using KNN and Naïve Bayesian." In Information Technology (ICIT), 2017 8th International Conference on, pp. 909-913. IEEE, 2017.

[27] Troussas, Christos, Maria Virvou, and Spyridon Mesaretzidis. "Comparative analysis of algorithms for student characteristics classification using a methodological framework." In Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on, pp. 1-5. IEEE, 2015.

Parallelizing Combinatorial Optimization Heuristics with GPUs

Mohammad Harun Rashid*, Lixin Tao

Pace University, New York, USA

ARTICLE INFO

Article history:

Received: 12 August, 2018

Accepted: 09 November, 2018

Online: 18 November, 2018

Keywords:

GPU

Combinatorial

Optimization

Parallel

Heuristics

ABSTRACT

Combinatorial optimization problems are often NP-hard and too complex to be solved within a reasonable time frame by exact methods. Heuristic methods which do not offer a convergence guarantee could obtain some satisfactory resolution for combinatorial optimization problems. However, it is not only very time consuming for Central Processing Units (CPU) but also very difficult to obtain an optimized solution when solving large problem instances. So, parallelism can be a good technique for reducing the time complexity, as well as improving the solution quality. Nowadays Graphics Processing Units (GPUs) have evolved supporting general purpose computing. GPUs have become many core processors, multithreaded, highly parallel with high bandwidth memory and tremendous computational power due to the market demand for high definition and real time 3D graphics. Our proposed work aims to design an efficient GPU framework for parallelizing optimization heuristics by focusing on the followings: distribution of data processing efficiently between GPU and CPU, efficient memory management, efficient parallelism control. Our proposed GPU accelerated parallel models can be very efficient to parallelize heuristic methods for solving large scale combinatorial optimization problems. We have made a series of experiments with our proposed GPU framework to parallelize some heuristic methods such as simulated annealing, hill climbing, and genetic algorithm for solving combinatorial optimization problems like Graph Bisection problem, Travelling Salesman Problem (TSP). For performance evaluation, we've compared our experiment results with CPU based sequential solutions and all of our experimental evaluations show that parallelizing combinatorial optimization heuristics with our GPU framework provides with higher quality solutions within a reasonable time.

1. Introduction

Combinatorial optimization is a topic that consists of finding an optimal solution from a finite set of solutions. Combinatorial optimization problems are often NP hard. It is often time consuming and very complex for Central Processing Unit (CPU) to solve combinatorial optimization problems, especially when the problem is very large. Metaheuristic methods can help finding optimal solution within a reasonable time.

Nowadays efficient parallel metaheuristic methods have become an interesting and considerable topic. Local search metaheuristics (LSMs) are single solution-based approaches, as well as one of the most widely researched metaheuristics with various types such as simulated annealing (SA), hill climbing, iterated local search, tabu search, and genetic algorithm etc. A common feature that local search metaheuristics can share is, a neighborhood solution is selected iteratively as a candidate

solution. We can use local search metaheuristics to solve combinatorial optimization problems such as graph bisection problem, Travelling Salesman Problem (TSP) etc. Many works have been done by using parallel computing technology to improve its performance from iteration level, algorithmic level and solution level. Therefore, parallelism is a good technique for improving the solution quality as well as reducing the time complexity.

Recently, Graphics Processing Units (GPUs) have been developed gradually to parallel/programmable processors from fixed function rendering devices. GPUs motivated by high definition 3D graphics from real time market demand have become many core processors, multithreaded, highly parallel with high bandwidth memory and tremendous computational power. So, more transistors are devoted to design GPU architecture in order to do more data processing than data caching/flow control. With the fast development of general purpose Graphics Processing Unit (GPGPU), some companies have promoted GPU programming

*Mohammad Harun Rashid, E-mail: harun7@yahoo.com

www.astesj.com

<https://dx.doi.org/10.25046/aj030635>

frameworks such as OpenCL (Open Computing Language), CUDA (Compute Unified Device Architecture), and direct Compute. GPU based metaheuristics have become much more computational efficient compared to CPU based metaheuristics. Furthermore, making local search algorithms optimized on GPU is an important problem for the maximum efficiency.

In the recent years, GPU computing has emerged as a very important challenge in the research areas for parallel computing. GPU computing is believed as an extremely useful technology for speeding up so many complex algorithms in order to improve solution quality. However, rethinking of existing parallel models as well as programming paradigms for allowing their deployment on GPU accelerators is one of the major challenges for metaheuristics. In fact the issue is, revisiting the parallel models as well as programming paradigms for efficiently considering the GPUs characteristics. However, some issues related to memory hierarchical management of GPU architecture need to be considered.

The contribution of this research is, to design a GPU framework which can efficiently deal with the following important challenges while parallelizing metaheuristics methods to solve large optimization problems:

1. Distribution of data processing between GPU and CPU with efficient CPU-GPU cooperation.
2. Thread synchronization and efficient parallelism control.
3. Data transfer optimization among various memories and memory capacity constraints with efficient memory management.

Our proposed parallel models on GPU architecture can be very useful and efficient in finding better optimized solution for large scale combinatorial optimization problems. We have made a series of experiments with our proposed GPU framework to parallelize some heuristic methods such as simulated annealing, hill climbing, genetic algorithm etc. for solving combinatorial optimization problems like Graph Bisection problem and Travelling Salesman Problem (TSP). All of our experimental evaluation shows that parallelizing heuristics methods with our GPU framework provides higher quality solutions in a reasonable computational time.

2. Background

As we contribute to the parallelization of heuristics methods for combinatorial optimization problems with GPU, below we discuss about combinatorial optimization problems, some optimization heuristics methods (hill climbing and simulated annealing), graph bisection problem/ travelling salesman problem as example optimization problems and GPU architecture/ computing.

2.1 Significance of Combinatorial Optimization

Combinatorial optimization can be defined as a mathematical discipline with the interplay between computer science and mathematics [1]. Very roughly, it deals with the problem of making optimal choices in huge discrete sets of alternatives. Combinatorial optimization is a way to search through a large

number of possible solutions for finding the best solution from them. When the number of possible solutions is really too large and it is impractical to search through them, we can apply different techniques to narrow down the set and speed up the search.

Many combinatorial optimization problems are known as NP hard. That means, the time needed for solving a problem instance to optimality grows exponentially with the size of the problem in the worst case. Hence, these problems are easy to understand and describe, but very hard to solve. Even it is practically impossible to determine all possibilities for problems of moderate size in order to identify the optimum. Therefore, heuristic approaches are considered as the reasonable way to solve hard combinatorial optimization problems. Hence, the abilities of researchers to construct and parameterize heuristic algorithms strongly impact algorithmic performance in terms of computation times and solution quality.

2.2 Combinatorial Optimization Heuristics

To deal with combinatorial optimization problems, the goal is to finding such an optimal/optimized solution, which can minimize the given cost function. The cost of the algorithms can exponentially increase while the complexity of the search space is growing up, and this can make the search of a solution not feasible.

Finding a suboptimal solution within a reasonable time is another way to address these problems. In some cases, We might even find the optimal solution in some cases. These techniques can be divided into two main groups: heuristics and metaheuristics.

A heuristic is an algorithm that tries to find optimized solutions to complex combinatorial problems, but there is no guarantee for its success. Most of the heuristics are based on human perceptions, understanding the problem characteristics and experiments, but they are not based on fixed mathematical analysis. The heuristic value should be based on comparisons of performance among the competing heuristics. The most important metrics for performance are quality of a solution, as well as the running time. The implementation of heuristic algorithms is easy and they can find better solutions with relatively small computational effort. However, a heuristic algorithm can rarely find the best solution for large problems.

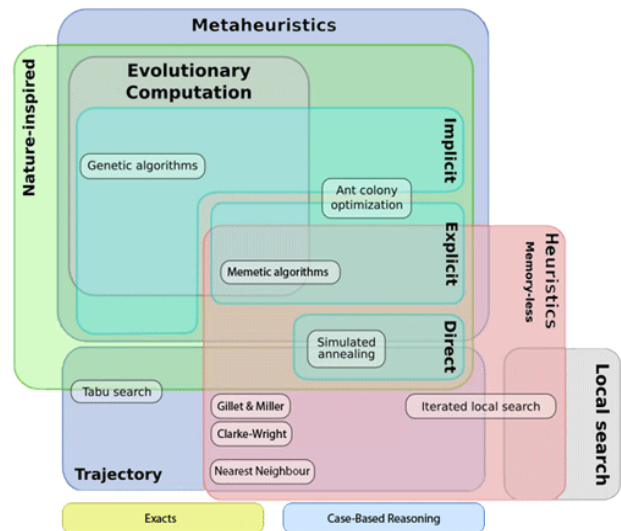


Figure 1: Different classifications of metaheuristics

Over the last couple of decades many researchers have been studying optimization heuristics for solving many real life NP hard problems, and some of the common problem solving techniques/methods underlying these heuristics came up as meta heuristics. A meta heuristic can be defined as a pattern or the major idea for a class of heuristics. Reusable knowledge in heuristic design can be represented by Meta-heuristics, which can provide us with important starting points in designing effective new heuristics for solving new NP hard problems.

Meta heuristics are not algorithms, nor based on theory. To effectively solve a meta heuristic based problem, we must have better understanding of the problem characteristics, and creatively designing as well as implementing the major meta heuristic components. So, it has become an action of research to use a meta heuristic for proposing an effective heuristic to solve an NP hard problem.

Another classification dimension is population based searches (P-Metaheuristics) vs single solution (S-Metaheuristics). P-Metaheuristics approaches maintain/ improve multiple candidate solutions, often using population characteristics to guide the search. P-Metaheuristics include genetic algorithms, evolutionary computation, and particle swarm optimization. S-Metaheuristics approaches focus on modifying/improving a single candidate solution. S-Metaheuristics include iterated local search, simulated annealing, hill climbing, variable neighborhood search etc. Below are some of the heuristics methods to solve combinatorial optimization problems:

2.2.1 Hill Climbing

Hill climbing is a mathematical optimization technique that belongs to the local search family and can be used for solving combinatorial optimization problems. The best use of this technique is in problems with “the property that the state description itself contains all the information needed for a solution”. Hill climbing algorithm is memory efficient, since it does not maintain any search tree. This algorithm mainly looks into the present state and the immediate future states only. By using an evaluation function, it tries to improve the current state iteratively.

Hill climbing can encounter a problem called “local maxima”. When the algorithm stops making progress towards an optimal solution, local maxima problem can occur because of the lack of immediate improvement in adjacent states. There are variety of methods to avoid Local maxima. Repeated explorations of the problem space could be one of the methods for solving this problem.

Hill Climbing Algorithm:

```

Get a random initial partition as the current partition.
While there is any improvement to the best cost seen so far
    Generate a random neighbor of the current partition.
    Evaluate the neighbor's cost.
    If the neighbor's cost improves the current cost
        Let the neighbor be the new current partition.
        If the neighbor's cost improves the best one seen so far, record it.
    End If.
End While.
Return the best partition visited.
    
```

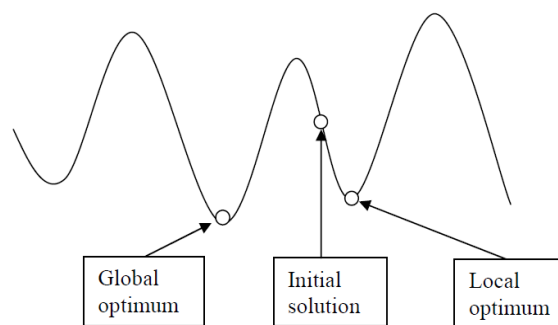


Figure 2: global solutions vs local solutions.

This algorithm starts from a random initial solution and then it keeps looking in the solution space in order to migrate to better neighbor solution. We might need to compare the current partition with all the neighbor's solutions before the algorithm is terminated. The algorithm terminates when all the neighbors' solutions are worse compared with the current partition. This technique can only find the local optimum solutions which are better solutions than all the neighbors, but the found solutions might not be global optimum solutions. The figure 2 shows difference between global and local solutions.

It is a time consuming process to maintain the visited neighbors of the current partition. That is why, it is necessary to parallelize the hill climbing algorithm for finding the best optimized solution.

2.2.2 Simulated Annealing

For graph bisection, simulated annealing heuristic starts with a high temperature t and a randomly selected initial partition as its current partition. After that, this heuristic starts the iterations with the same temperature and at each iteration, a neighbor partition is randomly generated. If the cost of the neighbor partition is better than the current cost, then the neighbor partition becomes the new current partition for the next iteration. If the neighbor partition does worsen the current cost, it can still be accepted with a probability as the new current partition. In case of high temperature, the probability is not sensitive to bad neighbor partition. But in case of low temperature, the probability for accepting a worsening neighbor will diminish with the extent of the worsening. The temperature is reduced by a very small amount after certain iterations are completed with the same temperature, and then, the iterations continue with the reduced temperature. The iteration process terminates once the termination criteria is met.

Many combinatorial optimization problems can be solved by applying simulated annealing heuristic. Unlike other meta heuristics, it has been mathematically proven that simulated annealing converges to the global optimum with sufficiently slow reduction of the temperature. As very few real world problems can afford such excessive execution time, this theoretical result does not interest much the practitioners. Below is the pseudocode for simulated annealing heuristic.

The simulated annealing and local optimization differ with the characteristics whether worsening neighbors will be accepted.

Simulated annealing heuristic starts with random walk in the solution space. If a random neighbor is better than the current solution, simulated annealing always accepts it. But, when the random neighbor is worse, the chance of accepting the worsen neighbor is slowly reduced. Simulated annealing can be reduced to local optimization with a very low temperature.

- Let $s = s_0$
- For $k = 0$ through k_{max} (exclusive):
 - $T \leftarrow \text{temperature}(k/k_{max})$
 - Pick a random neighbour, $s_{new} \leftarrow \text{neighbour}(s)$
 - If $P(E(s), E(s_{new}), T) \geq \text{random}(0, 1)$:
 - $s \leftarrow s_{new}$
- Output: the final state s

Figure 3: Simulated Annealing Algorithm

2.3 Optimization Problems

2.3.1 Graph Bisection Problem

The graph bisection problem can be defined as a data representation of a graph $G = (V, E)$ with a number of vertices= V and a number of edges= E , such that the graph G can be partitioned into smaller sections with some particular properties. For example, a k -way partition can divide the vertices into k smaller sections. When the number of edges between the separated components is relatively very small, it can be defined as a good partition. We can call a graph partitioning as uniform graph partition which divides the graph into smaller components in such a way that all the components are almost the same size as well as there are only small number of connections between the components. The important applications for graph partitioning include, but not limited to partitioning different stages for VLSI circuit design, scientific computing, clustering, task scheduling for multi-processing systems, and cliques detection in social networking etc.

Given a graph $G = (V, E)$ where $|V|$ is an even integer, find a partition of V into subsets L and R that minimizes the objective function

$$\text{cutSize}(L, R) = \sum_{(x,y) \in L \times R} \text{adj}(x, y)$$

under the constraint that $|L| = |R|$.

Here, $\text{adj}(x, y)$ is the adjacency matrix and cutSize represents the cost for the bisection of the given graph G .

A partition satisfying the above conditions is called an optimal solution to the problem.

Example: The following graph has been partitioned into two equal-sized subsets by a dotted line, and the partition has a cut size of 2, which is the minimal cut size possible. This partition is therefore called optimal.

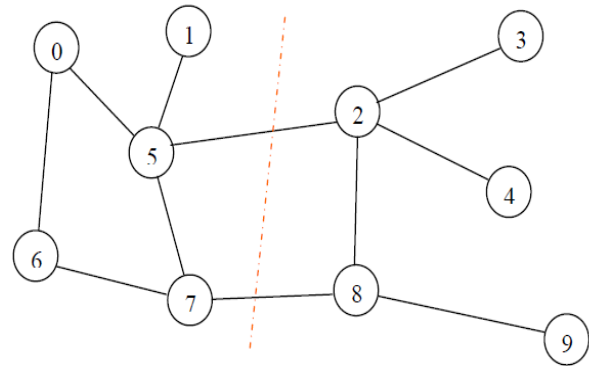


Figure 4: An optimal graph bisection

2.3.2 Travelling Salesman Problem

The Travelling Salesman Problem (TSP) is a combinatorial optimization problems which can be described easily, but it is very difficult to solve. A salesman starts with one city and will be visiting a number of cities with the condition that the salesman must visit each and every city only once and finally returns to first city. Selecting the sequence of the cities to be visited is the problem, because the salesman has to take the shortest path from a set of possible paths to minimize the path length. Exhaustive search can be used to find an optimal solution for a small instances (a few cities only) of TSP. But the problem is really critical for large number of cities, since with the increase in the number of cities, the number of possible paths increases exponentially. The number of possible paths for visiting n number of cities is the permutation of n which is $n!$. If the number of cities is increased by 1 only, the number of possible paths will become $(n+1)!$. Therefore, it will take too much time to compute the cost for all possible paths and find out the shortest path from them. TSP in known as a typical NP-hard problem.

TSP has a lot of applications in real world in different areas, like electronic maps, computer networking, Mailman’s job, VLSI layout, traffic induction, electrical wiring, etc.

TSP Algorithms:

TSP Greedy-Genetic Algorithm	
Step 1:	Generate a random initial population of chromosomes (travel path).
Step 2:	Calculate fitness (path length) of every chromosome of initial population.
Step 3:	Repeat (iterations). <ol style="list-style-type: none"> a) Apply local search heuristic to select parents from initial population for cross over. b) Apply a greedy algorithm to generate greedy children from parents. c) Calculate fitness of all newly generated children and add newly generated children in initial population. d) Sort combined population initial population + new children by path length and select best chromosomes (travel path) for next population. e) Set initial population = next population.
Until Terminating condition is satisfied (for given number of generations).	
Greedy-Children Algorithm	
input: an array parentPath containing n cities and an integer $k \leq n$:	
Step 1:	Copy $\text{parentPath}[1]$ into position 1 of a new array childPath of size n
Step 2:	Copy $\text{parentPath}[2, \dots, k]$ into a new list prefix .
Step 3:	Copy $\text{parentPath}[k+1, \dots, n]$ into $\text{childPath}[k+1, \dots, n]$
Step 4:	Set $\text{currentCity} = \text{parentPath}[1]$
Step 5:	For $i = 2$ to k : <ol style="list-style-type: none"> a) Extract from prefix the city nextCity at minimum distance from currentCity. b) Store nextCity in $\text{childPath}[i]$. c) Set $\text{currentCity} = \text{nextCity}$.
Step 6:	Return childPath .

2.4 GPU Architecture/Computing

Just a while ago, the conventional single core or multicore CPU processor was the only viable choice for parallel programming. Usually some of them were either loosely arranged as multicomputer in which the communication among them were done indirectly because of the isolated memory spaces, or tightly arranged as multiprocessors that shares a single memory space. CPU has a large cache as well as an important control unit, but it doesn't have many arithmetic logic units. CPU can manage different tasks in parallel which requires a lot of data, but data are stored in a cache for accelerating its accesses. Nowadays most of the personal computers have GPUs which offer a multithreaded, highly parallel and many core environments, and can potentially reduce the computational time. The performance of the modern GPU architecture is wonderful in regards to cost, power consumption, and occupied space.

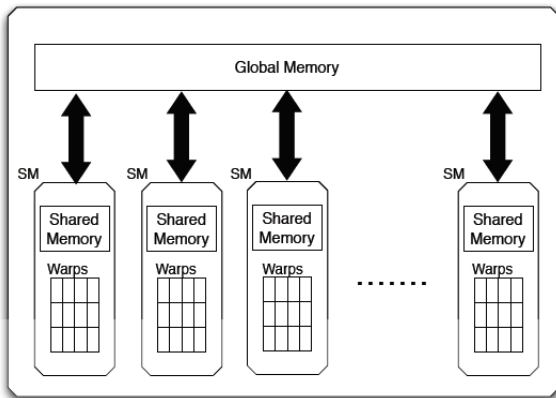


Figure 5: GPU architecture.

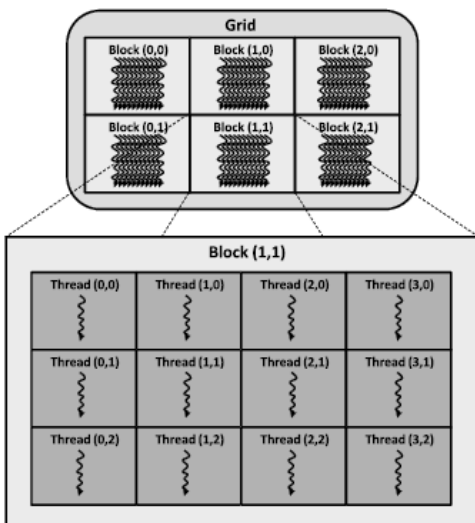


Figure 6: GPU thread blocks

A GPU includes a number of Streaming Multiprocessors (SMs). Each streaming multiprocessor contains a number of processing units which can execute thousands of operations concurrently. The warps inside a SM consist of a group of threads. A warp can execute 32 threads in a Single Instruction Multiple

Data (SIMD) manner, which means all the threads in a warp can execute same operation on different data points. GPUs have at least two kinds of memory: global memory and shared memory. Global memory allows to store a large amount of data (such as 8GB), whereas shared memory can usually store only few Kilobytes per SM.

A GPU thread can be considered as a data element to be processed. GPU threads are very lightweight in comparison with CPU threads. So, it is not a costly operation when two threads change the context among each other. GPU threads are organized in blocks. Equally threaded multiple blocks execute a kernel. Each thread is assigned a unique id. The advantage for grouping of threaded blocks is that simultaneously processed blocks are linked closely to hardware resources. The threads within the same block are assigned to a single multiprocessor as a group. So, different multiprocessors are assigned to different threaded blocks. Therefore, controlling the threads parallelism can be a big issue for meeting memory constraints. As multiprocessors are mainly organized based on the Single Program Multiple Data (SPMD) model, the threads can access to different memory areas as well as can share the same code.

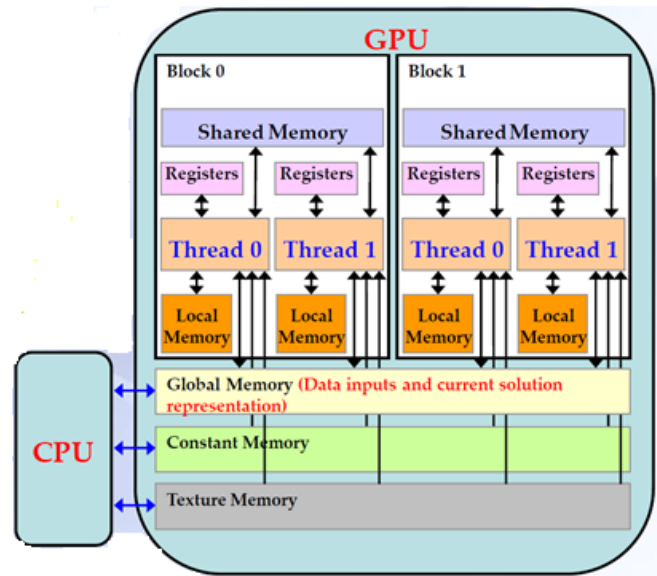


Figure 7: CPU - GPU communications.

GPU is used as a device coprocessor and CPU is used as a host. Each GPU has its own processing elements and memory which are separate from the host computer. Data is transferred between the host memory and the GPU memory during the execution of programs. Each device processor on GPU supports SPMD model, which means same program can simultaneously be executed on different data by multiple autonomous processors. To achieve this, we can define kernel concept. The kernel is basically a method or function which is executed by several processors simultaneously on the specified device in parallel and callable from the host as well. The Communications between CPU host and the device coprocessors are accomplished via the global memory.

3. Literature Review

In [2, 3, 4], the author proposed GPU based works with genetic algorithms. They proposed that the population evaluation as well as a specific mutation operator are to be performed in GPU. They implemented the selection and replacement operators in CPU. So, huge data transfers are performed between GPU and CPU. This kind of techniques can limit the performance of the solution.

In [5], the author proposed evolution strategy algorithm for solving continuous problems. According to his suggestions, multiple kernels can be designed for some of the evolutionary operators like selection, evaluation, crossover, mutation etc. and CPU can handle the rest of the search process. Later on, in [6], the author presented similar implementation with genetic algorithms. The additional contribution of their work was investigating the effect of problem size/ thread size /population size on GPU implementation comparing with sequential genetic algorithm.

In [7], the author proposed a memory management concept for an optimization problem. They implemented the concept for quadratic assignment problem where the global memory accesses were coalesced, the shared memory was used for storing as many individuals as possible, and the constant memory associated with matrices. For dealing with data transfers, their approach was a full parallelization based search process. In this regards, they divided the global genetic algorithm into multiple individual genetic algorithms, such that a thread block is represented by each sub population. Because of the poor management of data structures, the speed-ups obtained in their solution for combinatorial problems are not convincing.

In [8], the author proposed a framework of the automatic parallelization with GPU for the evaluation function. Only the evaluation function code need to be specified in their approach and the users don't need to know CUDA keywords. This approach allows evaluating the population on GPU in a transparent way. But, this strategy has some problems, such as it lacks flexibility because of transferring the data and nonoptimized memory accesses. In addition, the solution is limited to the problems where no data structure is required.

In [9], the author proposed an implementation of an evolutionary algorithm which is a GPU based full parallelization of the search process. Without any problem structures they implemented this approach to make an application for continuous and discrete problems. Later on, the authors also submitted an implementation of their algorithm with multi GPUs [10]. However, since there are some challenging issues of the context management such as global memories of two separate GPU, their implementation with multiple GPUs does not really provide with any significant performance advantages.

In [11], the author implemented a model for continuous optimization problems in which is very similar to the previous

model. In this model, shared memory is used to store each sub population and organized based on ring topology. Although the speed-ups for the obtained solution are better compared with a sequential algorithm, the implementation of this model was dedicated to few continuous optimization problems only. As by considering the two previous models no general methods were outlined, in [12], the author made some investigation on the parallel island model on GPU. By involving different memory managements, they designed three parallelization strategies and were able to address some issues.

In [13], the author proposed a multi start tabu search algorithm and implemented to the TSP as well as the flow shop scheduling problem. The parallelization is performed on GPU by using shared libraries, and one tabu search associated with each thread process. However, this approach requires so many local search algorithms for covering the memory access latency and so, this type of parallelization is not much effective. In [14], the author proposed similar approach with CUDA. In this approach, the memory management for optimization structure is done in the global memory and they implemented this to the quadratic assignment problem. However, as one local search associated with each thread, the solution performance is limited to the size of instance.

For designing of multi-start algorithms, in [15], the author provided general methodology which are applicable to local search methods like simulated annealing, hill climbing, or tabu search. They also have contribution regarding the relationship between available memories and data mostly used for the algorithms. But, the application of the GPU accelerated multistart model is very limited, because it requires so many local search algorithms at run time to become effective. In [16], the author proposed a GPU based hybrid genetic algorithm. In their approach, they implemented an island model where a cellular genetic algorithm is represented by each population. Also, the mutation step in their hybrid genetic algorithm is followed by hill climbing algorithm. They performed their implementation for the maximum satisfiability problem.

According to the previous work, the hill climbing need to be integrated with the island model as per the investigation of the full parallelization. In this regard, in [17] the author proposed the redesign of GPU based hybrid evolutionary algorithms which performs a hybridization with a local search. Their focus was on different neighborhoods generation on GPU, correlating to each individual to be mutated in the evolutionary process. This kind of mechanism may guarantee more flexibility.

In [18], the author introduced a GPU accelerated multi objective evolutionary algorithm. In his approach, he implemented some of the multi objective algorithms on GPU, but not with the selection of non-dominated solutions. There are more works on P-metaheuristics for GPU parallelization are proposed. The parallelization strategies used for these implementations are similar to the prior techniques mentioned above. These works include particle swarm optimization [19, 20, 21], genetic

programming [22, 23, 24, 25] and other evolutionary computation techniques [26, 27, 28].

3.1 Research issues and contributions

Most of the approaches in the literature are mainly based on either iteration level or algorithmic level. In other words, the approaches are based on basically either the simultaneous execution of cooperative/independent algorithms, or GPU accelerated parallel evaluation of solutions. Regarding the cooperation between CPU and GPU, there are some implementations which also consider the GPU parallelization of other treatments such as selection/variation operators for evolutionary algorithms. We may argue on the validity of these choices, since an execution profiling may show that such treatments are negligible compared to the evaluation of solutions. As mentioned above, for reducing the data transfer between GPU and CPU, a full GPU parallelization of metaheuristics may also be performed. The original semantics of the metaheuristic are altered in this case to fit the GPU execution model.

Regarding the control of parallelism, a single thread with one solution are associated in most of the implementations. Besides, some of the cooperative algorithms associate one threads block with one sub population and may take advantage of the threads model. However, so far we've not found any work that has been investigated for managing parallelism of the threads efficiently to meet the memory constraints. The previous implementations may not be robust while dealing with large problem instances or a large set of solutions. In Chapter 4, we will show how an efficient control of thread may allow introducing fault tolerance mechanism for GPU applications.

For the memory management, so far there is no explicit efforts made in most of the implementations for memory access optimizations. For example, one of the most important elements for speeding up is memory coalescing, and additionally, could consider local memories for reducing non coalesced accesses. However, some of the authors proposed the simple way of using the shared memory to cache, but there is no performance improvement guarantee in those approaches. Some other authors also put some explicit efforts for handling optimization structures with the different memories, but still there is no general guideline/outline from those works. In fact, a lot of time, the associations of memory strictly depend on the target optimization problem such as small problem instances and/or no data inputs.

The contribution of this research is: to design a GPU framework with a set of efficient algorithms which can efficiently address the above mentioned challenges by parallelizing major metaheuristics for combinatorial optimization problems on the CPU-GPU architecture and also, to validate the solution quality with graph bisection problem as well as Travelling Salesman Problem (TSP).

4. Proposed Method and Contribution

Below are the research methodologies that we followed for proposing our methods and adding contributions to our GPU framework.

www.astesj.com

- Studying data processing distribution between CPU and GPU, and finding the challenges for efficient CPU-GPU cooperation.
- Studying thread synchronization, parallelism control on GPU threads, and finding the challenges for efficient parallelism control.
- Designing algorithms to optimizing data transfer between various memories and memory capacity constraints with efficient memory management.
- Developing parallel combinatorial optimization algorithms/frameworks for the CPU-GPU architecture to efficiently deal with the above mentioned GPU challenges.
- Validating the solution quality and efficiency of the proposed frameworks/algorithms relative to those of the best sequential meta-heuristics with extensive experimental design for graph bisection problem and TSP.

4.1 Difficulties in parallelizing optimization heuristics on GPU

Most of the time the performance of a parallel algorithm may depend on how well the communication structure of the target parallel system is matched with the communication structure of the algorithm. Nowadays one class of parallel processing systems consists of a number of processors, each with its own private memory. In addition, each of these processor memory pairs are connected to a small number of other pairs in a fixed topology. In these systems, if two processor-memory pairs must share data, a message is constructed and sent through the interconnection network. Such a message must be forwarded through one or more intermediate processors in the network. This forwarding introduces delay and hence reduces the amount of speedup achieved.

In science and industry, local search methods are heuristic methods for solving very large optimization problems. Even though these iterative methods can reduce the computational time significantly, the iterative process can still be costly when dealing with very large problem instances. Although local search algorithms can also reduce the computational complexity for the search process, still it is very time consuming for CPU in case of objective function calculations, especially when the search space size is too large. Therefore, instead of traditional CPUs the GPUs can be used to find efficient alternative solutions for calculations.

It is not straightforward to parallelize combinatorial optimization heuristics on GPU. It requires a lot of efforts at both design level and implementation level. We need to achieve few scientific challenges which are mostly related to the hierarchical memory management. The major challenges are: the CPU-GPU data processing with efficient distribution, synchronization of different threads, the data transfer optimization between different memories and their capacity constraints. Such challenges must be considered in redesigning of parallel metaheuristic models on GPU-CPU architectures for solving large optimization problems.

The following major challenges are identified for designing parallel combinatorial optimization algorithms efficiently on CPU-GPU architecture:

1. **CPU-GPU Cooperation:** It is important to optimize the data transfer between GPU and CPU to achieve the best performance. For efficient CPU-GPU cooperation, repartition of task must be defined in metaheuristics.
2. **Parallelism control on GPU threads:** In order to satisfy the memory constraints, it is important to apply the control of threads efficiently, since the order of the threads' execution is unknown for parallel multithreading in GPU computing. Also, it is important to define the mapping efficiently between each of the candidate solutions and a single GPU thread which is designated with a unique thread ID assigned at runtime.
3. **Management of different memories:** The performance optimization of GPU accelerated applications sometimes depend on data access optimization that includes the proper use of different GPU memory spaces. In this regard, it is important to consider the sizes and access latencies of different GPU memories for efficient placement of different optimization structures on different memories.

Below are the contributions of this research that address the challenges mentioned in section 4.1:

4.2 Efficient cooperation between GPU and CPU

An efficient GPU-CPU cooperation requires sharing the work as well as optimizing data transfer between two components.

4.2.1 Task repartition on GPU

The iteration level parallel model focuses on the parallelization of each iteration of metaheuristics. Indeed, the most time consuming task in a metaheuristic is the evaluation of the generated solutions. The concerns for the parallelization is the search techniques/mechanisms which are problem independent operations (For example, the evaluation of successive populations for P-metaheuristics and the generation/evaluation of the neighborhood for S-metaheuristics). As the iteration level model does not change heuristic's behavior, it can be defined as a low level Master Worker model. The following Figure 8 illustrates this Master Worker model. A set of solutions generated by the master at each iteration need to be evaluated. Each worker receives a partition of the solutions set from the master. The solutions are then evaluated by the worker and sent back to the master. In case of S-metaheuristics, the workers can generate the neighbors. Each worker receives the current solution from the master, generates neighbors for evaluation and then return this to master. This model is generic and reusable, since it is problem independent.

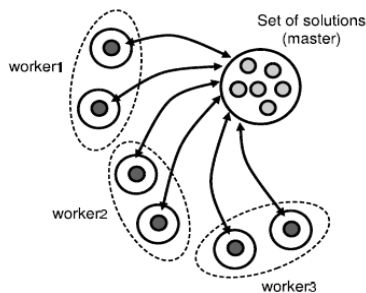


Figure 8: The parallel evaluation of iteration-level model.

As mentioned above, the most time consuming task of metaheuristics is often the evaluation of solution candidates. So, in regards with the iteration level parallel model the evaluation of solution candidates should be performed in parallel. According to the Master Worker model, the solutions can be evaluated in parallel with GPU. We can design the iteration level parallel model based on the data parallel SPMD (single program multiple data) model to achieve this. As showed in Figure 9, the main concept for GPU-CPU task partitioning is that CPU is responsible to host as well as execute the whole sequential part of the handled metaheuristic. On the other hand, the GPU is responsible for the solutions' evaluation at each iteration. The function code in this model called "kernel" to be executed on a number of GPU threads is sent to GPU. The number of threads per block determines the granularity of each partition.

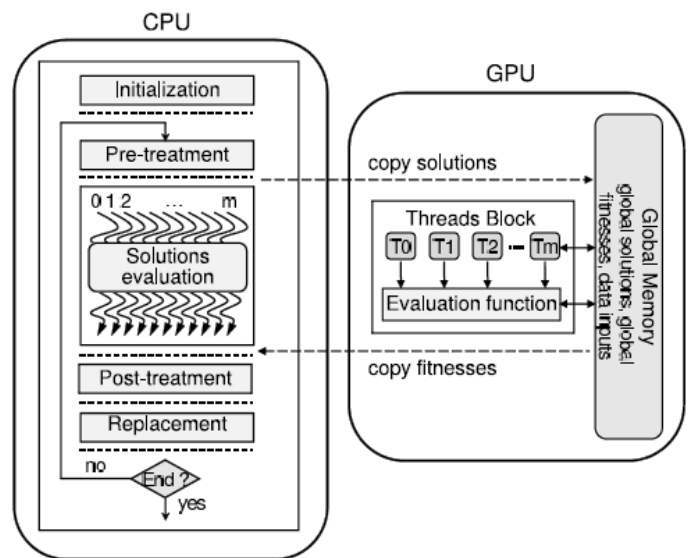


Figure 9: The parallel evaluation of solutions on GPU

4.2.2 Optimization of data transfer

Both GPU and the host computer have their own separate memories and processing elements. So, data transfer between GPU and CPU via PCI bus can be performance bottleneck for GPU applications. A higher volume of data to be copied while repeating the process thousands of times, definitely has a big impact on the execution time. For metaheuristics, the data to be copied are basically the solutions to be evaluated as well as their resulting fitnesses. For most of the P-metaheuristics, the solutions at hand are usually uncorrelated, but for S-metaheuristics each neighboring solution varies slightly compared to the initial candidate solution. So, for parallelization, the data transfers optimization is more prominent in case of S-metaheuristics.

In deterministic S-metaheuristics (such as hill climbing, variable neighborhood search, tabu search), the generation as well as evaluation of the neighborhood can be performed in parallel, which is indeed the most computation intensive. One challenge for data transfer optimization between GPU and CPU is to define where in S-metaheuristics the neighborhood should be generated. Below are two fundamental approaches for this challenge:

- **Neighborhood generation on CPU, but evaluation on GPU:** The neighborhood is generated on the CPU at each iteration of the search process, and the structure associated with this to store the solutions is then copied to GPU. It is pretty straightforward, as a neighbor representation of a thread is associated automatically with it. Usually this is something that can be done to parallelize P-metaheuristics with GPU. So, in this approach the data transfers are basically: 1) Copying neighbor solutions from CPU to GPU 2) Copying fitnesses structures from GPU to CPU.
- **Both neighborhood generation and evaluation on GPU:** The generation of neighborhood happens dynamically on GPU and thereby, there is no need to allocate any explicit structures. A little variation with the candidate solution that generates the neighborhood can be considered as a neighbor. So in this case, only the candidate solution is copied from CPU into GPU. The main advantage is that the data transfers is reduced drastically because only the resulting fitnesses structure need to be copied back from GPU to CPU, but the entire neighborhood does not need to be copied. However, this approach has a problem of determining the mapping between a thread and a neighbor which might be challenging in some cases.

Although the first approach is straightforward, implementing this in S-metaheuristics with GPU will require a large volume of data transfers in case of large neighborhood. This approach can be implemented in P-metaheuristics because the whole population is generally copied from the CPU to the GPU. The first approach might affect the performance because of the external bandwidth limitation. That is why, we consider the second approach i.e. both generation of neighborhood and evaluation on GPU.

The Proposed GPU accelerated Algorithm:

It is not a simple task to adapt traditional S-metaheuristics to GPU. We propose an algorithm 4.2.2 in a generic way to rethink S-metaheuristics on GPU. Memory allocations are made on GPU at the initial stage and also, data inputs as well as candidate solution are allocated initially.

Algorithm 4.2.2: GPU accelerated S-metaheuristic Template

- 1: Select an initial solution and also, evaluate this solution
- 2: Initialize specific variables if needed
- 3: Allocate problem data inputs, a solution, a neighborhood fitness's structure and additional structures of solution on GPU device memory,
- 4: Copy data inputs of the problem, a solution, and additional structures of solution on GPU device memory
- 5: **repeat**
- 6: **for each** neighbor on GPU in parallel
- 7: Evaluate the candidate solution
- 8: Add the resulting fitness into the neighborhood fitness's structure
- 9: **end**
- 10: Copy neighborhood fitness's structure on CPU host memory
- 11: Specific strategy for solution selection on the neighborhood fitness's structure
- 12: Specific post-treatment
- 13: Copy the chosen solution and additional structures of solution on GPU device memory
- 14: **until** the stop criterion is satisfied

As previously said, heavy computations are required by GPUs with predictable accesses of memory. Hence, we need to allocate a structure to store the results for evaluating each neighborhood fitness's structure at different addresses. To facilitate the computation of neighbor evaluation, we can also allocate additional solution structures that are problem dependent. The data inputs of the problem, initial candidate solution and additional solution structures need to be copied onto GPU (line#4). The data inputs of the problem are read only structure which doesn't change at the time of all executions of the S-metaheuristic. So, during all the execution their associated memories are copied for only once. In the parallel iteration level, the neighboring solutions are evaluated and resulting fitnesses are then copied to the neighborhood fitnesses structure (line #6 to #9). The neighborhood fitnesses structure need to be copied into CPU host memory, since it is not defined in which order the candidate neighboring solutions are evaluated (line #10). Then, a particular strategy for the selection of the solution is implemented on the neighborhood fitness's structure (line #11): CPU explores the neighborhood fitnesses structure in a sequential way. Finally, the chosen solution and additional structures of solution are copied into GPU device memory (#13). This process repeats until some stop criteria is met.

4.2.3 Additional optimization of data transfer

In some S-metaheuristics, the selection criteria to find the best solution are based on maximal or minimal fitness. So, only one value (maximal fitness or minimal fitness) can merely be copied from GPU to CPU. However, it is not straightforward to find the appropriate maximal/minimal fitness's, as the read/write memory operations are performed asynchronously. The traditional parallel techniques that strongly suggests the global synchronization of hundreds of threads can decrease the performance drastically. Therefore, the techniques for the parallel reduction of each thread block should be adapted to address this issue. The following algorithm describes the techniques for the parallel reduction of each thread block.

Algorithm: parallel reduction techniques

Input Parameter: InputFitnesses on Global memory;

- 1: SharedMem[ThreadId] := InputFitnesses[id]
- 2: Synchronize locally
- 3: for n := NumOfThreadsPerBlock/2 ; n > 0; n := n / 2
- 4: if ThreadId < n
- 5: SharedMem[ThreadId] :=
- 6: Compare(SharedMem[ThreadId], SharedMem[ThreadId + n])
- 7: Synchronize locally
- 8: end if
- 9: end for
- 10: if ThreadId = 0
- 11: OutputFitnesses[blockId]:= SharedMem[0]
- 12: end if

Output: OutputFitnesses

One element of input fitnesses from global memory is basically loaded into shared memory by each thread (line #1 and line #2). The array elements are compared by pairs at each iteration of the loop (line #3 to #7). As threads operate on different

memory addresses, the maximum/minimum of a given array can be found via the shared memory by applying local threads synchronizations in a given block. We can find the maximum or minimum fitness for all neighbors after a number of iterations are operated on GPU reduction kernel.

In some S-metaheuristics (such as simulated annealing), indeed the best neighbor is selected by the selection of maximal fitness or minimal fitness at each iteration. Therefore, the entire fitness's structure doesn't need to be transferred for these algorithms, and also, further optimizations might be possible.

4.3 Efficient control of parallelism

The efficient parallelism control on GPU for the iteration level is mainly focused here. For parallel multithreading in GPU computing, the order for the execution of threads is unknown, since it is indeed hyper threading based. First, it is important to apply the control of threads efficiently in order to satisfy the memory constraints. This allows to improve the overall performance by adding some robustness in the developed metaheuristics on GPU. Second, it is important for S-metaheuristics to define the mapping efficiently between GPU thread and each neighboring candidate solution. Therefore, at runtime each GPU thread is assigned with a unique thread ID for this purpose.

The key components for the parallelism control are the heuristic for controlling threads and the efficient mappings of the neighborhood structures. New S metaheuristics can be designed on GPU by considering these key components. The difficulty arises from the sequential characteristics of the metaheuristics that are first improvement based. In case of traditional parallel architectures, the neighborhood is generally divided into separate partitions with equal size. Then the generated partitions are evaluated and when an improved neighbor is found, the exploration stops. The whole neighborhood doesn't need to be explored, as the parallel model is asynchronous. When the computations become asynchronous, GPU computing is not efficient to execute such algorithms because GPUs' execution model is basically SIMD. Moreover, as the execution order of GPU threads is not defined, no such inherent mechanism exists for stopping the kernel in its execution.

We can deal with this type of asynchronous parallelization by transforming these algorithms into a data parallel regular application. So, we can consider the previous iteration-level parallelization scheme on GPU, which means that instead of applying to the entire neighborhood, we can apply to a sub set of solutions which need to be generated and evaluated on GPU. A specific post treatment on this partial set of solutions is performed on CPU after this parallel evaluation. This approach is considered as a parallel technic for simulating the first improvement based S-metaheuristic. Regarding implementation, this is similar to the Algorithm 4.2.2 that is proposed in the previous section. The sub set that has to be handled is the only difference concerns, in which the neighbors are randomly selected. The heuristic of thread

control can adjust the remaining parameters automatically once the number of neighbors are set.

Although it may be normal to deal with such an asynchronous algorithm, but compared to an S-metaheuristic this approach may not be efficient, because a full neighborhood exploration is performed on GPU in case of S-metaheuristic. Indeed, memory accesses need to constitute an adjacent range of addresses to get coalesced in order to get a better global memory performance. But this cannot be achieved for exploring a partial neighborhood, because neighbors are chosen randomly.

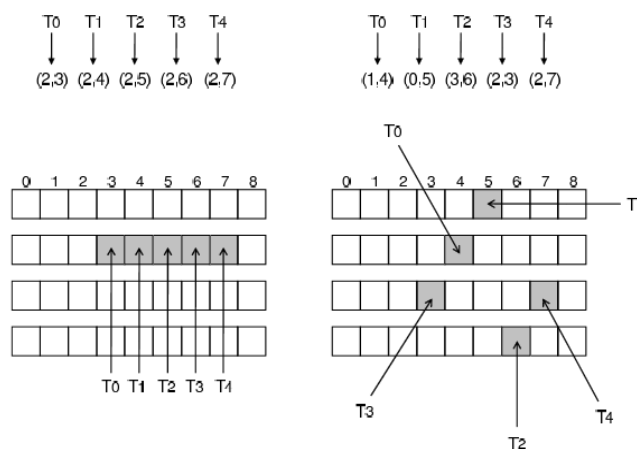


Figure 10: Illustration of a memory access pattern for both full exploration and partial exploration.

Figure 10 shows a memory access pattern for both full exploration and partial exploration of the neighborhood. In case of full exploration (left side of the Figure), all the neighbors are generated and many thread accesses get coalesced. In case of partial exploration of the neighborhood it does not happen (right side of the Figure), because no connection is available between the elements to get accessed.

4.4 Efficient memory management

For efficient implementation of parallel metaheuristics on GPU, it's important to understand the hierarchical organization of different memories. However, global memory coalescing can be done for some of the optimization structures which are specific to given GPU thread. This is usually the case for the large local structures that are used in P-metaheuristics for the evaluation function, or organization of data for a population.

Data accesses optimization that includes proper utilization of different memory spaces in GPU is important for optimizing the performance of GPU accelerated applications. The texture memory can certainly provide an amazing aggregation capabilities such as caching global memory. Each unit in texture memory gets some internal memory which buffers data from global memory. We can consider texture memory like a relaxed technique/mechanism of global memory access for the GPU threads, since coalescing is not required to accesses to this memory. For metaheuristics, the utilization of texture memory can be well adapted due to the following reasons:

- As no write operations are possible to perform on Texture memory, it is considered as a read only memory. This memory can be adapted in metaheuristics, because the inputs of the problem are read only values as well.
- In computation of evaluation methods, data accesses are frequent. The texture memory can make some high performance improvement by reducing the number of memory transactions
- In order to provide the best performance for 2D/1D access patterns, cached texture data is laid out. From a spatial locality perspective when the threads of a warp read locations are close together, then the best performance can be achieved. As the inputs of the optimization problems are generally 1D solution vectors or 2D matrices, we can bind the optimization structures to texture memory. Using of texture memories in place of global memory accesses is totally a mechanical transformation.

For parallelizing a metaheuristic, reducing the search time is one of the main goals and also, a fundamental aspect when there are some hard requirements on search time in some types of problems. In this regard, the parallel evaluation of solutions can be a concern for the iteration-level model. It can be considered as an acceleration model for the evaluation of independent as well as parallel computations. This is usually the case of S-metaheuristics that improve a single solution iteratively. There is no direct interaction between different neighborhood moves in these algorithms.

In case of P-metaheuristics, things are little different. During the search process, the solutions that represents a population can cooperate. For example, the solutions that compose the population are selected/reproduced by using variation operators in evolution based P-metaheuristics. A new solution can be constructed with different attributions of solutions which belong to the current population. Participating in constructing a common or shared structure (for example, ant colonies) is another example that concerns P-metaheuristics. The main input for generating the new population of solutions will be this shared structure, and the solutions that are generated previously participate in updating this type of common structure. Unlike S-metaheuristics, P-metaheuristics can provide additional cooperative aspects which is much more important while running multiple metaheuristics in parallel. The challenging issue here is the exploitation of these cooperative properties on GPU architectures.

To the best of our knowledge, these cooperative algorithms are never investigated much for CPU-GPU architecture. For P-metaheuristic, it is indeed the costliest operation to evaluate the fitness for each solution. Therefore, it is important to clearly define the task distribution in this scheme: for each cooperative algorithm the CPU is responsible for managing the whole sequential search process, whereas the GPU is responsible for evaluating the populations in parallel. The CPU sends a set of solutions through the global memory to be evaluated by GPU, and then, these solutions are processed on GPU. The same evaluation function kernel is executed on each GPU thread associated with

one solution. Finally, the results of the evaluation function are sent to CPU through global memory.

Algorithm: GPU accelerated Cooperative algorithm for the parallel evaluation of populations	
1:	Select initial populations
2:	Initialize specific variables if needed
3:	Allocate problem data inputs, the different populations, fitness's structures, additional structures of solution on GPU
4:	Copy the problem data inputs to GPU
5:	repeat
6:	for each P-metaheuristic
7:	particular pre-treatment
8:	Copy different populations as well as additional structures of solution on GPU device memory
9:	for each solution on GPU in parallel
10:	Evaluating Solution
11:	Adding resulting fitnesses to corresponding fitness's structure
12:	end for
13:	Copy fitness's structures on CPU (hosts memory)
14:	particular post-treatment
15:	Population replacement
16:	end for
17:	Possible transfers between different P-metaheuristics
18:	until some stop criteria is met

According to the above algorithm, memory allocations on GPU are made first i.e. problem data inputs, different populations and corresponding fitness's structures are allocated first (line #3). Additional structures of solution that are problem dependent can be allocated as well in order to make the computation of solution evaluation easier (line #3). Secondly, the data inputs of the problem need to be copied onto GPU device memory (line #4). The structure of these problem data inputs are read only, and also, for all the execution their associated memory need to be copied for one time only. Thirdly, the algorithm mainly describes that at each iteration different populations as well as the associated/ additional structures need to be copied (line #8). Then, the solutions are evaluated on GPU in parallel (lines #9 to #12). Fourthly, the structures of the fitnesses need to be copied into CPU (#13) and then a particular post treatment as well as population replacement are performed (line #14 and #15). Lastly, a possible migration can be performed on CPU at the end of each generation for information exchange between different P-metaheuristics (line #17). This process repeats until some stop criteria is met.

In this algorithm, GPU is utilized synchronously as a device coprocessor. However, as previously mentioned, copying operations (like population and fitnesses structures) from CPU to GPU can be a serious performance bottleneck. Accelerating the search process is the main goal of this scheme which does not alter the meaning of the algorithm. Hence, compared to the classic design on CPU, the policy of migration between the P-metaheuristics remains unchanged. This scheme is essentially

devoted to cooperative algorithms (synchronous), as the GPU is utilized as a device coprocessor for parallel evaluation of all individuals.

5. Experimental Validation

For our experimental validation, we have considered the following optimization problems to parallelize some heuristic methods with our proposed GPU framework in order to find higher quality solutions.

- Graph Bisection Problem
- Travelling Salesman Problem

5.1 Graph Bisection Problem

For Graph Bisection Problem, we have made experiments to parallelize hill climbing algorithm as well as simulated annealing algorithm with our GPU framework as follows:

5.1.1 Experimental Environment

OpenCL programming environment was setup on a NVIDIA CUDA GPU using C++ as follows:

- NVIDIA CUDA GPU (GeForce GTX 1050 Ti)
- OpenCL Driver for NVIDIA CUDA GPU
- VISUAL STUDIO 2017
- Windows 10 (64-bit Ultimate edition)

Created Visual Studio OpenCL projects (for both Hill Climbing and Simulated Annealing) using Visual C++.

5.1.2 Experimental Data

We've considered the following problem for the experiments:

Problem:

Prepare 200 GPU threads and run the heuristic method (Hill Climbing/Simulated Annealing) of graph bisection for the same problem instance to find the best solution/cost by considering the following factors:

- Each thread generates its initial random solution.
- Repeat 50 times to 100 times for each of the 200 threads to run heuristic method (Hill Climbing/Simulated Annealing) of graph bisection for the same problem instance.
- Keeps the running minimal cost and solution. They don't communicate.
- All the 200 threads pause. Find the smallest solution and its cost. Broadcast them to all the 200 threads as initial solution. Report the best solution and its cost.
- Many threads run in parallel. When the execution comes to a barrier method call, all threads suspend itself until all the threads have reached the barrier. Then they all do one operation (find minimum solution and broadcast to all threads), then resume their execution.
- Here, numbers 200, 50, and 100 are just random example numbers and we can change them based on the number of GPU threads that we can use.

Also for experimental validations, various problem instances (such as 20 vertices, 50 vertices, 100 vertices etc.) for graph

bisections were used. An example of graph bisections problem for 20 vertices is given below:

```

0
1 1
0 0 1
0 1 0 0
0 0 0 1 0
0 1 0 1 0 0
1 0 0 0 0 1 0
0 1 0 1 0 0 0 1
1 0 0 0 0 0 1 1 0
0 1 0 0 0 0 1 0 1 0
0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 1 1 1 0 0 0 1
0 0 0 0 1 0 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 1 0
1 1 0 0 1 1 0 0 0 1 0 1 1 1 0
1 0 1 0 0 1 0 1 0 1 0 0 0 1 0 1
0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 1
0 1 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0
0 0 0 0 1 1 1 0 0 0 0 1 0 0 0 1 0 0 0
    
```

5.1.3 Experimental Data Presentation and Analysis

Below are the experimental presentation and analysis for both parallel Hill Climbing and simulated annealing algorithm with our GPU framework for Graph Bisection Problem. It is noted that we've used the following formula for calculating cut size as explained in section 2 (background).

$$cutSize(L, R) = \sum_{(x,y) \in L \times R} adj(x, y)$$

under the constraint that $|L| = |R|$.

Here, $adj()$ is the adjacency matrix and $cutSize$ represents the cost for the bisection of a given graph $G = (V, E)$; where $|V|$ is an even integer and we find the partition of V into subsets L and R that minimizes the objective function.

Data Presentation and Analysis for parallel Hill Climbing

As we ran both CPU based sequential hill climbing and our GPU based parallel hill climbing solution for multiple problem instances (For example: 20 vertices, 50 vertices, 100 vertices etc.) with the above mentioned scenarios and data, we've found the following results:

Table 1: Experiment Results for hill climbing algorithm on GPUs

Problem Instance	Best Cost for Sequential Solution	Best Cost for Parallel Solution	Improvement
Graph10.txt	4	3	25%
Graph15.txt	18	13	27.77%
Graph20.txt	25	18	28.00%
Graph30.txt	21	17	23.52%
Graph50.txt	18	15	16.66%
Graph100.txt	30	20	33.33%

We can see from the above results in Table 1 that for each problem instance the GPU based parallel solution got some good

improvement on cut size (cost) compared to CPU based sequential solution and thus, the average improvement on cut size (cost) for multiple problem instances is 25.71%. Therefore, much better optimized solution is found for Graph Bisection problem by parallelizing hill climbing algorithm on GPUs.

Data Presentation and Analysis for parallel Simulated Annealing

As we ran both CPU based sequential Simulated Annealing and GPU accelerated parallel Simulated Annealing solution for multiple problem instances ((For example: 30 vertices, 100 vertices etc.)) with the above mentioned scenarios and data, we’ve found the following results:

Table 2: Experiment Results for simulated annealing algorithm on GPUs.

Problem Instance	Best Cost for Sequential Solution	Best Cost for Parallel Solution	Improvement
Graph10.txt	5	5	0%
Graph15.txt	14	13	7.14%
Graph20.txt	19	15	21.05%
Graph30.txt	44	32	27.27%
Graph50.txt	133	60	54.88%
Graph100.txt	663	140	78.88%

We can see from the above results in Table 2 that for each problem instance the GPU based parallel solution got some good improvement on cut size (cost) compared to CPU based sequential solution. Thus, the average improvement for multiple problem instances on cut size (cost) is 31.53%. This improvement looks better when comparatively large problem instances are considered (such as: for 100X100 adjacent matrix, the improvement is 78.88%). Therefore, we can say that we’ve found a better optimal solution as we parallelize Simulated Annealing algorithm with GPUs.

5.2 Travelling Salesman Problem

5.2.1 Experiment Design

We built an experiment environment with the followings:

- CUDA programming model
- C++
- Visual Studio 2017
- NVIDIA GPU (GeForce GTX 1050 Ti)
- CPU (Core i7 9300 quad-core processor)
- Windows 10 (64-bit Ultimate edition)

For experimental validation of Travelling Salesman Problem (TSP), we’ve considered a CPU based sequential solution [29] that we previously proposed at an IEEE conference in 2017. For parallelization with our GPU framework, we’ve considered the following steps:

- Step-1: Allocate 200 GPU threads.
- Step-2: For each GPU threads.
 - a) Run our Main TSP Greedy-Genetic Algorithm in parallel to find the Path Length.
 - b) Send this thread obtained path length to CPU.

Step-3: CPU compares all the path lengths sent by all 200 GPU threads and determine the best path length.

It is noted that our Main TSP Greedy-Genetic algorithm that we previously implemented for a sequential solution [29] is illustrated in background section. We’ve parallelized the same proposed heuristics with our GPU framework, used the same data and compared the two results to confirm that we find the better optimized solution with GPU.

We developed our simulator by producing the inputs for Euclidean TSP and simplified the simulator with the following assumptions:

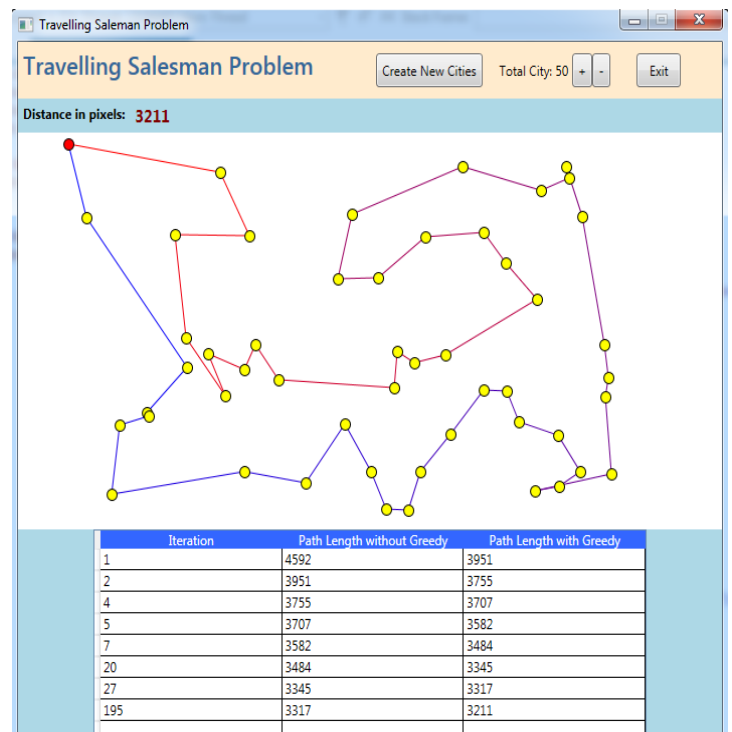
- The cities are located on the plane
- The distance between the cities is the Euclidean distance
- Each city is able to reach all other cities

We generated the inputs in such a way that the cities were uniformly placed on a grid at random with 600 columns and 350 rows. Then, by using the columns and rows as unit of Euclidean distance, the path length was obtained after so many numbers of iterations.

5.2.2 Experimental Data Tabulation and Visualization

We made 10+ repeated runs with the same instance in order to observe the behavior of our GPU accelerated parallel TSP solution. Table 3 shows the path lengths for n = 50 cities with 200 iterations. The input for each of this execution was generated randomly as explained above.

The following visual simulator developed for our previously proposed CPU based solution [29] shows the path length in different iteration levels which finally provide us with one minimum path length after completing all iterations.



We also randomly captured the following GPU thread results from our GPU based parallel solution for one single execution.

Table 3: Path lengths on different GPU threads for TSP

GPU Threads	Path length	Best Path length	Improvement
1	2086		
2	2019		
10	2162		
25	1895		
40	2362		
42	2108		
45	1965		
47	1888		
50	1919		
56	2169		
60	2009	1776	12.68%
65	2195		
68	2023		
75	2380		
85	1929		
88	1977		
100	1809		
115	1990		
160	1776		
200	2019		

After 10 times repeated run of our TSP heuristics on both GPU and CPU for the same number of cities (n=50), we obtained the following results:

Table 4: Comparison of path lengths between CPU and GPU for TSP

Run#	Path Length (CPU)	Path Length (GPU)	Improvement
1	2034	1776	12.68%
2	1965	1702	13.38%
3	2370	1904	19.66%
4	1881	1607	14.56%
5	2018	1780	11.79%
6	2243	1945	13.28%
7	1980	1756	11.31%
8	1777	1578	11.19%
9	2400	2019	15.87%
10	2018	1745	13.52%

5.2.3 Experiment Result Explanation

In our experimental validation, we involved 200 GPU threads to run the same TSP metaheuristics simultaneously by using our GPU framework. We captured different thread results which is illustrated in Table 4. We can see that the best path length calculated by thread 1 is 2086, thread 2 is 2019,, thread 100 is 1809....., thread 200 is 2019 and so on. As all the thread results are sent to CPU for comparison, CPU finds the best path length (shortest) for TSP is 1776. The average path length obtained from different threads is 2034 and so, the improvement is 12.68% (approximately) because of the parallelization.

As we made 10 times repeated run of our TSP heuristics on both GPU and CPU for the same number of cities (n=50), both results are illustrated in Table 4. We can see that for run#3, CPU based best path length is 2370, whereas GPU best path length is 1904 which is much shorter than CPU based path length and thus the improvement on run#3 is 19.66%. Similarly, if we also observe the results for other repeated runs (#1, #2.....#10), we

can easily notice that GPU based path length is definitely shorter than CPU based path length and thus there are some improvements in the solution quality for each execution on GPU.

The presented results show that the path lengths are shorter up to 19.66%, with an average of 13.72%. This improvement for finding the shortest path length in TSP is due to GPU parallelization with our framework.

6. Conclusion

In combinatorial optimization, parallel metaheuristic methods can be helpful to improve the effectiveness and robustness of a solution. But, their exploitation might make it possible to solve real world problems by only using important computational power. GPUs are based on high performance computing and it has been revealed that GPUs can provide such computational power. However, we have to consider that GPUs can have many issues related to memory hierarchical management, since parallel models' exploitation is not trivial. In this paper, a new guideline is established to design parallel meta heuristics and efficiently implement on GPU.

An efficient mapping of the GPU based iteration level parallel model is proposed. In the iteration level, CPU is used to manage the entire search process, whereas GPU is dedicated to work as a device coprocessor for intensive calculations. In our contributions, to achieve the best performance an efficient cooperation between CPU and GPU is very important because it minimizes the data transfer. Also, the goal for the parallelism control is, controlling the neighborhood generation to meet the memory constraints and also, finding the efficient mapping between the GPU threads and neighborhood solutions.

The redesigning of GPU based iteration level parallel model is suitable for most of the deterministic metaheuristics like Tabu search, Hill climbing, Simulated Annealing, or iterative local search. Moreover, we applied an efficient thread control to prove the robustness of our approach. This allows GPU accelerated metaheuristics preventing from crash when a large number of solutions are considered for evaluation. Also, this kind of thread control can provide some improvements with additional acceleration.

Redesigning of the algorithm for an efficient management of the memory on GPU is another contribution. Our contribution is basically the redesigning of GPU accelerated parallel metaheuristics. More specifically, we proposed multiple different general schemes to build efficient GPU based parallel metaheuristics as well as cooperative metaheuristics on GPU. In one scheme, the parallel evaluation of the population is combined with cooperative algorithms on GPU (iteration level). In regards to implementation, this approach is a very generic approach because we only considered the evaluation kernel. However, the performance is little limited in this approach because of data transfer between GPU and CPU. To address this issue, GPU based two other approaches operate on the complete distribution of search process, involving the appropriate use of local memories.

Applying such a strategy allows to extremely improve the performance. This approaches might experience some limitations because of the memory limitations with some of the problems which can be possibly more demanding with respect to resources. We have proved effectiveness of the proposed methods with a set of experiments in a general manner.

Furthermore, our experiments show that not only GPU computing exploits the parallelism to improve the solution quality, but also it can speed up the search process. In the future, we'll try to extend the framework with further features to be validated on a wider range of NP-hard problems in various fields like deep neural network, data science, artificial intelligence, computer vision, machine learning etc. including current industry challenges.

Conflict of Interest

We have no conflicts of interest to disclose.

References

- [1] Dr. Lixin Tao, "Research Incubator: Combinatorial Optimization" Pace University, NY, February 2004
- [2] Man Leung Wong, Tien-Tsin Wong, Ka-Ling Fok, "Parallel evolutionary algorithms on graphics processing unit" IEEE Congress on Evolutionary Computation, 2005. <https://doi.org/10.1109/CEC.2005.1554979>
- [3] Man-Leung Wong, Tien-Tsin Wong, "Parallel hybrid genetic algorithms on consumer-level graphics hardware" IEEE International Conference on Evolutionary Computation, 2006. <https://doi.org/10.1109/CEC.2006.1688683>
- [4] Ka-Ling Fok, Tien-Tsin Wong, Man-Leung Wong, "Evolutionary computing on consumer graphics hardware" IEEE Intelligent Systems, 22(2), 69–78, 2007. <https://doi.org/10.1109/MIS.2007.28>
- [5] Weihang Zhu, "A study of parallel evolution strategy: pattern search on a gpu computing platform" GEC '09 Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation, 765–772, 2009. <https://doi.org/10.1145/1543834.1543939>
- [6] Ramnik Arora, Rupesh Tulshyan, Kalyanmoy Deb, "Parallelization of binary and real-coded genetic algorithms on gpu using cuda" IEEE Congress on Evolutionary Computation, 2010. <https://doi.org/10.1109/CEC.2010.5586260>
- [7] Shigeyoshi Tsutsui, Noriyuki Fujimoto, "Solving quadratic assignment problems by genetic algorithms with gpu computation: a case study" GECCO '09 Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, 2523–2530, 2009. <https://doi.org/10.1145/1570256.1570355>
- [8] Ogier Maitre, Laurent A. Baumes, Nicolas Lachiche, Avelino Corma, Pierre Collet, "Coarse grain parallelization of evolutionary algorithms on gpgpu cards with easea" GECCO '09 Proceedings of the 11th Annual conference on Genetic and evolutionary computation, 1403–1410, 2009. <https://doi.org/10.1145/1569901.1570089>
- [9] Pablo Vidal, Enrique Alba, "Cellular genetic algorithm on graphic processing units" Springer Juan Gonz'alez, David Pelta, Carlos Cruz, Germ'an Terrazas, and Natalio Krasnogor, editors, Nature Inspired Cooperative Strategies for Optimization (NICSO 2010), 2010. https://doi.org/10.1007/978-3-642-12538-6_19
- [10] Pablo Vidal, Enrique Alba, "A multi-gpu implementation of a cellular genetic algorithm" IEEE Congress on Evolutionary Computation, 2010. <https://doi.org/10.1109/CEC.2010.5586530>
- [11] Petr Pospichal, Jir'ı Jaros, Josef Schwarz. "Parallel genetic algorithm on the cuda architecture", In Cecilia Di Chio, Stefano Cagnoni, Carlos Cotta, Marc Ebner, Anik'o Ek'art, Anna Esparcia-Alc'azar, Chi Keong Goh, Juan J. Merelo Guerv'os, Ferrante Neri, Mike Preuss, Julian Togelius, and Georgios N. Yannakakis, editors, EvoApplications, Springer, 2010. https://doi.org/10.1007/978-3-642-12239-2_46
- [12] Th'e Van Luong, Nouredine Melab, El-Ghazali Talbi. "GPU-based Island Model for Evolutionary Algorithms. Genetic and Evolutionary Computation Conference" GECCO '10 Proceedings of the 12th annual conference on Genetic and evolutionary computation, 1089–1096, ACM, 2010. <https://doi.org/10.1145/1830483.1830685>
- [13] Adam Janiak, Wladyslaw A. Janiak, Maciej Lichtenstein. "Tabu search on gpu", J. UCS, 14(14):2416–2426, 2008.
- [14] WZhu, J Curry, A Marquez. Simd, "tabu search with graphics hardware acceleration on the quadratic assignment problem" International Journal of Production Research, 2008.
- [15] Th'e Van Luong, Nouredine Melab, El-Ghazali Talbi. "GPU-based Multi-start Local Search Algorithms" Coello C.A.C. (eds) Learning and Intelligent Optimization, Springer, 2011. https://doi.org/10.1007/978-3-642-25566-3_24
- [16] Asim Munawar, Mohamed Wahib, Masaharu Munetomo, Kiyoshi Akama, "Hybrid of genetic algorithm and local search to solve maxsat problem using nvidia cuda framework." Genetic Programming and Evolvable Machines, 10(4), 391–415, 2009. https://doi.org/10.1007/978-3-642-25566-3_24
- [17] Th'e Van Luong, Nouredine Melab, El-Ghazali Talbi, "Parallel Hybrid Evolutionary Algorithms on GPU" IEEE Congress on Evolutionary Computation, 2010. <https://doi.org/10.1109/CEC.2010.5586403>
- [18] Man Leung Wong, "Parallel multi-objective evolutionary algorithms on graphics processing units" GECCO '09 Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, 2515–2522, ACM, 2009. <https://doi.org/10.1145/1570256.1570354>
- [19] Luca Mussi, Stefano Cagnoni, Fabio Daolio. "Gpu-based road sign detection using particle swarm optimization" IEEE Ninth International Conference on Intelligent Systems Design and Applications, 2009. <https://doi.org/10.1109/ISDA.2009.88>
- [20] You Zhou, Ying Tan, "Gpu-based parallel particle swarm optimization" IEEE Congress on Evolutionary Computation, 2009. <https://doi.org/10.1109/CEC.2009.4983119>
- [21] Boguslaw Rymut, Bogdan Kwolek, "Gpu-supported object tracking using adaptive appearance models and particle swarm optimization" In Leonard Bolc, Ryszard Tadeusiewicz, Leszek J. Chmielewski, and Konrad W. Wojciechowski, editors, ICCVG, Springer, 2010. https://doi.org/10.1007/978-3-642-15907-7_28
- [22] Simon Harding, Wolfgang Banzhaf, "Fast genetic programming on GPUs" In Proceedings of the 10th European Conference on Genetic Programming, 90–101. Springer, 2007.
- [23] Darren M. Chitty, "A data parallel approach to genetic programming using programmable graphics hardware" GECCO '07 Proceedings of the 9th annual conference on Genetic and evolutionary computation, 1566–1573, ACM, 2007. <https://doi.org/10.1145/1276958.1277274>
- [24] William B. Langdon, Wolfgang Banzhaf, "A SIMD interpreter for genetic programming on GPU graphics cards" EuroGP Proceedings of the 11th European Conference on Genetic Programming, 73–85, Springer, 2008. https://doi.org/10.1007/978-3-540-78671-9_7
- [25] William B. Langdon, "Graphics processing units and genetic programming: an overview. Soft Comput" Springer, 2011. <https://doi.org/10.1007/s00500-011-0695-2>
- [26] Asim Munawar, Mohamed Wahib, Masaharu Munetomo, Kiyoshi Akama, "Theoretical and empirical analysis of a gpu based parallel bayesian optimization algorithm" IEEE International Conference on Parallel and Distributed Computing, Applications and Technologies, 2009. <https://doi.org/10.1109/PDCAT.2009.32>
- [27] Lucas de P. Veronese, Renato "A. Krohling. Differential evolution algorithm on the gpu with c-cuda" IEEE Congress on Evolutionary Computation, 2010. <https://doi.org/10.1109/CEC.2010.5586219>
- [28] Mar'ia A. Franco, Natalio Krasnogor, Jaume Bacardit, "Speeding up the evaluation of evolutionary learning systems using gpgpus" GECCO '10 Proceedings of the 12th annual conference on Genetic and evolutionary computation, 1039–1046, ACM, 2010. <https://doi.org/10.1145/1830483.1830672>
- [29] Mohammad Rashid, Dr. Miguel A. Mosteiro, "A Greedy-Genetic Local-Search Heuristic for the Traveling Salesman Problem" IEEE ISPA/IUCC, 2017. <https://doi.org/10.1109/ISPA/IUCC.2017.00132>

Appendix:

Sample code for Kernel execution:

```
const char *kernelBytes = "\n" \
    "__kernel void SimulatedAnnealing(
    " __global int* inBestPartition,
    " __global int* inBestCost,
    " __global int* outBestPartition,
    " __global int* outBestCost,
    " const int NumOfThreads)
    {
    " int gid = get_global_id(0);
    " int* FinalbestPartition;
    " int FinalBestCost;
    " barrier(CLK_GLOBAL_MEM_FENCE);
    " FinalbestPartition = inBestPartition[gid];
    " int newCost = inBestCost[gid];
    " if(newCost < FinalBestCost)
    " {
    "     FinalBestCost = newCost;
    "     FinalbestPartition = inBestPartition[gid];
    " }
    " if(gid < NumOfThreads)
    " {
    "     outBestPartition[gid] = FinalbestPartition;
    "     outBestCost[gid] = FinalBestCost;
    " }
    " }
```

Sample code for parallel simulated annealing:

```
/******
Initialization of input data
*****
const int vertexNumber = 100;
size_t sizeOfBuffers = vertexNumber*sizeof(int);
int* inputBestPartition = (int*)malloc(sizeOfBuffers);
int* inputBestCost = (int*)malloc(sizeOfBuffers);

//=====SIMULATED ANNEALING=====
int bestPartition[vertexNumber];
randomPartition(bestPartition, vertexNumber);
int currentCost = cutSize(bestPartition, vertexNumber);
int bestCost = currentCost; // p[] is the

int neighbor[vertexNumber]; // Allocate space for a n
double t = 10.0; // Initial temperature; parame
```

```
while (t > 0.01) { // While not frozen; parame
    copyArray(bestPartition, neighbor);
    randomSwap(neighbor, vertexNumber); //
    int newCost = cutSize(neighbor, vertexNumber);
    int delta = newCost - currentCost;

    // Probability to accept a worsening neighbor
    double acceptProbability = exp(-delta / t); //
    // Otherwise take it with probability acceptPr

    if (delta <= 0) {
        // Accept the neighbor
        copyArray(neighbor, bestPartition);
        currentCost = newCost;

        // If the new solution is the best seen so
        if (currentCost < bestCost) {
            bestCost = currentCost;
        }
    }

    t = 0.95*t; // Reduce temperature
}
```

Sample code for parallel hill climbing:

```
/******
Initialization of input data
*****
const int vertexNumber = 20;
size_t sizeOfBuffers = vertexNumber*sizeof(int);
int* inputBestPartition = (int*)malloc(sizeOfBuffers);
int* inputBestCost = (int*)malloc(sizeOfBuffers);

//Initial hill climbing solution
int bestPartition[vertexNumber];
randomPartition(bestPartition, vertexNumber);
int bestCost = cutSize(bestPartition, vertexNumber);
//cout << "\nInitial Best Cost :" << bestCost << "\n" <<

// Loop terminates if we see 100 successive non-improving
int neighbor[vertexNumber]; // Allocate space for a ne:
for (int i = 0; i < 100; i++) {
    copyArray(bestPartition, neighbor);
    randomSwap(neighbor, vertexNumber); // neighbor[] :

    int newCost = cutSize(neighbor, vertexNumber);

    if (newCost < bestCost) { // If new cost
        copyArray(neighbor, bestPartition);
        bestCost = newCost;

        i = 0; // retry 100 ti:
    }
}

for(int i=0; i<vertexNumber; i++) //We put some r
{
    inputBestPartition[i] = bestPartition[i];
}

inputBestCost[0]=bestCost;
```


Probabilistic Method for Anomalies Detection Based on the Analysis of Cyber Parameters in a Group of Mobile Robots

Elena Basan*, Alexander Basan, Oleg Makarevich

Department of Information Security Southern Federal University, Taganrog, 347922, Russia

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 26 September, 2018

Online: 18 November, 2018

Keywords:

Mobile robots

Abnormal behavior

Probabilistic methods

Attack

Detection

ABSTRACT

This article is devoted to the issues of ensuring the security of a group of mobile robots in the implementation of attacks aimed at the property of accessibility of information and the availability of network nodes. The article presents a method for detecting an abnormal behavior of a network node based on the analysis by the group members of the parameters: residual energy and network load. Analysis of the behavior of individual robots relative to general behavior using probabilistic methods avoids the problem of creating a reference distribution for describing the behavior of a node, as well as creating a signature database for detecting anomalies. The developed method demonstrates high detection rate of denial of service attack and distributed denial of service attack with the number of malicious nodes not exceeding or slightly exceeding the number of trusted nodes. It also provides detection of the Sybil attack.

1. Introduction

This paper is an extension of work originally presented in CyberC 2017, "9th International Conference on Cyber-enabled distributed computing and knowledge discovery" titled 'A Trust Evaluation Method for Active Attack Counteraction in Wireless Sensor Networks' [1]. Mobile robot networks are quite vulnerable to attacks both over the network and physical properties of nodes. The article [2] presented a threats model for the network of mobile robots. The authors also analyzed the attacks for a group of mobile robots. Based on the analysis, it was revealed that the main set of attacks that an attacker can implement for a group of mobile robots is denial of service (DoS), distributed denial-of-service (DDoS) attack, a man in the middle (MITM) and a Sybil attack, and exhaustion resources. In addition, there are a number of attacks aimed at the robot positioning system and on other elements of the sensor system, which are not considered in this study. The main purpose of this study is to detect these attacks with a minimum of resources of mobile robots. The offender, implementing an active attack, can influence any physical parameter of the mobile robot through a network or physical impact.

1.1. Maintaining the Integrity of the Specifications

The standard methods used in intrusion detection systems (IDS) may not always have a positive effect when anomalous behavior of mobile robot network nodes is detected. This is due to

several factors: 1. IDS, as a rule, work with the TCP / IP protocol stack, which is not always applicable to mobile robots that transmit data over the radio channel and can use any radio modules and any proprietary protocols; 2. Signature analysis, which is often used in the IDS [3], may be ineffective when an intruder is detected for a group of robots. The behavior of the group robots can vary significantly depending on the task being performed, including the level of network activity, it is quite difficult to create a signature database for the behavior of nodes in the context of each individual task. 3. The computing power of mobile robots is much lower than for standard computer systems for which the IDS is developed [4]. In addition, as a rule, mobile robots can either not be equipped with an operating system or have a "cut-down" version of an operating system with limited capabilities [5]. If the first two problems can be solved by writing their own software for IDS, then the problem of limited energy resources (in the form of insufficiently capacious batteries) makes the use of standard IDS almost impossible for mobile robots [6]. In [7], the authors considered an attack detection system based on the decision tree using the C5.0 algorithm applied to a group of robotic vehicles. The advantage of the presented approach is that for detecting cyber-attacks, the authors, along with four features for analyzing the process of communication and information processing, called cyber input functions, use four parameters for analyzing the physical properties of the robot, which the authors call the physical characteristics of the input signal. Next, the authors conduct 5 types of destructive impact on the robot and get a set of rules for building a decision tree. The

* Elena Basan, E-mail: ele-barannik@yandex.ru

disadvantage of the approach is that in this paper, attacks on only one robot, and not a network of robots were considered. At the same time, the authors considered a limited set of attacks: denial-of-service attacks and attacks aimed at violating physical parameters. In addition, in such systems, there is a need to constantly add rules to detect new attacks. This system is aimed at ensuring the availability of the transmitted data. In [8] an intrusion detection system based on the signature analysis is considered. The authors conducted a series of experiments to create a standard template describing the normal behavior of the robot in the absence of any external influence, as well as random behavioral anomalies. Then a number of situations in which abnormal behaviors occurred caused by environmental conditions were simulated. A normal behavior pattern of the robot based on the collected data with the weighting coefficients calculated on the basis of the frequency of occurrence of a particular type of abnormal behavior. This approach demonstrates greater efficiency in detecting a malicious node than a simple signature analysis; however, there are some disadvantages:

- The need to constantly update the signature database to control data from the new sensors of the mobile robot.
- Conducting analysis of changes only physical parameters of the node and the absence of network analysis of data.

The article [9] considers the system for detecting attacks on unmanned aerial vehicles. The development of this system used the approach based on the creation of a signature database. The system works as follows. Each node of the network has a monitor node, which may be a neighboring unmanned device that fixes the behavior of the trusted node and writes it to the matrix. The monitor node constantly monitors the behavior of the ward node and presents it with estimates. These estimates depend on how much the behavior of the ward deviates from the normal pattern of behavior. Then a rule database is created and the behavior of the node is evaluated. At the same time, the assessment is made on 7 parameters. The authors claim that their system is adaptive and demonstrates a low level of errors of the 1st and 2nd kind when detecting attacks. The disadvantage of the system is the need to constantly monitor the nodes one by one and analyze their behavior, which involves the computational load and network bandwidth.

The article [10] considers the system for detecting the abnormal behavior of robots of the Internet-robots network. The peculiarity of this system is that it has two subsystems. One is a group of robots that collect data using a sensor system and transmit it to the central node that is connected to an external mobile network. The second is a mobile network, where the following modules are available: a data acquisition module, an anomaly classification module, a control command module. The disadvantage of this system is that it is completely centralized; robots do not communicate with each other and act only through an intermediary. Anomaly detection occurs via classifier, which is trained by using training samples preformed. Thus, as a result of studying the works devoted to the topic of detecting attacks on robots, there are three main drawbacks in the existing approaches:

- Most systems based on signature analysis, either on a rules-based system. In this regard, there are the following limitations: the difficulty of detecting new attacks that are not

related to the fixed patterns of the attacker's behavior, as well as the need to keep the database of rules or sets of signatures up-to-date.

- Systems based on fully distributed detection methods require additional energy costs, computational power costs from nodes and increase bandwidth. In addition, if the distributed system is used in conjunction with the signature analysis, the information about the abnormal behavior must be constantly updated, which uses the already limited resources of the robot's memory.
- When using centralized methods, a node that performs basic functions for detecting abnormal behavior is a vulnerability of the system.

In this article, a method for detecting an abnormal behavior of an attacker or several intruders within a group of mobile robots based on probabilistic methods is being developed [11]. The main difference of this method is that it does not require the creation of a standard probability distribution, like other probabilistic methods. The absence of the need to build a reference distribution is due to the fact that the current indications of the node group are taken to reveal the anomalous behavior, then the normal distribution function is constructed and the confidence interval of values is calculated. To estimate the behavior of the Ni node, the probability of the current node indicators entering the confidence interval is calculated, based on the indices of all nodes of the group. Thus, it becomes possible to estimate the probability of the node deflection behavior of the overall behavior of a group of nodes.

2. Method for detecting abnormal behavior

The peculiarity of the proposed method for a group of robots is that for the formation of the normal distribution function it is necessary to obtain data from several nodes performing a similar set of actions. To more accurately determine the degree of deviation of the current indications of a node from a group of nodes, it is necessary that the indicators of a group of nodes are in the same range. In the case of a group of mobile robots that exchange information in one task, this method will work most efficiently. An attacker can affect both the network connection between nodes and the physical parameters of the network node. Table 1 shows the parameters that can be affected by the attacker and the attack by which he can do it.

The parameter packets with data - here it is understood the fact of transfer or redirection of the packet, that is, the availability of the transmitted information is estimated. If there is any impact on the network from the attacker, then there may be situations when packets are discarded, duplicated, etc.

Table 1. The correlation of network parameters and the attacks affecting them

Parameters (indicators of anomalies)	Attacks
Data packets	Black Hole, Gray Hole, False Redirection, Denial of Service, Packet Delay
Remaining battery power	Denial of service, depletion of resources
Network load	Denial of service, resource depletion, the Sybil attack, Flood-attack, Wormhole
Package Integrity	Modification, substitution messages, Man in the middle

The battery charge parameter is the current consumption of the battery (or power consumption), as well as the remaining energy

reserve in the battery pack, which allows the device to function in the network [12].

The Network Load - the total number of packets transmitted on the network. Either the number of packets transmitted through one of the nodes of the network [13].

When detecting attacks such as denial of service and attacks aimed at depleting resources, it is necessary to select those parameters that will be evaluated according to the claimed method. Thus, in the case of a denial of service attack changing network load for a malicious node and network load for the victim. Therefore, it is advisable to estimate the network load, which is the total number of received, sent, and redirected packets of the network node. When an attacker implements an attack aimed at depleting the resources of the node, there will be a sharp decrease in the residual energy level of the node. In addition, an attacker can have superiority in the reserves of energy resources. In the implementation of the Sibyl attack or attack redirecting the impact on themselves, the main purpose of the attacker is to change the processes and routes of data exchange in the network. In other words, an attacker achieves such a situation that all or most of the traffic of neighboring nodes passes through him. Further, the attacker can simply drop received packets, or send them to the wrong nodes. The attacker can achieve this situation in various ways, in this case it is important that the level of incoming traffic will be much higher than that of other nodes of the network. Therefore, it is also necessary to consider the network boot parameter to detect this attack. In the previous work of the authors [14], in addition to the parameters, network loading and residual energy, the parameter of the number of discarded packets P1 was considered. In this study, it will not be considered. Thus, consider two parameters: the network load P2 and the residual energy P3. Changing these parameters affects the state of both the nodes of the network and the entire group of robots in general. The state of the nodes of the network can be described as follows: S1 - the state when the node is not subject to attack and does not conduct the attack itself, i.e. is authentic at the current time; S2 - the state when the behavior of the mobile robot deviated from the behavior of the greater part of the robot group can be observed provided that the node became the victim of the attack, i.e. the node is undefined; S3 - when the behavior of the mobile robot is significantly different from the nodes of the group, i.e. most likely the site is malicious. Figure 1 shows the transition graph from one state to another, and also reflects the effect of parameters and attributes on each other.

The following attributes of the node affect the parameter $P2 = L$: A_{21} - the total number of packets sent by the node containing data. In this model, data transfer uses the UDP protocol and the CBR traffic type. $A_{21} = scbr$; A_{22} is the total number of management packs, or beacons, for testing connections. In this model, packets sent via the ARP protocol act in this role. $A_{22} = sarp$; A_{23} is the total number of packets sent over the routing protocol. This model uses the AODV routing protocol. $A_{23} = saodv$; A_{24} is the total number of received CBR packets. $A_{24} = rcbr$; A_{25} - the total number of ARP packets received. $A_{25} = rarp$; A_{26} is the total number of received AODV packets. $A_{26} = raodv$; A_{27} is the total number of dropped CBR packets. $A_{27} = dcbr$; A_{28} is the total number of dropped ARP packets. $A_{28} = darp$; A_{29} is the total number of discarded AODV packets. $A_{29} = daodv$.

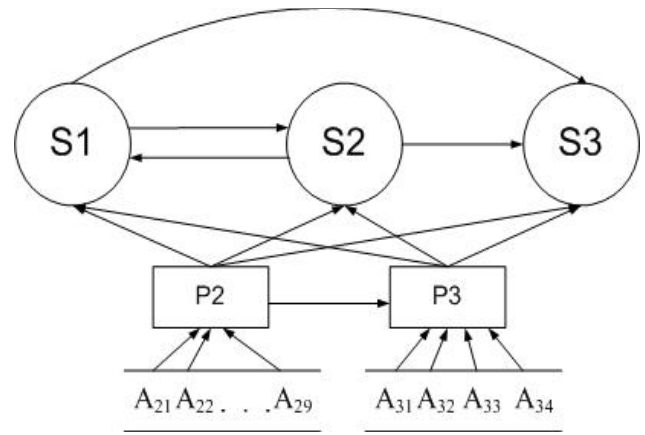


Figure.1. Graph of state of nodes in groups of mobile robots

Thus, the parameter L can be represented by the following equation:

$$L = A_{21} + A_{22} + A_{23} + A_{24} + A_{25} + A_{26} + A_{27} + A_{28} + A_{29} \quad (1)$$

Parameter $P3 = e$ can be characterized by a finite set of attributes A_{3j} . The following attributes affect the amount of residual energy of the node, but apart from the attributes described below, the amount of residual energy is affected by the node's load, i.e. parameter P2: A_{31} is the initial energy reserve of the network node. $A_{31} = initialEnergy$; A_{32} is the power of the transmitted signal. $A_{32} = rxPower$; A_{33} - signal reception power, $A_{33} = txPower$; A_{34} - speed of moving the node. $A_{34} = speed$.

The residual energy is calculated by reducing the level of initial energy A_{31} for each transmitted and each received packet per unit time:

$$\begin{aligned} e_{tx} &= A_{31} - (A_{33} * txtime) \\ e_{rx} &= A_{31} - (A_{32} * rcvtime) \end{aligned} \quad (2)$$

where e_{tx} and e_{rx} this is the level of residual energy after receiving the packet and after the transmission of the first packet; $rcvtime$ - time of packet transmission; $txtime$ - time of reception of a package. A formal description of the method is presented in Section 3, which provides an algorithm implemented in a simulation system for conducting an experimental study.

3. Implementation of the method for detecting abnormal behavior

The simulation of the developed method was carried out in the simulator NS-2.35. The procedure for detecting abnormal behavior and outputting results is called regularly at regular intervals. The start and end time of this process is set by the user in the script using a special command. The command handler plans to start a special timer for the start time of the process, and the timer and end time are written to the timer parameters [15]. To account for the parameters e and L of the mobile robot in the assembly model was added a special object that counts the number of packets transmitted / sent / forwarded node and the residual energy of the node [16]. The procedure for calculating trust works according to the following algorithm representing at the table 2.

Table 2: Calculation of the trust level for nodes in the group of mobile robots

No.	Name of equation	Equation	Description
1 The calculation of the confidence interval boundaries			
1.1	Variance for L parameter	$D_{Li} = \left(\sum_i^N (L_i - \bar{L})^2 \right) / n$	D_{Li} - variance of the parameters L calculated for the group of nodes in the current time interval
1.2	Variance for e parameter	$D_{ei} = \left(\sum_i^N (e_i - \bar{e})^2 \right) / n$	D_{ei} - variance of the parameters e calculated for the group of nodes in the current time interval
1.3	The standard deviation for L parameter	$\sigma_{Li} = \sqrt{D_{Li}}$	σ_{Li} - is the standard deviation of the parameter L which calculated for the group of nodes in the current time interval.
1.4	The standard deviation for e parameter	$\sigma_{ei} = \sqrt{D_{ei}}$	σ_{ei} - is the standard deviation of the parameter e which calculated for the group of nodes in the current time interval.
1.5	Argument of the Laplace function - t	$\Phi(t) = \frac{\alpha}{2}$	$\Phi(t)$ - is the Laplace function; α is a given reliability, in this study the value of the coefficient is equal to $\alpha = 0.98$, so the argument $t = 2.34$;
1.6	The limits of the confidence interval for the e parameter	$a_{min} = \bar{e} - t \cdot \sigma_e / \sqrt{n},$ $a_{max} = A_{31}; a_{min} < a_{max}$	The upper bound of the confidence interval of the parameter e is always equal to the maximum permissible energy value, that is, $a_{max} = initialEnergy$. This is due to the fact that nodes can migrate from one group to another; new nodes may appear with a residual energy value equal to the initial value. a_{min} - lower bound of confidence interval. $t \cdot \sigma / \sqrt{n}$ - is the accuracy of the estimation. n - Total number of nodes.
1.7	The limits of the confidence interval for the L parameter	$b_{min} = L_{min},$ $b_{max} = \bar{L} + t \cdot \sigma_L / \sqrt{n}$	The lower bound for the parameter L is equal to the minimum required number of packets passed through the node in one time interval L_{min} . These measures are taken because the mobile robot can exhibit selfish behavior, that is, refuse to participate in the network to save energy, which can artificially "understate" the boundaries of the interval. b_{max} - the upper bound of confidence interval for L parameter
2	Determination of the probability of anomalous behavior of the mobile robot on the basis of the calculated confidence intervals.		In order to calculate the mean square deviation and mathematical expectation, it is necessary to shorten the interval for which the value is calculated and take into account only the node parameters in the previous time interval L_{i-1}, e_{i-1} and L_i, e_i the node parameters for the current interval. Note: If you take the parameter values over the entire time interval, the standard deviation is too large, due to the large difference between the start and end values.
2.1	The mathematical expectation for the e parameter	$\bar{e}_g = (e_{i-1} + e_i) / 2$	\bar{e}_g - mathematical expectation of the values of the e parameter for the sampling interval, which calculated for individual node
2.2	The mathematical expectation for the L parameter	$\bar{L}_g = (L_{i-1} + L_i) / 2$	\bar{L}_g - mathematical expectation of the values of the L parameter for the sampling interval, which calculated for individual node
2.3	The variance for the e parameter	$D_{e_g} = \left(\sum_i^N (e_i - \bar{e}_g)^2 \right) / n$	D_{e_g} - variance for the sampling interval for e , which calculated for individual node.
2.4	The variance for the L parameter	$D_{L_g} = \left(\sum_i^N (L_i - \bar{L}_g)^2 \right) / n$	D_{L_g} - variance for the sampling interval for L , which calculated for individual node.
2.5	The standard deviation for the e parameter	$\sigma_{e_g} = \sqrt{D_{e_g}}$	σ_{e_g} - the standard deviation for the sampling interval for the residual energy, which calculated for individual node.
2.6	The standard deviation for the L parameter	$\sigma_{L_g} = \sqrt{D_{L_g}}$	σ_{L_g} - the standard deviation for the sampling interval for the L parameter, which calculated for individual node.

2.7	The probability of the value for parameter e falling into the confidence interval	$P_e(a_{\min} < e_i < a_{\max}) =$ $= \Phi\left(\frac{a_{\max} - \bar{e}_e}{\sigma_{e_e}}\right) - \Phi\left(\frac{a_{\min} - \bar{e}_e}{\sigma_{e_e}}\right)$	P_{e_s} - the probability of deviations the network load from confidence interval, which calculated for individual node. Φ is a Laplace function.
2.8	The probability of the value for parameter L falling into the confidence interval	$P_L(b_{\min} < L_i < b_{\max}) =$ $= \Phi\left(\frac{b_{\max} - \bar{L}_e}{\sigma_{L_e}}\right) - \Phi\left(\frac{b_{\min} - \bar{L}_e}{\sigma_{L_e}}\right)$	P_L - the probability of deviations the residual energy from confidence interval, which calculated for individual node.
2.9	The resulting probability value that the node is trusted	$P_{sum} = P_e * P_L$	To obtain the resulting probability value, it is necessary to use a combination of the values of P_{e_s} , P_L of the direct value of trust P_{sum} in [17], an algorithm for combining confidence values using the Bayes theorem is presented.
3.	Deciding on the degree of trust in the node	$P_{sum} > 0,5;$ $P_{sum} = 0,5;$ $P_{sum} < 0,5.$	Assume threshold probability that the node is abnormal equal to 0.5. When the node reaches a value of 0.5, it is necessary to reduce its residual energy level by half, then the node is considered in an undefined state . Further, if the value of the confidence level reaches the level of 0.4, then it is necessary to consider the node malicious and reduce its energy level to zero, thus, the node is excluded from the network [18].

4. Experimental study, evaluation of the effectiveness of the developed method.

The model of a robot group in the simulation environment NS-2.35 was developed. Robots communicate with each other via wireless communication and use the TCP / IP protocol stack to transfer information. In particular, the UDP protocol is used for data transmission at the transport level, the ARP protocol is used to transmit control commands at the data link layer, the AODV protocol is used for routing the packets [19]. Figure 2 shows a group of mobile robots in the modeling system, which includes 10 nodes. Of these, one N4 node is a base station or a central server. The node N0 is the group leader and performs the functions of gathering information from the other robots and redirects it to the central server. The nodes of the group exchange information with each other and with the group leader [20]. In this case, nodes N6, N7, N8, N9 will conduct a DDoS attack starting from 50 seconds of network operation.

4.1. Implementation and detection of a DoS attack.

Conducting denial of service attack, the attacker creates a situation where the network node becomes unavailable to other nodes and cannot respond to their requests and work normally. An attack aimed at depleting resources, as a rule, creates such conditions for a node that it begins to lose more energy than in the absence of an attack. These attacks are interrelated. In fact, the goal of a DoS attack can be to completely disable a node by exhausting the node's resources. The developed model of a group of mobile robots is assumed that the nodes spend energy on the transmission and reception of packets. Therefore, an attacker "forces" trusted nodes to spend more energy than when working in normal mode, sending a large number of packets to the network.

Three types of situations were simulated. In the first case, an estimate was made of the energy consumed by network nodes in the absence of an attack. In Figure 3, this situation is represented by a blue chart marked with rhombuses.

In the second case, an attack was conducted on the network, while the traffic of the malicious node is $I_t \leq I_m \leq 2I_t$, where I_m is the traffic of the malicious node, I_t is the traffic of the authentic node. The third graph represents a situation where an attack is carried out intensively and $I_m > 2I_t$.

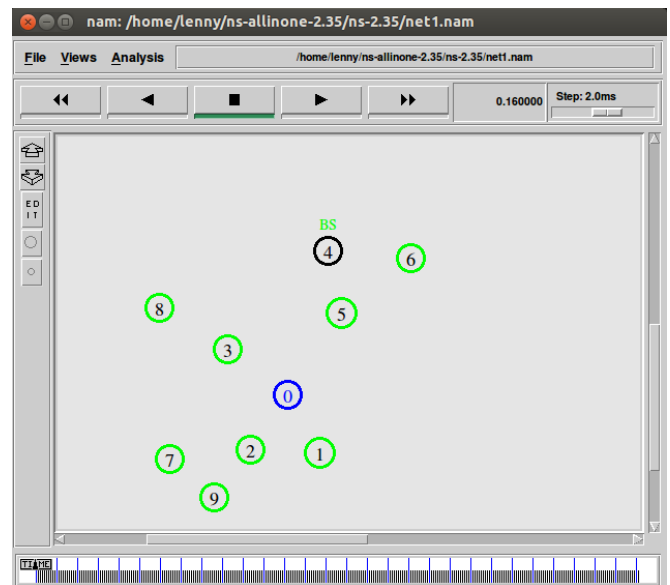


Figure 2. Group of robots in the simulation system NS-2.35

Figure 3 shows that during a non-intensive attack, the energy level of the nodes will remain almost the same as for the case when the attack is not carried out. That is, in this case, the attack can be considered ineffective. The graph showing the change in the energy level during an intense attack shows a sharp drop in the energy level, which confirms the effectiveness of the attack. At the same time, the load of the attacker's node is more than twice the workload of authentic nodes. In this case, the developed method for detecting abnormal behavior allows us to identify a malicious node.

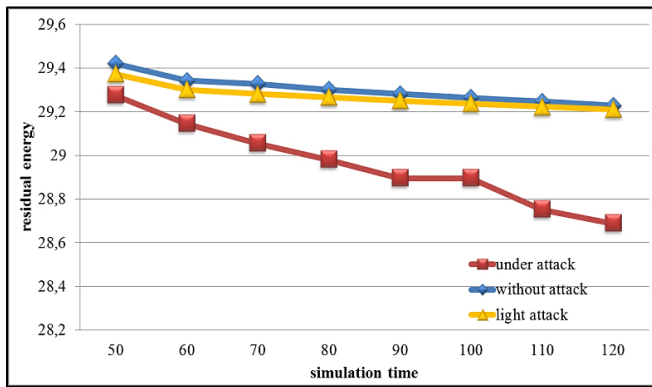


Figure 3. Change in the level of residual energy, depending on traffic intensity of network nodes.

The proposed method allows in a few seconds to detect a DoS attack, if it has a significant impact on the resources of network nodes and helps to increase the level of traffic. In Figure 4, a graph showing the level of detection of an attack, given that an attacker starts an attack after the 50th second of simulation, we can say that the attack is almost immediately detected. Since on an interval of time between 50-60th seconds the malicious node has a hit level in the confidence interval of 0.4 and is already blocked by the system for 60 seconds.

4.2. Implementation and detection of DDoS attacks.

The detection of a distributed denial of service attack is more difficult. This is due to the fact that when the number of malicious nodes prevails over the number of authentic nodes, the boundaries of the confidence interval are significantly expanded. Especially if malicious nodes conduct an attack with varying intensity. Nevertheless, the developed method is quite effective in detecting this attack. When the ratio of malicious nodes to trusted hosts is 4 to 5, the method allows to immediately detect all malicious nodes and block them already in the second time interval.

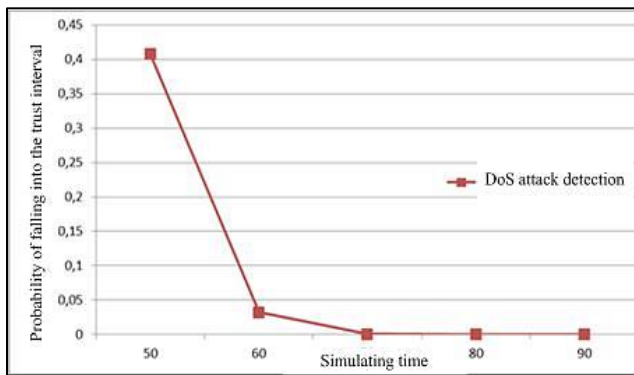
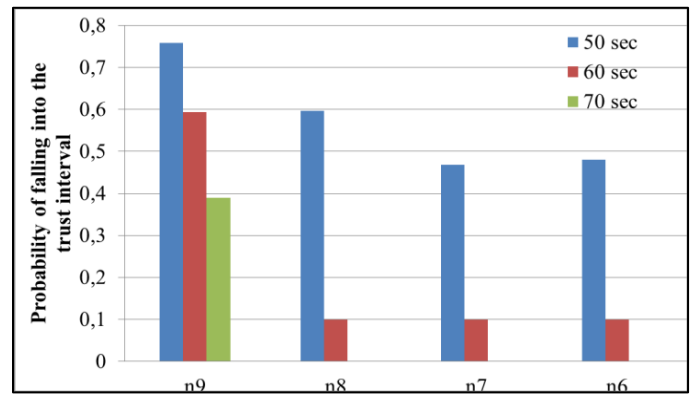


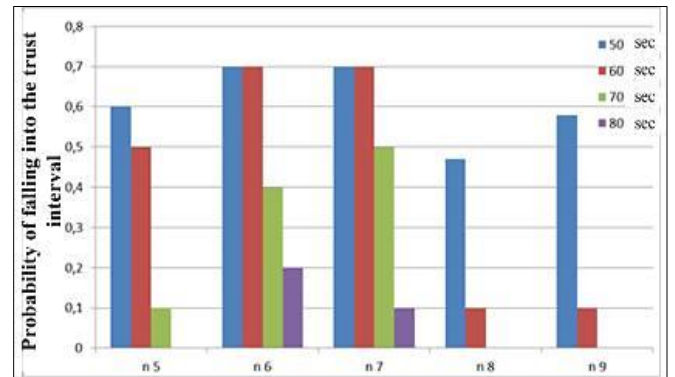
Figure 4. The probability of hit by the load and residual energy values of the malicious node in the trust interval

Figure 5 (a) presents a histogram showing the level of hit of current indicators e and L of malicious nodes in the confidence interval. When malicious and trusted hosts are in an equal ratio of 5 to 5, the quality of detection becomes worse.

Figure 5 (b) shows a histogram showing the detection level of malicious nodes. N9 and N8 nodes were also detected in the second interval, nodes N5 and N6 were detected in the third interval and node N7 in the fourth time interval, starting from the moment when the attack began. In general it can be said that the detection rate of 100%, but the rate of detection decreased.



(a)



(b)

Figure 5. The level of detection of an attacker in a distributed denial of service attack for (a) four malicious hosts (b) for five malicious nodes and five authentic hosts

When the number of malicious nodes exceeds the number of trusted in the ratio of 6 malicious to 5 authentic, the detection level is 83%, i.e. one malicious node remains undetected. Figure 6 shows a histogram of the level of hit of current values of malicious nodes in the confidence interval. In this case, three nodes: N8, N9, N5 - are detected in the second time interval. One node N6 in the third interval and one node N10 in the 4th interval, only node n7 remains undetected on the fourth interval, most likely if the attack continues at the same rate, then this node will be detected on the 5th interval. But this time is high enough to detect an attack [21]. Nevertheless, the developed method shows a sufficiently high speed of detection of attack, even if the number of malicious nodes is more than the number of trusted ones.

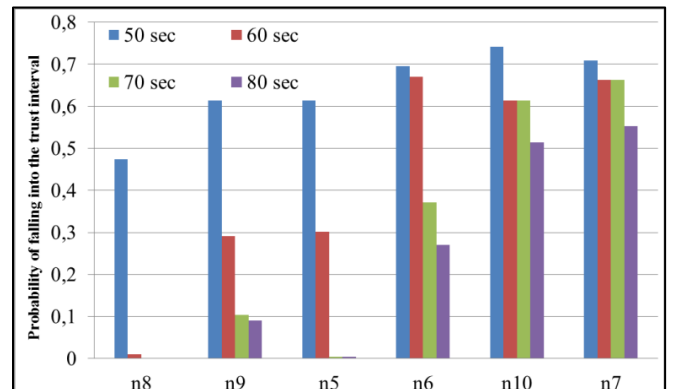


Figure 6. The detection level of nodes with distributed attack denial of service for 6 malicious nodes and 5 trusted ones

4.3. Implementation and detection of the Sibyl attack.

The Sibyl attack is that an attacker is represented by several network nodes and tries to redirect most of the traffic to itself [22]. At the same time, it can make a destructive impact on the network by discarding messages, redirecting them to the wrong nodes, violating the routing scheme, or can passively listen for traffic.

At the same time to detect an attack, when an attacker does not have a destructive effect is quite difficult. In the works of the authors [23], as a rule, there are methods using hard protection: password protection, cryptographic protection, as well as signature analysis and group detection. These methods are used in networks MANET, IoF, P2P [24], which are not so much limited in resources as groups of mobile robots.

The developed method for detecting abnormal behavior shows the effectiveness of detection of the Sibyl attack, even if the attacker redirects the traffic to himself and does not take any further action. In this case, detection is possible by changing the load level of nodes that conduct an attack on neighboring nodes. In addition, the level of residual energy of the attacking nodes is significantly reduced. To assess the method, malicious N7-N11 nodes were added to the robot group in the NS-2.35 simulation system, which are called (Sybil1-Sybil5). Figure 7 shows the topology of the network, taking into account malicious nodes. The figure shows that the number of malicious nodes and the number of trusted ones, excluding the base station (BS) and the group leader (GL), corresponds to half of the network nodes.

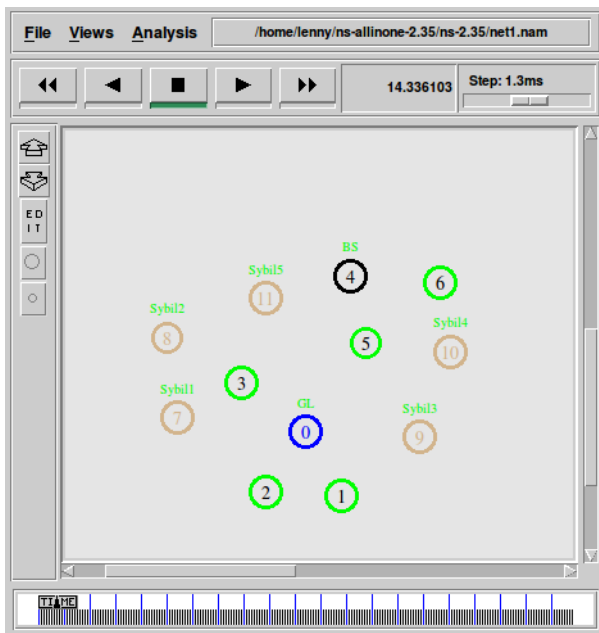


Figure 7. The network topology for the implementation of the Sibyl attack.

N7-N11 nodes redirect packets from neighboring mobile robots to themselves starting from 50 seconds, thus disrupting the network operation scheme. Initially, mobile robots will send packets to the group leader in a predetermined pattern; the leader of the base station sends packets. Thus, in the first 10 seconds of the attack, the method allows you to identify 2 malicious nodes N10 and N11. This is due to the fact that these nodes redirect more traffic to themselves. Further, starting from the 60th second, the detected nodes are blocked and nodes N7 and N8 are detected at

the 70th second. The most difficult for detection was the node N9, this is due to its relatively low activity for redirecting traffic, at the time of detection the level of congestion of this node is less than twice the level of congestion of other nodes. Figure 8 shows a histogram representing the detection level of malicious nodes.

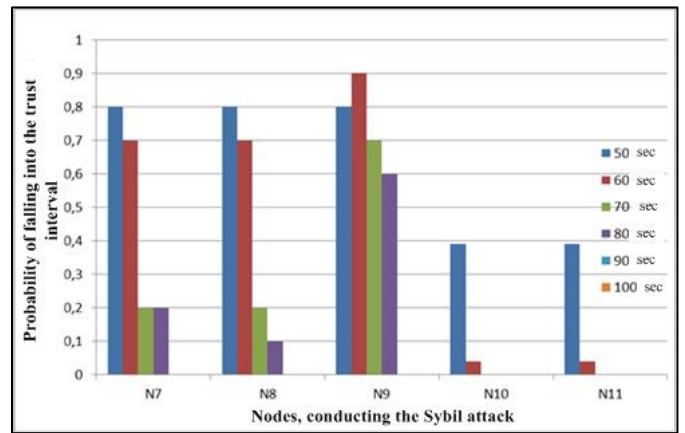


Figure 8. detection level nodes conductive the Sibyl attack

5. Conclusion

The issues related to the security of mobile robots and, in particular, group management of mobile robots, are currently being addressed by a limited number of scientists and institutions. Nevertheless, the subject of research is quite relevant, in connection with the widespread use of robotic systems. The developed method is a versatile tool for detecting anomalous behavior for a group of nodes, when it is possible to conduct an analysis of the behavior of most nodes and identify single or mass deviations from general behavior. Due to this, it is possible to increase the number of analyzed parameters for expanding the range of attacks. The method demonstrates a sufficiently high level of detection, namely it detects all malicious nodes within 30-40 seconds. Limitations of the method consist in the number of malicious nodes that conduct an attack on the network; their number should not exceed 60%, compared to the number of trusted ones. And also this method is applicable to a group of nodes that perform a similar task. In contrast to existing methods of detecting abnormal behavior, where one has to be comprehensive signature database or rules databases, store them, and then update the developed method makes it possible to detect anomalies in the current time, and depending on the current situation. This advantage is quite important because mobile robots can be used in different environments and for various tasks. In this case, in addition to networking protocols between mobile robots can change and environmental conditions, which in turn also influence the occurrence of anomalies and errors.

This method takes into account the threshold values, that is, the permissible level of anomalies, which reduces the number of false positives. In this case, the developed method allows to reduce the load on the mobile robot, it does not need to constantly exchange messages, monitor neighboring nodes and update, store the signature database. In the future study, it is proposed to add the node mobility parameter, to estimate the coordinates and speed of its movement. Evaluation of these parameters will allow to detect attacks such as interception by management, when an attacker, captures a trusted node and manages it independently.

Estimation of position and speed of movement will detect abnormal behavior nodes.

Acknowledgment

The work was supported by the Ministry of Education and Science of the Russian Federation (Initiative Science Projects No. 2.6244.2017 / 8.9).

References

- [1] A. Basan, E. Basan, O. Makarevich. "A Trust Evaluation Method for Active Attack Counteraction in Wireless Sensor Networks". in 9th International Conference on Cyber-enabled distributed computing and knowledge discovery. Nanjing, China.2017.P. 369-372. DOI: 10.1109/CyberC.2017.14.
- [2] Basan A.S., Basan E.S. "Threat model for mobile robot group management systems". System synthesis and applied synergetics. Collection of proceedings of the VIII All-Russian Scientific Conference. - Sistemy sintez i prikladnaya sinergetika. Sbornik nauchnykh trudov VIII Vserossiyskoy nauchnoy konferentsii. Rostov-on-Don: Southern Federal University. 2017. C. 205-212. - (In Russ.)
- [3] H.S. Kim, S.W. Lee. "Enhanced novel access control protocol over wireless sensor networks". IEEE Transactions on Consumer Electronics. 2009. № 55 (2). pp. 492 - 498.
- [4] Branitsky A.A., Kotenko I.V. "Analysis and classification of methods for detecting network attacks", Information security. Proceedings of SPIIRAN - Informatsionnaya bezopasnost'. Trudy SPIIRAN.2016. №2(45). pp.207-244.(In Russ.)
- [5] Petrovsky O. "Attack on the drones". Virus bulletin conference. 2015. pp. 16-24.
- [6] Garber L. Robot OS: "A New Day for Robot Design". Computer. № 46 (12). 2013. pp. 16-20.
- [7] Kozhemyakin I.V., Putintsev I.A., Semenov N.N., Chemodanov M.N. "Development of an underwater robotic complex, using open simulation tools, supplemented with a model of hydroacoustic interaction". News of SFedU. Technical science. Section II. Marine robotics. - Izvestiya YUFU. Tekhnicheskkiye nauki. Razdel II. Morskaya robototekhnika. 2016. № 1 (174). pp.88-99.(In Russ.)
- [8] Vuong T. P., Loukas G., Gan D., Bezemskij A. "Decision tree-based detection of denial of service and command injection attacks on robotic vehicles". 2015 IEEE International Workshop on Information Forensics and Security (WIFS) 2015. pp. 1-6.
- [9] Bezemskij A., Loukas G., Richard J. Gan D. "Behavior-based anomaly detection of cyber-physical attacks on a robotic vehicle". 15th International Conference on Ubiquitous Computing and Communications and 2016 8th International Symposium on Cyberspace and Security. 2016. pp. 61-68.
- [10] Mitchell R., Chen I.R. "Adaptive Intrusion Detection of Malicious Unmanned Air Vehicles Using Behavior Rule Specifications". IEEE Transactions on Systems, Man, and Cybernetics: Systems. №44 (5). 2014. pp. 593 -599
- [11] Monshizadeh M., Khatri V., Kantola R., Yan Z. "An Orchestrated Security Platform for Internet of Robots". Springer International International Conference on Green, Pervasive, and Cloud Computing. 2017. pp. 298–312.
- [12] Basan A.S., Basan E.S., Makarevich O.B. "Method of counteracting active attacks of an attacker in wireless sensor networks". News of SFedU. Technical science. №5 (190).2017. pp. 16-25. (In Russ.)
- [13] Basan A., Basan E., Makarevich O. "Methodology of Countering Attacks for Wireless Sensor Networks Based on Trust". 8th International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC).Chengdu.2016. pp.409 – 412.
- [14] Schoch E., Feiri M., Kargl F., Weber M. "Simulation of Ad Hoc Networks: ns-2 compared to JiST/SWANS SIMUTools". First International Conference on Simulation Tools and Techniques for Communications, Networks and Systems. 2008. pp. 34 - 41.
- [15] He W., Yang S., Teng D., Hu Y. "A Link Level Load-Aware Queue Scheduling algorithm on MAC layer for wireless mesh networks". International Conference on Wireless Communications & Signal Processing.2009. Nanjing, China.pp.1-16
- [16] Basan A.S., Basan E.S., Makarevich O.B. "Development of the Hierarchical Trust management System for Mobile Cluster-based Wireless Sensor Network". Proceeding SIN '16 Proceedings of the 9th International Conference on Security of Information and Networks. 2016. pp. 116-122.
- [17] Mohammad Momani. Bayesian Fusion Algorithm for Inferring Trust in Wireless Sensor Networks // Journal of Networks 5(7). 2010; C. 815-822. DOI: 10.4304/jnw.5.7.815-822
- [18] Abramov E.S., Basan E.S. "Development of a model of a protected cluster wireless sensor network". News of SFedU. Technical science. - Izvestiya YUFU. Tekhnicheskkiye nauki.2013. № 12(149). pp. 48-56. (In Russ.)
- [19] Ferronato J. J. Sandini Trentin M. A. "Analysis of Routing Protocols OLSR, AODV and ZRP in Real Urban Vehicular Scenario with Density Variation". IEEE Latin America Transactions. 2017. № 15 (9). pp. 1727 – 1734.
- [20] Pshikhopov V.Kh., Soloviev V.V., Titov A.E., Finaev V.I., Shapovalov I.O. "Group management of mobile objects in uncertain environments". Ed. V.H. Pshihopova. M.: FIZMATLIT. - Pod red. V.KH. Pshikhopova. M.: FIZMATLIT IoT 2015. – pp. 233-270. (In Russ.)
- [21] Sargeant I., Tomlinson A. "Maliciously Manipulating a Robotic Swarm". Int'l Conf. Embedded Systems, Cyber-physical Systems, & Applications. ESCS'16. 2016. pp. 122- 128.
- [22] Abbas S., Merabti M., D. Llewellyn-Jones, K. Kifayat. "Lightweight Sybil Attack Detection in MANETs". IEEE system journal, №. 7, (2). 2013. pp 236-248.
- [23] Patel S.T., Mistry N.H. "A review: Sybil attack detection techniques in WSN". 4th International Conference on Electronics and Communication Systems (ICECS). 2017. pp. 184 – 188.
- [24] Wang G., Musau F., Guo S., Abdullahi M. B. "Neighbor Similarity Trust against Sybil Attack in P2P E-Commerce". IEEE transactions on parallel and distributed systems. № 26 (3). 2015. pp. 824-833.

An Empirical Study of Icon Recognition in a Virtual Gallery Interface

Denise Ashe, Alan Eardley*, Bobbie Fletcher

School of Computing and Digital Technologies, Staffordshire University, ST4 2DE, U.K.

ARTICLE INFO

Article history:

Received: 13 September 2018

Accepted: 07 November 2018

Online: 29 November 2018

Keywords:

Interface usability

Icon recognition

Icon intuitiveness

Icon design

ABSTRACT

This paper reports on an empirical study (an extension of a pilot study) that analyses the design of icons in a German 3-D virtual art gallery interface. It evaluates the extent to which a sample of typical computer users from a range of ages, educational attainments and employments can interpret the meaning of icons from the virtual interface taken 'out of context' and 'in context'. The study assessed a sample of 21 icons representing the 'action', 'information' and 'navigation' functions of the virtual interface using a new Icon Recognition Testing method (IRT) developed by the researchers from existing usability test methods. The Icon Recognition Rate (IRR) of the icons was calculated and they were classified as 'identifiable', 'mediocre' or 'vague' in a novel and useful classification system. The IRT results show that the IRR of almost a quarter of the icons was below the 'identifiable' standard, which could seriously compromise the usability of a virtual interface. A comparison is made, using textual and thematic analysis, between the participants' understanding of the icons' meaning in and out of context and of the effect of positioning icons in relation to their virtual surroundings and of grouping them in tool bars. From the findings of the study, conclusions are drawn, and recommendations are made for economical icon redesign and replacement. It is suggested in the conclusions that further research is needed into how designers' conceptual models can be better matched to users' mental models in the design of virtual interfaces by bringing user profiles into the study.

1. Introduction

This paper is an extension of a pilot study by Ashe *et al.* [1] into the effectiveness of icon design in a virtual gallery interface that was presented in the e-Tourism stream of the International Conference in Information Management at Oxford University in May 2018 (ICIM2018). Experience with that pilot study and feedback from reviewers informed a fuller research project, which forms the extended work in this paper. The second part of the research project used a larger, more representative sample of participants, took account of the context within which the icons are understood and added textual and thematic analysis to the research. The present paper therefore contains more detail about the research methodology and the data analysis and its results, which will allow the research to be replicated. The pilot study examined a sample of virtual tours of museums and galleries, including the Smithsonian Natural History Museum in Washington, D.C. [2], the Louvre in Paris [3], Oxford University Museum of Natural History [4] and the portal Virtual Tours [5,] that currently includes more than 300 'Museums, exhibits, points of special interest and real-time

journeys' [6]. The study showed that icons are an important part of this generation of virtual interfaces as the main way of performing interactive tasks such as navigation, initiating actions and obtaining information [7]. The virtual interface itself is a complex sign system [8] containing components (e.g. buttons, icons and scroll bars) through which the user interacts with the system [9]. The icons can be symbols, images or pictures [10] that communicate meaning [8] without textual description [11-12].

This provides icon-based interfaces with the potential to overcome language barriers [10, 13], which can be important in an international context such as a cultural attraction. Icons used as shortcuts to a function (e.g. a printing icon in a word processing package) should provide the user with a memory aid to increase his or her ability to recall and to recognize the intended function without needing further instructions [14-15]. Successful recognition depends on the user's familiarity with that type of interface and experience of using that icon [1] and greater familiarity and experience should therefore allow a more abstract (i.e. less concrete) icon to be used in the design of the interface.

*Corresponding Author: Alan Eardley, Email: w.a.eardley@staffs.ac.uk

www.astesj.com

<https://dx.doi.org/10.25046/aj030637>

Gatsou [15] cites the work of Nadin [16], who uses a calculator icon to demonstrate the principles of concreteness and abstraction, as shown in Figure 1.

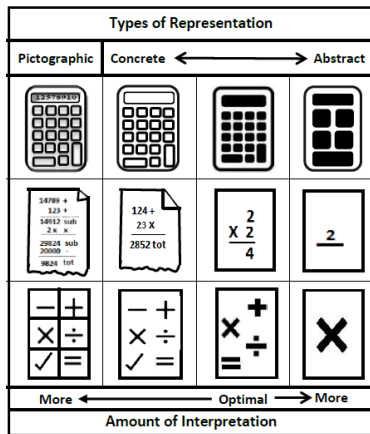


Figure 1: Types of icon representation
Adapted from Nadin [16] and Gatsou, et al., [15]

Scalisi [17] suggests that users need an initial period of learning the interface to understand the icons through ‘visual codification’. This may come easily with an office package that is used every day but may not be possible with a rarely-used interface such as a virtual gallery [1]. Icons may resemble to a greater or lesser extent the objects or functions that they represent [17] and the closeness of this relationship is the ‘semantic distance’, which is “important in determining the success of icon usability” [18].

Arnheim [19] discusses the relationship between ‘concreteness and abstraction’ stating that, “Images can serve as pictures or as symbols; they can also be used as mere signs”, implying that increased user familiarity can allow an icon to be simplified yet still allow the user to understand it. The pilot study supported this view, suggesting that the closer the semantic distance, the more likely the users were to understand the icon’s function and meaning. Conversely, the more abstract the icon (i.e. the greater its semantic distance) the more generally useful it could be in a variety of contexts, although correct recognition of the icon’s meaning could be more difficult in a specific case.

For example, the icon for printing a document could be a photograph of the actual printer to which the file could be sent. That would promote easy recognition and could be useful, for instance to locate the correct printer in a room, but would involve having a different icon for every available model of printer. This would make it difficult for users to learn the general meaning of the ‘printer’ icon in other instances and applications.

2. Icon Usability

The pilot study [1] reviewed the literature on icon usability testing and took the definition of Ferreira et al. [20] who cite the work of Barr, et al. [14] stating that an icon is successful, “...if the interpretant of the user [i.e. the user’s understanding] matches the object that the designer had intended with that sign, and [it is] unsuccessful otherwise” [20, p 2]. In other words, a recognizable (i.e. ‘identifiable’) icon should be easy to interpret and be unambiguous in order for it to succeed. This formed the baseline ‘measure of success’ used in the pilot study.

A range of different icon usability testing methods were reviewed in the pilot study, such as Icon Understandability Testing [21], [12], Test with Comparison [13], Matching Method [22], Icon Intuitiveness Testing [23] and Standard Usability Icon Testing [23]. From this review the Icon Intuitiveness Test (IIT) was selected for the study. The method was felt to be the most suitable as it seeks to find out how well users interpret and recognize icons using their existing insight and experience. Nielsen and Sano [23] describe a paper-based IIT as used by Sun Microsystems. Ferreira et al. [20] used a paper-based IIT and Foster [24] suggests that the IIT can be administered on a computer or on paper. Bhutar et al. [13] conducted a similar ‘test without context’ using an MS PowerPoint® presentation and paper-based questionnaires.

3. Pilot Study

Extending a previous study by Bhutar et al. [13], the modified IIT used in the pilot study adhered to the following guidelines:

- With one exception (i.e. Icon 1) the icons did not have text labels attached [23, 25] so their effectiveness relied entirely on their functioning as signs;
- The icons were not displayed in the actual interface (i.e. they were taken out of context), so the participants had no external visual cues to their meaning;
- Only one icon was made visible at a time so participants had no clues to their meaning from their sequence or by association.

Previous studies by Ferreira et al. [20] had used the standard ISO 9186:2001 benchmark [26] of 66% for successful icon recognition. Gatsou et al. [15] adopted the more stringent standard ISO 3864:1984 [27] which has a slightly higher benchmark, in which a success rate above 66.7 % was considered as ‘good’ and below that as ‘low’. A similar scale by Howell & Fuchs [28], was adapted for use in the pilot study. With this scale, icons achieving 60% Icon Recognition Rate (IRR) or above are classed as ‘identifiable’, whereas icons scoring less than that are ‘unsuccessful’ in conveying their meaning.

The adaptation for the pilot study further divided these ‘unsuccessful’ icons into ‘mediocre’ (30% - 59% IRR) and ‘vague’ (0% - 29% IRR) as shown in Table 1.

Table 1: Icon Recognition Rates and Classifications

Icon recognition rate (IRR) classification	
IRR (%)	Classification
60 – 100	Identifiable
30 – 59	Mediocre
0 – 29	Vague

The research required as a subject an advanced interface containing icons that are capable of a number of different interpretations and which carry out defined functions. A virtual art gallery was felt to meet these requirements and a search on the World Wide Web identified more than 100 possible candidates. A German 3-D virtual art gallery was eventually selected for the test, as it was felt to be representative of its type [1]. For ethical reasons the site is referred to as ‘Artweb.com’. The test examined the

users’ understanding of the icons when taken ‘out of context’ (i.e. without reference to their use in the actual interface).

3.1. Icon Intuitiveness Test

All 21 icons used in the pilot study IIT were selected from the ‘Artweb.com’ virtual art gallery interface [1], which is close to the recommended number of 20 used in a previous study by Nielsen and Sano [23]. These icons were taken at random either individually or from grouped toolbars from various parts of the interface. The icons were designed for various basic interface functions (i.e. carrying out navigation, initiating an action and obtaining information) and are depicted in Table 2, labelled according to their function or purpose.

Table 2: The 21 Icons Evaluated in the Tests

Images of 21 evaluation icons								
1		2		3		4		
	Action		Action		Information		Navigation	
5		6		7		8		
	Information		Action		Action		Navigation	
9		10		11		12		
	Navigation		Action		Action		Action	
13		14		15		16		
	Action		Navigation		Navigation		Information	
17		18		19		20		
	Information		Action		Navigation		Navigation	
21								
	Navigation							

3.2. Test Sample

Five users consented to take part in the pilot study [1] to evaluate icons by participating in the IIT of icons displayed ‘out of context’. The choice of a small sample size in this type of research was based on the studies of icon usability by Nielsen and Sano [23] to collect rich data. The pilot sample included one female and four males - a ratio that is proportionate to the gender balance of the organization in which the tests were conducted. All the participants fell within the age range 20 - 29 years and all had good eyesight and no obvious disabilities.

None of the participants had previously used the ‘Artweb.com’ virtual art gallery, although 80% had experience of using another virtual tour and had used other 3-D virtual worlds. All the participants had more than ten years’ experience of using personal computers and most of the participants fell within the range of 10 to 14 years’ experience, as shown in Table 3. This may be because

most of the participants in the study were university students undertaking a technology-related degree course.

Table 3: Experience of Using Computers in Years

Years’ experience of using a personal computer	
Range of experience	No. of users
0 – 4 years	0
5 – 9 years	0
10 -14 years	4
15 -19 years	1
20 -25 years	0

Most of the subjects fell into the range of 30 – 44 hours of weekly computer use, with one subject exceeding 60 hours, as shown in Table 4.

Table 4: Computer Use per Week in Hours

Hours of computer use per week	
Range of No. of hours	No. of users
65	0
15 – 29	0
30 – 44	4
45 – 59	0
60 +	1

3.3. Test Procedure

The IIT in the pilot study used a variety of the commonly-used ‘card sorting’ technique [29]. The participants were provided with brief details of the test scenario as in previous studies of this type [30]. The test administrator then conducted the IIT with the participants individually, each session lasting approximately forty-five minutes [1]. This procedure was repeated and the participant’s interpretation of the icons’ meaning or function was noted until all 21 cards had been displayed. An overall results table was produced by calculating the IRR expressed as a percentage for each of the icons using the formula:

$$\frac{\text{(No. of correct responses / No. of participants)} \times 100}{\text{= Icon Recognition Rate \%}}$$

3.4. Results for Icons ‘Out of Context’

The IIT results for all 21 icons tested ‘out of context’ were placed into the chosen icon classification (i.e. ‘identifiable’, ‘mediocre’ and ‘vague’) based on the participants correctly interpreting their meanings or functions. In the pilot test, fifteen icons (i.e. 71.4% of the set of 21 icons) were classed as ‘identifiable’, one was classed as ‘mediocre’ (i.e. 4.8% of the set) and five were classed as ‘vague’ (i.e. 23.8%). This high proportion of ‘identifiable’ icons could suggest that the designs were generally successful in this interface. However, the meaning of 28.6% of the icons (i.e. the ‘mediocre’ and ‘vague’ classes) was misinterpreted or confused, which could seriously compromise

the usability of the interface in practice. For the purposes of the pilot study [1] a ‘traffic light’ system was used to indicate the icons’ classification according to their IRR score, from best to worst, (i.e. green applies to ‘identifiable’ icons, amber to ‘mediocre’ icons and red to ‘vague’ icons) as in Table 5.

Table 5: Classification of Icons by IRR Score

Classification of icons as identifiable, mediocre or vague					
No.	Image	Meaning	Score	%	Class
1		Start Virtual Tour	5/5	100.0	Ident.
2		Previous tour position, pause tour, next position.	5/5	100.0	Ident.
3		Exhibition information	5/5	100.0	Ident.
8		Previous artwork to the left	5/5	100.0	Ident.
10		Play animation button to circle artwork	5/5	100.0	Ident.
11		Pause animation button to circle artwork.	5/5	100.0	Ident.
13		Pan and zoom image.	5/5	100.0	Ident.
16		Information on artwork.	5/5	100.0	Ident.
17		Contact the exhibitor (by email).	5/5	100.0	Ident.
19		Navigation arrow buttons	5/5	100.0	Ident.
5		Help information for navigation.	4/5	80.0	Ident.
14		Next artwork to the right	4/5	80.0	Ident.
6		Full screen of virtual exhibition.	3/5	60.0	Ident.
7		Return to screen to window size.	3/5	60.0	Ident.
18		Close window button.	3/5	60.0	Ident.
12		Slider to zoom in & out of image.	2/5	40.0	Med.
9		Rotate left (anti-clockwise)	1/5	20.0	Vague
15		Rotate right (clockwise).	1/5	20.0	Vague
20		Fast jump to location.	1/5	20.0	Vague
21		Jump to next room.	1/5	20.0	Vague
4		Back to start point of virtual art exhibition.	0/5	0.0	Vague

3.5. Findings from the Pilot Study

The pilot study [1] showed that ‘universal’ icons from applications with which participants were already familiar were easily recognized. Icons that resembled those used in other interfaces and packages, but which had different functions, were confusing to the respondents and did not match their expectations. It was concluded from the pilot study [1] that icons that closely resemble their intended function and therefore do not require prior learning or experience achieve a higher IRR score. The pilot study

also showed that icons taken out of context or which have been encountered previously in another context can be confusing to the user. This appears to depend on the user’s experience, knowledge and familiarity with that type of interface.

Some icons in the interface *appeared* to be common to most applications (e.g. the ‘question mark’ suggests a general help function) but were used in this case for an unusual purpose (i.e. specific navigation help) contrary to the user’s expectations. Therefore, adding more visual detail to the icons to make them more concrete [19] may help users by reducing their ambiguity. However, it may take longer initially for the users to process the icon’s meaning cognitively [16]. In fact, the pilot study suggests that designers’ adaptation of the same icon for different purposes appears to be creating misinterpretation. There are also other factors which may influence icon recognition, including the icons’ grouping in tool bars, their location on the screen, their function, distinctiveness, color and boldness.

3.6. Implications of the Pilot Study

The purpose of a pilot study is to provide pointers and guidelines so that further research can be carried out more effectively. The pilot study found that although most of the icons tested (15/21 or 71.4%) are ‘identifiable’, a significant proportion of them are not functioning effectively (see Table 5). Of the icons tested ‘out of context’ 28.6% (6/21) failed to meet the adopted level of identifiability, which is lower than the ISO standard for signs in general. Of these ‘unsuccessful’ icons, one was classed as ‘mediocre’ (scoring 40% IRR) and 23.8% of the total (5/21) were in the lowest ‘vague’ class, having an IRR of 20% or lower. The meaning of one icon was not recognized by any of the participants (scoring 0% IRR). If these findings are extended to virtual interfaces in general, this lack of recognition could have serious consequences for the effectiveness of icon-driven virtual interfaces in terms of usability, the quality of the users’ experience and their satisfaction. It was therefore decided to explore the possibility of extending the research.

Reflection by the researchers and feedback from reviewers offered the following insights into ways in which the pilot study could be extended:

- The small sample size (five participants) inhibited the data analysis. A larger test sample would improve the statistical validity of the recognition test and make it more representative of the real users of a virtual interface. However, the larger sample could make it more difficult to capture the same ‘richness’ in the data. Nielsen and Sano [23], who devised the tests, justify the use of a sample of five for this reason. In fact, the small sample size means that some values were so marginal that one correct or incorrect interpretation of the icon could increase or decrease the IRR by as much as 20%.
- All the participants were expert computer users, and all had used virtual tour software. This may not be representative of the typical users of a virtual gallery. Similarly, the age range of the participants could be expanded to be more representative of such users. In the pilot study all the participants were in the 20 to 29 age group. A similar study by Gatsou *et al.* [15] that included participants from 20 to 79

suggests that icon recognition declines consistently with age. It would be interesting to test this. The extended research using the same icons should therefore include novice users and older users, which would provide an interesting comparison of the way in which experts and novices and different age groups interpret icon types.

- The test ‘out of context’ was felt to be a fair assessment of the ability of an icon to convey its meaning, but also to be unrealistic as a test of its success ‘in action’. Further tests should therefore be carried out to assess the users’ understanding of the meaning and purpose of the same icons when placed in context, which was felt to be a more realistic evaluation of their function in an interface through environmental clues and positioning. The extended research therefore includes more detailed tests of icons and records more data about the ways in which users understand and interpret icons both in and out of context.
- Little was recorded in the pilot study [1] about the factors which may affect individual participants’ performance in the test. The findings suggest that a user’s personal profile, including factors such as prior knowledge and experience and cognition and learning style, can affect the usability of the interface as well as the degree of ‘immersion’. The extended study therefore includes some of these factors and examines them as influences on icon recognition success.

4. The Extended Study

The testing method used in the pilot study [1] was developed from Icon Intuitiveness Testing by Nielsen and Sano [23]. The study indicated that an IIT is a useful tool for assessing how accurately an icon expresses its intended meaning. However, it was felt that the extended study should provide richer data through which the icons could be evaluated in more depth. Experience of the IIT in practice suggested that improvements could be made. The testing method used therefore draws to some extent on all the other methods explored in the pilot study [1] but is adapted for the extended study. The chosen testing method is therefore termed Icon Recognition Testing (IRT) to avoid confusing it with other testing methods.

4.1. Choice of Subject

It was decided that the extended IRT required as a subject an advanced virtual interface with icons having the following features:

- The icons should be capable of different interpretations in and out of context and be used to carry out a range of functions. Ideally these should include 3D navigation and ‘jumping’ from one location to another, obtaining information about the interface and exhibits and performing action functions such as ‘zooming’ and rotation. They should also initiate sophisticated user-driven interface functions such as screen and window manipulation.
- The icons should be capable of being tested ‘out of context’ and ‘in context’ by using small icon cards and still ‘screen shots’ from the virtual art gallery interface. It was not intended that a fully functional interface should be used, as this may suggest the function of the icons too readily to the participants in the study.

- Some of the icons should be grouped in tool bars as well as being displayed individually and some should only appear when they are usable (i.e. ‘toggled’).
- The icons are used for the basic activities that a visitor would carry out in a ‘real’ art gallery (e.g. navigation around the exhibits and obtaining information about the gallery and artworks) as well as virtual ‘action’ functions (e.g. closing a ‘pop-up’ menu).

After a selection process failed to identify a superior candidate site, it was decided to use the desktop version of the same German 3-D virtual art gallery (i.e. ‘Artweb.com’) that had been used for the pilot study. The website is a more ‘traditional’ type of virtual gallery, using a selection of different styles of room layout based on ‘real’ art gallery architectural plans. It uses an interactive virtual environment, in which users can navigate through a 3-D space using a mouse and keyboard to access an array of icons to carry out tasks using buttons, cursor pointers and interface metaphors.

This website may be less immersive than some that use high-end interactive technology (e.g. VR headsets, helmets and gloves) but it includes a larger selection of icon types and functions [31]. This makes it more useful for an icon recognition test than some of the later generation of virtual tour interfaces that rely on techniques such as ‘swiping’ for some of their navigation actions.

It is important to state that the extended study is not a critical test of this specific site, but a general test of the extent to which certain icons convey meaning and of the usability of this generation of virtual gallery sites of which it is typical. The rationale behind the IRT was to gain an insight into how participants from different backgrounds with varying levels of experience and alternative perspectives would perceive the meaning of the icons. Also, it was intended to see if there is a difference in IRR score between the icons seen ‘out of context’ and ‘in context’. In this study, an icon is taken to be ‘in context’ if two factors apply:

1. There are visual cues in the virtual environment to aid the user in understanding the meaning and/or function of the icons including landmarks, points of reference (e.g. non-interactive objects), contours and boundaries (e.g. walls and doorways), routes around landmarks (e.g. pathways) and room layouts of exhibits [32].
2. Control tool bars are used, with a hierarchical structure, having icons grouped according to their purpose, which change according to the virtual ‘position’ of the user in the interface or the function being requested.

Although the tests identified in the literature review examined similar aspects of icon understandability [15,21,12] as far as the researchers can ascertain no test has examined the same properties of icon design using the same measures of icon recognition. This extended study is therefore an original contribution to the field of icon design as well as to the construction of virtual interfaces. One implicit purpose of the study is to understand how misconceptions arise and to derive recommendations or guidelines for a more effective way of designing icons, allowing virtual interfaces to be developed that enhance ease of use and improve the quality of the user’s experience.

4.2. Test methodology

The complete IRT used in the extended study consists of two recognition tests and two questionnaires, one administered before the tests and one after both tests had taken place, as follows:

- **A pre-test questionnaire**, which contained 13 basic questions to record the participants’ demographic data and level of experience of computing in general and virtual interfaces specifically.
- **Test One (‘out of context’)**, in which participants were shown a range of icons from the interface without any visual cues to their function and were asked to interpret the meaning of each icon. They answered in their own words and their responses were recorded in an Icon Recognition Booklet as brief notes by the Test Administrator.
- **Test Two (‘in context’)**, in which the participants revisited the icons but were shown the context of the art gallery and the environment in which the icon would be seen. As with Test One the responses were recorded in the Icon Recognition Booklet. The responses to Tests One and Two were then analyzed for themes and are reported as Thematic Analysis 1 & 2.
- **A post-test questionnaire**, which contained a series of ‘yes/no’ questions in two sections:

Section 1 (‘out of context’) relating to Test One

Question 1. *Were any of the 21 icons easier to recognize when out of context?*

Question 2. *Were any of the 21 icons harder to recognize when out of context?*

Section 2 (‘in context’) relating to Test Two

Question 1. *Did viewing any of the 21 icons in context (Test Two) change their meaning from Test One (i.e. out of context)?*

Question 2. *Are you familiar with any of the 21 icons in other contexts (i.e. software and applications)?*

Question 3. *Does grouping icons into tool bars make their meaning clearer?*

The verbal responses to both sets of questions were recorded *verbatim* in brief form by the Test Administrator in the Icon Recognition Booklet. The two tests lasted around forty-five minutes to one hour with each participant and the initial tests were completed within a one-week period, followed by a further round of tests with a different sample following comments from a reviewer. Six participants without postgraduate qualifications who were employed in non-computer related work were tested and their results replaced six postgraduate expert participants. The test environment in all cases was a quiet room with adequate lighting, free from distractions. A description of the IRT procedure was read out from a Briefing Instruction Sheet and participants were informed about the test scenario as in previous studies [20], before being asked to complete the consent form and pre-test questionnaire.

4.3. Icon classification

Three categories of icon were identified according to their intended function, such as; initiating action (e.g. zooming in and out, opening and closing a window), obtaining information (e.g.

about an exhibit or the gallery itself) and navigating around the gallery (e.g. moving to the left and right, going forward and back). The set of icons contained some ‘familiar’ icons, which resembled those used in other interfaces, as well as some more ‘obscure’ icons, which would be less familiar to the participants. This combination would test whether experienced users could employ existing conventions to aid their recognition and whether misconceptions could arise because of their existing knowledge and familiarity.

4.4. Pre-test questionnaire – participant demographics

All 21 participants in the tests declared themselves to have good eyesight for computer work and all were competent English speakers, although they had different cultures and nationalities. All were regular users of computers for a variety of purposes. The balance of age, gender, education level and employment (including a category for students) in the opinion of the researchers made the sample representative of the probable range of users of a typical virtual art gallery interface. The responses to the demographic questions are described in the following section:

Questions 1 & 2. *What is your age group? What is your gender?*

The age of the participants was noted, as previous research suggests that the ability to recognize icons declines with age [15] and this was to be tested again. For ethical reasons minors (defined as persons under eighteen) were omitted from the study but apart from that the age range and proportions (from 18 to 69 years) broadly reflect that of visitors to UK galleries in 2016 - 2017 [33]. The gender balance was approximately equal (i.e. 10 males and 11 females) which is also representative of the UK population, as shown in Table 6.

Table 6: Ages and gender of the participant sample

Participant sample by age range and gender			
Age range	No. of users	Male	Female
18 - 25 years	6	1	5
26 - 33 years	6	5	1
34 - 41 years	3	2	1
42 - 49 years	2	2	0
50 - 59 years	2	0	2
60 - 69 years	2	0	2
Totals	21	10	11

Question 3. *What is the highest academic qualification you have obtained?*

The participants were asked to declare their highest level of academic qualification (i.e. school certificate, college diploma, bachelor’s degree, master’s degree or doctoral degree) as it was felt that this may have some bearing on their ability to interpret the meaning of the icons. This is depicted as a ‘pie chart’ in Figure 2 with the proportion of participants’ highest level of academic qualification expressed as a number (in brackets) and a percentage. One participant (4.8% of the sample) had only a school level

qualification, 38.1% had a college Diploma, 28.6% a Bachelors' degree, 23.8% a Masters' degree and 4.8% a Doctoral degree. It is assumed for the purposes of this research that a sample of adults visiting a virtual art gallery will have a similar educational profile.

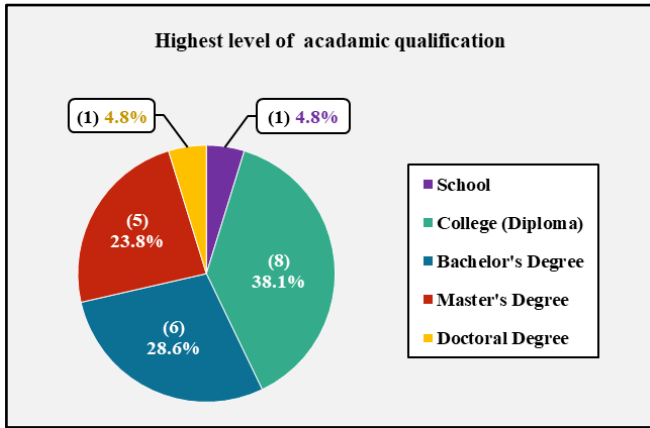


Figure 2: Highest levels of academic qualification

Question 4. *If you are a current or a past student, please state your course title and main area of study.*

The participants' relevant areas of study (e.g. Computing or Art and Design) were recorded briefly in 'free-form' and were placed into seven categories (as shown in Figure 3) to check whether the subjects studied may have some effect on icon recognition. The largest proportion of participants (28.6%) was in the Computing category, with Art and Design the second largest (19.0%), and Sciences constituting the smallest proportion with 4.8%.

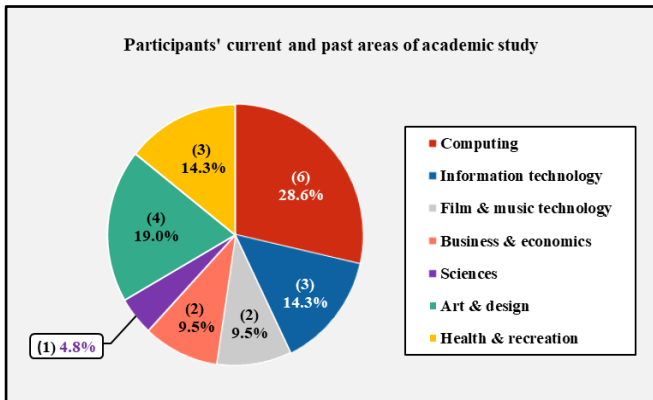


Figure 3: Current and past areas of academic study

Questions 5 & 6. *Which category best describes your occupation? If you are employed, please state your job title.*

Each participant's occupational status (i.e. employed, student, retired or home maker) was recorded with the job category where relevant, to find out if there was a correlation between the participant's employment and his or her ability to interpret icons. It was suspected that certain occupations could develop traits that could affect icon recognition. The primary pie chart on the left of Figure 4 shows the participants' occupational status expressed as a proportion, number and percentage. It is not known whether this employment profile represents the visitors to an actual virtual gallery, but it represents a cross-section of the population.

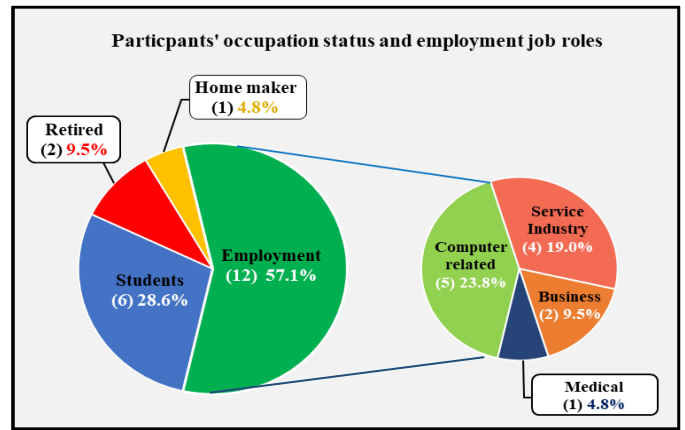


Figure 4: The participants' occupational status

The largest proportion (57.1%) was in employment, while less than a third (28.6%) were students, two people (9.5%) were retired and one person (4.8%) was a home maker. The 'employed' segment was then expanded into a secondary pie chart on the right of Figure 4, which was further divided into job categories, again expressed as a number and percentage. The most common employments were related to the use of computers and the service industries. This implies that less than a quarter of the sample would be regular computer users through their work.

Question 7. *Have you ever worked as an icon designer or a webmaster?*

It was assumed that either of these roles would provide the jobholder with a distinct advantage in terms of icon recognition both 'out of context' and 'in context'. As the sample used in the IRT was intended to be representative of typical virtual gallery visitors, it may be expected that they would have experience as users, rather than as icon creators or designers, so as not to bias the results. It was found that 9.5% of the participants had this type of experience, which was not felt to be excessive. The analysis would show whether experience of icon or website design improved the respondents' ability to recognize the icons.

4.5. Participants' computer experience

Question 8. *Typically, how often do you use a computer interface with icons and for what purpose?*

It was felt that regular use of icon-driven interfaces may have a bearing on the IRR score, so participants were asked to indicate how frequently they used icon-based interfaces and the purpose for which the computer was used, as shown in Figure 5. All the participants used a computer interface daily for Leisure, Home, Work and Study, which constituted the most frequent purpose (i.e. 61.9% of participants). As most packages (e.g. MS Office ®) are icon-driven, this suggests that all the participants would be competent at icon recognition. It should be noted that the operating systems of many commonly-used mobile devices also use an icon-based interface, including Android® and iPhone® mobile 'phones'. The retired participants (9.5% of the sample) characteristically did not use computers at all for Work or Study, but used them for Home and Leisure.

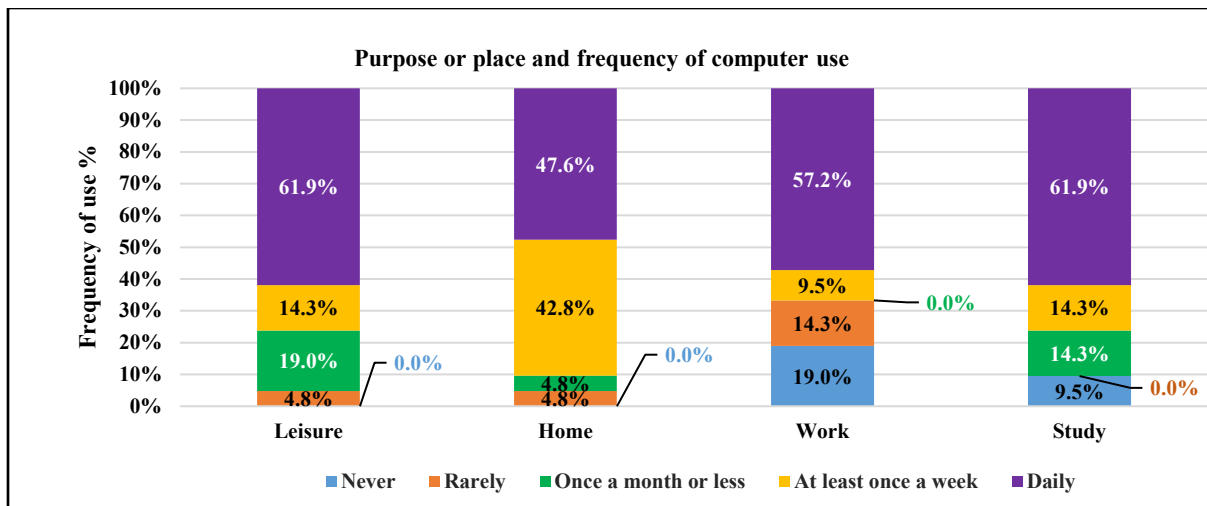


Figure 5: Frequency and purpose of computer use

The responses were given scores of 4 points for daily use, 3 points for use at least once a week, 2 points for at least once a month, 1 point for rarely used and ‘0’ for never used. The point scores for each of the four ‘purposes of use’ categories were accumulated and the totals were ranked in descending order, as shown in Table 7.

Table 7: Users ranked by frequency and purpose of computer use

Users ranked in order by frequency of computer use					
Users	Purpose for which computer used				TOTAL (pts)
	Leisure	Home	Work	Study	
U.3	4	4	4	4	16
U.6	4	4	4	4	16
U.7	4	4	4	4	16
U.14	4	4	4	4	16
U.15	4	4	4	4	16
U.16	4	4	4	4	16
U.21	4	4	4	4	16
U.2	4	4	3	4	15
U.11	4	4	4	3	15
U.5	4	4	4	2	14
U.10	4	2	4	4	14
U.1	2	3	4	4	13
U.19	2	3	4	4	13
U.12	4	3	1	4	12
U.8	3	3	1	4	11
U.20	2	3	3	3	11
U.4	4	3	0	2	9
U.9	3	3	0	3	9
U.17	2	3	1	2	8
U.13	3	3	0	0	6
U.18	1	1	0	0	2
TOTAL	70	70	57	67	264

The highest total score for all categories was the maximum of 16 points (colored green) for Users 3, 6, 7, 14, 15, 16 and 21. The lowest total score was User 18 with 2 points out of a maximum of 16 points (colored red). The median score was 14 and Users 5 and 10 fell into this range, as highlighted by the bold lines. The joint highest frequency of use of computer interfaces was for Leisure and Home (i.e. scoring 70 points) followed closely by Study (scoring 67 points) while Work scored 57 points. The participants overall scored a total of 264 points (78.6%) out of a possible total of 336 points (100%). This indicates that, depending on the types of applications, programs and browsers used, in general the users tested had a significant exposure to a range of icons.

Question 9. How would you describe your level of computer skills?

The participants were asked to rate qualitatively their own level of computer skill (rather than their quantitative experience of using computers) as the user’s general experience with computers may not necessarily equate with his or her competence in using a virtual interface. The self-described level of computer skill showed that all of the participants had some experience of using computers, 42.9% of the sample describing themselves as ‘advanced’ and equal percentages (28.6%) rating themselves as ‘intermediate’ and ‘basic’ as shown in Figure 6. This can be said to represent a typical range of the computer expertise that would be found in visitors to a virtual gallery.

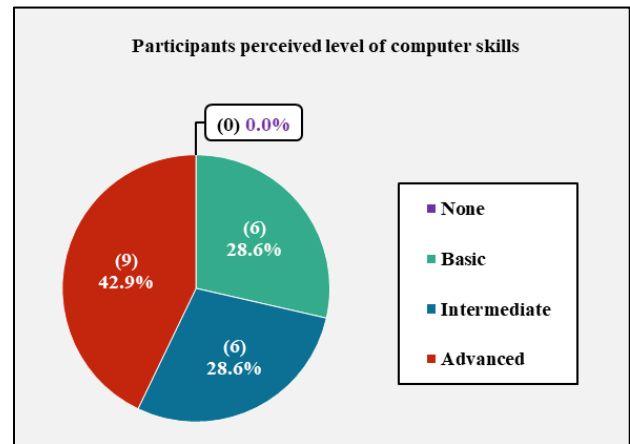


Figure 6: Self-described levels of computer skills

Question 10. Which of the following devices do you use to access the internet?

The participants were asked to indicate which of the ten most common computing devices they used to access the Internet, with the opportunity to record less common devices in free-form as ‘other’. The responses were given scores of ‘1’ for a ticked box or ‘0’ for an unticked box. The scores for the number of devices for each of the 21 users were added and this total score was ranked in descending order as shown in Table 8.

Table 8: Ranked order of participants by the number of hardware devices used

Users ranked in order by number of hardware devices used													
Users	Hardware devices											TOTAL	
	Smart Watch	Smart Phone	PDA	E-reader	Tablet	Notebook	Laptop	Desktop	Games Console	Smart TV	Other		
U.6	0	1	1	1	1	1	1	1	1	1	1	0	9
U.2	0	1	0	0	1	0	1	1	1	1	1	0	6
U.10	0	1	0	0	1	0	1	0	1	1	1	0	5
U.11	0	1	0	0	1	0	1	1	1	0	0	0	5
U.1	0	1	0	0	0	0	1	1	0	1	0	0	4
U.14	0	1	0	1	1	0	1	0	0	0	0	0	4
U.16	0	1	0	0	1	0	1	1	0	0	0	0	4
U.3	0	0	0	0	1	0	1	1	0	0	0	0	3
U.5	0	1	0	0	0	0	1	1	0	0	0	0	3
U.7	0	1	0	0	0	0	1	1	0	0	0	0	3
U.8	0	1	0	0	1	0	1	0	0	0	0	0	3
U.9	0	1	0	0	1	0	1	0	0	0	0	0	3
U.12	0	1	0	0	1	0	1	0	0	0	0	0	3
U.15	0	1	0	0	1	0	1	0	0	0	0	0	3
U.17	0	1	0	0	0	0	1	0	0	1	0	0	3
U.20	0	0	0	0	0	0	1	1	0	1	0	0	3
U.21	0	1	0	0	1	0	0	1	0	0	0	0	3
U.4	0	0	0	1	0	0	1	0	0	0	0	0	2
U.13	0	0	0	0	0	0	1	0	0	1	0	0	2
U.18	0	0	0	0	0	0	1	0	0	0	0	0	1
U.19	0	1	0	0	0	0	0	0	0	0	0	0	1
TOTAL	0	16	1	3	12	1	19	10	4	7	0	0	73

The participant with the highest score was User 6 with 9/10 devices (colored green) and the joint lowest were Users 18 and 19 with 1/10 devices (colored red). The median score was three devices and Users 3, 5, 7, 8, 9, 12, 15, 17, 20 and 21 fell into this range, as highlighted by the bold lines. In terms of the devices, more participants used laptops (19 users) followed by smartphones (16 users) and tablets (12 users). No-one used the smartwatch and older devices such as PDAs also achieved low numbers (one user). It was noted that ‘smartphone’ interfaces tend to be icon-driven which could affect the results of the IRT. The desktop version of the virtual interface was chosen for the test as it was felt by the researchers that this version was most likely to be used for virtual tours of a gallery or museum due to the size and quality of the

monitor. It is unlikely that artwork would be viewed in detail on a smartphone or even a tablet by discerning art lovers.

Question 11: Which of the following types of computer application have you used and how frequently?

The respondents were asked to indicate how frequently they used nine types of computer application, (i.e. regularly, occasionally or never used) to establish their familiarity with different types of interface and their experience of viewing icons in different contexts. The responses were given scores of two points for ‘regular use’, one point for ‘occasional use’ and zero for ‘never used’. The point scores for each of the nine categories were added and this total was ranked in descending order, as in Table 9.

Table 9: Ranked order of participants by the number of computer applications used

Users ranked in order by number of computer applications used										
Users	Computer Applications									TOTAL (pts)
	Interactive websites	Virtual Worlds	Virtual Tours	Social Media	Navigation	Web Browsers	Media Players	Office Apps	Gaming	
U.20	0	1	2	2	2	2	2	2	2	16
U.5	2	1	1	2	1	2	2	2	2	15
U.6	2	2	1	1	1	2	2	2	2	15
U.2	2	0	1	1	2	2	2	2	2	14
U.3	2	1	1	1	2	2	2	2	1	14
U.11	1	1	1	2	1	2	2	2	2	14
U.1	1	0	1	2	2	2	2	2	1	13
U.19	1	1	1	2	1	2	2	2	1	13
U.21	2	0	1	2	1	2	2	2	1	13
U.7	2	0	0	2	1	2	2	2	1	12
U.10	0	0	1	2	2	2	2	2	1	12
U.15	2	0	0	2	2	2	2	2	0	12
U.13	1	0	1	2	1	2	2	0	2	11
U.14	2	0	0	1	1	2	1	2	2	11
U.16	2	0	1	2	1	2	1	2	0	11
U.17	1	0	0	2	2	2	2	1	1	11
U.8	1	0	0	2	1	2	2	1	1	10
U.9	2	0	0	2	2	2	2	0	0	10
U.12	2	0	0	2	1	2	2	1	0	10
U.4	2	0	1	0	0	2	0	0	0	5
U.18	1	0	0	1	0	1	1	0	0	4
TOTAL	31	7	14	35	27	41	37	31	23	246

The highest total score was User 20 with 16 out of a maximum 18 points (colored green) and the lowest score was User 18 with four points out of 18 (colored red). The median score was 12 points and Users 7, 10 and 15 fell into this range, as highlighted by the bold lines. Most users used Web Browsers (scoring 41 points) and Media Player frequently (scoring 37 points) while Virtual Worlds (scoring 7 points) and Virtual Tours (scoring 14 points) were used less frequently. This suggests that the participants would approach the IRT as average users of a virtual interface rather than as experts, which had been identified as a drawback to the pilot study [1]. The researchers noted that the subject interface uses a mixture of icons that would be familiar to the user and ones that had been created specially or adapted that would be unfamiliar.

Question 12: *Have you ever been to a public or private art gallery before?*

All participants except one had visited a real art gallery before and therefore it was felt that a sufficient number would be familiar with the layout and setting within which the ‘in context’ IRT would take place.

Question 13: *Have you ever visited the German ‘Artweb.com’ virtual online art gallery interface before?*

None of the participants had visited the virtual gallery site before and so all undertook the tests on an equal footing in this respect. Participants were given the real name of the gallery.

5. Experimental procedure

As stated in Section 4.2 the IRT consisted of two parts. In the first part, the icons were evaluated ‘out of context’. In other words, they were not associated with the interface and there were no contextual clues to their function or purpose. In the second part of the test, the icons’ context was indicated by using still ‘screen shots’ taken from the virtual tour of the gallery, but the interface was not accessed [34]. This would place an emphasis on understanding the icon in its context and would be a fairer test of the icons’ success in communicating its meaning. The experimental procedures for each test are described below:

5.1. Experimental Procedure – Test One

The test used a variant of the ‘card sorting’ technique [29] using icon cards each measuring 28mm by 28mm, depicting images of the icons. An example of the test set-up is shown in Figure 7.

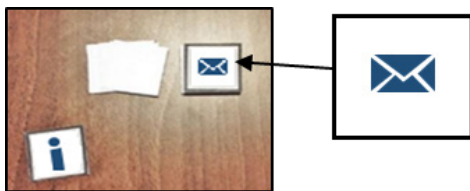


Figure 7: Setup for the paper-based IRT

In carrying out Test One the following principles were observed:

- The icons included no text [23,25] except for Icon 1;
- The icons were displayed without reference to the actual interface (to preserve the lack of context).

- Only one icon was made visible to the user at a time to avoid giving clues to its use.

The test administrator shuffled the pack of cards to ensure that the icons were not grouped in any way (e.g. by spatial association) before placing them face down on the table as a pack [35]. The administrator then picked up one card at a time from the top of the pile and showed this card to each participant at approximately the same viewing angle and ‘reading distance’ as it would be in the virtual interface. Each participant was then prompted verbally to attempt a ‘free-form’ or ‘thinking aloud’ interpretation of the meaning of each icon [34] as specified in ISO 9186 [26] and following the pattern set by Duarte [36]:

Question 1. *What do you think is the meaning of this icon?*

Question 2. *What function do you think would occur if you clicked on this icon?*

Question 3. *Does this icon resemble any sign or symbol you have seen or used before?*

The test administrator noted the responses in the appropriate column of the icon recognition booklet *verbatim*. If a participant was not able to interpret the meaning of the icon within one minute, he or she was encouraged to move on to the next icon card and ‘don’t know’ was recorded. It was felt that if users needed this length of time to interpret the meaning of an icon its use in the interface would be compromised. Participants could provide more than one answer and these were noted for later interpretation. After a response was recorded, the test administrator discarded the icon card onto a separate pile, and the participant was not allowed to revisit any of the icons. This process was repeated for all 21 cards.

5.2. Experimental Procedure – Test Two

In Test Two, the same 21 icons were evaluated again but ‘in context’ (i.e. in their ‘natural surroundings’). The participants were shown ten screenshots from the Artweb.com interface on A4 coloured photographic sheets. These screenshots were still images with no interactive functionality and icons were depicted either individually or grouped in toolbars. The participants were therefore able to use visual clues to derive more understanding and meaning from the icons. No text was included, although Icon 1 contained the English word ‘Tour’. The ten A4-sized screenshots were shuffled to avoid their functionality being revealed by their sequence or by association. The icons to be identified (singly and in groups) were indicated by red rings [34] as shown in Figure 8.

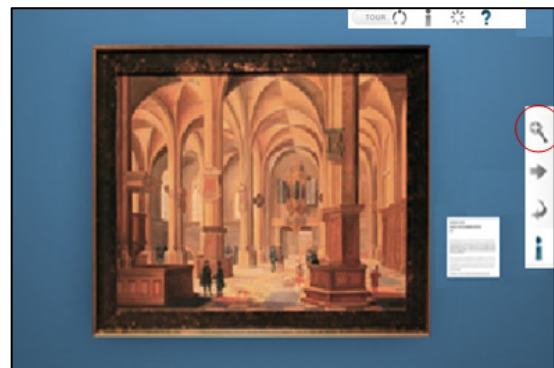


Figure 8: A screenshot of an art exhibit with tool bars

An Icon Reference sheet showing all the numbered icons was available to the participants and the same testing environment was used as for Test One. Each participant was asked what he or she thought the icon meant (as in Test One) and what purpose the icon had. Participants were encouraged to examine the icon's surroundings for additional clues (i.e. from the gallery room or exhibit) and, where relevant, from other icons that were associated when grouped into tool bars. The test administrator noted the participants' responses in the icon recognition booklet. After a response was recorded, the administrator discarded the screenshot onto a separate pile, face down to avoid influencing the next choice. At the end of the test the participants could use the Icon Reference Sheet to help them fill in the open-ended questionnaire.

6. Scoring criteria for Tests One and Two

After the IRT sessions, the researchers assessed the participants' responses according to the following scoring criteria, adapted from a method developed by Rosenbaum and Bugental [37]:

1. **Completely correct** - the participant's response matches both the object and the function, if not the exact description of the icon's meaning (scored as +2);
2. **Partially correct** - the participant's response matches either the object or the function but not both (scored as +1);
3. **Incorrect** - the participant's response matches neither the object nor the function or the answer is completely different from the intended meaning of the icon (scored as 'zero'). The following cases were included in this category:
 - a. Respondent gave 'don't know', 'not sure' or 'no idea';
 - b. No response given;
 - c. Opposite response given to the true meaning of the icon (e.g. in the case of movement or rotation).

If a participant's entry was not completely clear, a discussion was undertaken by the researchers to interpret the response [35]. In extreme cases the participant was consulted about the meaning. An overall results table giving the IRR score for each icon (shown in Appendix B) was produced by using the following formula, where the maximum possible score for each icon is 42.

$$\frac{\text{(Actual Score / Maximum Possible Score)} \times 100}{\text{= Icon Recognition Rate \%}}$$

The IRT results for all 21 icons 'out of context' and 'in context' were separated into classes adapted from a study by Howell and Fuchs [28] with the difference that one class was renamed 'mediocre' instead of 'medium' as it was felt to be a clearer term. The range boundaries differ from those in ISO 3864-2:2016, [38] which refers to general signs rather than computer icons and rates 66.7% and above as 'good'.

According to the Howell and Fuchs stereotypy, icons achieving 60% IRR or above are classed as 'identifiable', whereas icons scoring less than 60% IRR are felt to be 'unsuccessful' in conveying their meaning. The adaptation of this technique that was developed for this research further divides these 'unsuccessful' icons into 'mediocre' (scoring 30% - 59% IRR) and 'vague' (scoring 0% - 29% IRR) as shown in Table 10.

Table 10: Icon Recognition Rate classification

Icon recognition rate (IRR) classification	
IRR (%)	Classification
60 – 100	Identifiable
30 – 59	Mediocre
0 – 29	Vague

7. Results of Test One - 'out of context'

The 21 icons used in Test One 'out of context' were given an IRR score according to the procedure described above and were classed as 'identifiable, (60% - 100%), 'mediocre' (30% - 59%) or 'vague' (0% - 29%) according to the adapted classification system. Where 'identifiable' icons also reached the ISO 3864-2:2016 [38] standard of 66.7% for signs, this was also noted in the results table for interest but was not included in the formal classifications.

The textual comments made by the participants were examined to see if they expressed confidence in their interpretation, for instance by giving several alternative answers or by indicating uncertainty in the hesitant way they provided their responses. This was felt to be important in the 'out of context' test as the participants had no other clues to guide them, so the form of the icons alone had to indicate their meaning.

7.1. Test One – 'identifiable' icon results (60% - 100% IRR)

In total, the 'out of context' test produced eight 'identifiable' icons (i.e. Icons 1, 2, 8, 10, 11, 13, 14 and 19) which is 38.1% of all the icons evaluated in the IRT, as shown in Table 11. The icons are presented in the table in numerical order with the score out of a maximum total of 42 (i.e. 2 points for an icon that is 'completely successful' in conveying its meaning) in the fourth column and the IRR% in the fifth column.

Table 11: 'Identifiable' Icons 60% - 100% IRR










Identifiable icons scoring an IRR 60% -100% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score/ Max. score	IRR %
1		Start Virtual Tour	33/42	<u>78.6%</u>
2		Previous tour position, pause tour, next tour position.	28/42	<u>66.7%</u>
8		Previous artwork to the left	25/42	60.0%
10		Play animation button to circle artwork	28/42	<u>66.7%</u>
11		Pause animation button to circle artwork.	30/42	<u>71.4%</u>
13		Magnifying glass - Pan and zoom image	25/42	60.0%
14		Next artwork to the right	25/42	60.0%
19		Navigation arrow buttons	33/42	<u>78.6%</u>

NB: Icons with an IRR meeting the ISO 3864-2:2016 [38] standard of 66.7% and above are underlined.

7.2. Test One – ‘mediocre’ icon results (30% - 59% IRR)

In total, there were nine ‘mediocre’ icons (i.e. Icons 3, 5, 6, 7, 12, 16, 17, 18, 20) which is 42.9% of all icons evaluated in the IRT. All the results for the ‘mediocre’ icons are listed in Table 12.





Table 12: ‘Mediocre’ Icons 30% - 59% IRR ‘out of context’

Mediocre icons scoring an IRR 30%-59% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score/Max. score	IRR %
3		Exhibition information	22/42	52.4%
5		Help - with navigation of system or interface	17/42	40.5%
6		Full screen view of the virtual exhibition	23/42	54.8%
7		Return screen to window size (smaller view)	21/42	50.0%
12		Slider to zoom in and out of image.	17/42	40.5%
16		Information on artwork or exhibit	24/42	57.1%
17		Email - contact the exhibitor or gallery.	23/42	54.8%
18		Close window button.	18/42	42.9%
20		Fast jump to go to location	16/42	38.1%

7.3. Test One – ‘vague’ icon results (0% to 29% IRR)

In total, there were four ‘vague’ icons (i.e. Icons 4, 9, 15 and 21) which is 19.0% of all the icons evaluated in the IRT. All the results for ‘vague’ icons are listed in Table 13.

Table 13: ‘Vague’ icons 0% - 29% IRR

Vague icons scoring an IRR 0% - 29% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score/Max. score	IRR %
4		Back to start point of virtual art exhibition	0/42	0.0%
9		Rotate left (anti-clockwise)	7/42	16.7%
15		Rotate right (clockwise).	9/42	21.4%
21		Jump to next room.	9/42	21.4%

7.4. Summary of IRT ‘out of context’

The IRT ‘out of context’ showed that eight icons of the 21 icons (i.e. 38.1%) achieved an average IRR above 60%. These icons were therefore classed as ‘identifiable’. Nine icons (42.9%) scored an average IRR% between 30% and 59% and were classed as ‘mediocre’. Four icons (19.0%) failed to reach 30% IRR and were therefore classed as ‘vague’ (see Table 14). That is not to say

that the icons would not function, but it is a strong indication that the user experience would be confusing and less than satisfactory.

Table 14: Summary of icon classes, IRR ranges and results ‘out of context’

Summary of icon classes, IRR ranges and results			
Class of icons	IRR Range	Icon Nos.	No. & % of icons
‘Identifiable’	60% to 100%	1, 2, 8, 10, 11, 13, 14, 19	8/21 = 38.1%
‘Mediocre’	30% to 59%	3, 5, 6, 7, 12, 16, 17, 18, 20	9/21 = 42.9%
‘Vague’	0% to 29%	4, 9, 15, 21	4/21 = 19.0%

8. Results of Test Two – ‘in context’

All 21 icons shown ‘in context’ were given an IRR score in the same way as the ‘out of context’ test and were classified as in Test One. Icons that reached the ISO 3864-2:2016 [38] standard of 66.7% were also noted but not included in the formal classifications. In this case, as with Test One, the verbal responses from the participants were analysed for the degree of confidence they showed in their interpretation of the icons’ meaning, for instance by giving several different answers, by the length of time they pondered while providing a response or by the degree of uncertainty they showed in coming to a decision.

This was felt to be important in the ‘in context’ test as the participants now had clues (e.g. the position of an icon in relation to a landmark or the association of an icon with an exhibit) to guide them. The researchers were interested to see if the inclusion of contextual clues improved the participants’ confidence in their decision-making process. However, confidence in reaching a decision about the meaning of an icon is not necessarily associated with the correctness of the interpretation. It is possible to be confident and incorrect. This could apply particularly to icons that are used in a different context from that with which the participants are familiar. This is discussed in the Textual Analysis in Section 9.

8.1. Test Two – ‘identifiable’ icon results (60% - 100%)

Icons which achieved an IRR score within the 60% - 100% range are classed as ‘identifiable’. In total, there are sixteen ‘identifiable’ icons (i.e. Icons 1, 2, 3, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19 and 20) which is 76.2% of all icons evaluated in the IRT. The ‘out of context’ IRT had already shown that 29.0% of icons were in the this category. Eight icons have therefore improved their IRR score and moved up to the ‘identifiable’ category from the ‘mediocre’. In the ‘in context’ test only one of the ‘identifiable’ icons (i.e. Icon 20) failed to meet the more stringent ISO standard of 66.7% IRR. The results show that twice the number of icons were classed as ‘identifiable’ in context (16) when compared to out of context (8).

This increase in the participants’ ability to recognize the purpose of the icons when the context is known (even in a limited way by showing a screenshot) implies that contextual knowledge makes a significant difference to a users’ understanding of an icon’s meaning and function. All the results for ‘identifiable’ icons are listed in Table 15 and Appendix B.

Table 15: ‘Identifiable’ icons ‘in context’

Identifiable icons scoring an IRR 60% -100% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score / Max. score	IRR %
1		Start Virtual Tour.	35/42	<u>83.3%</u>
2		Previous tour position, pause tour, next tour position.	33/42	<u>78.6%</u>
3		Exhibition information	30/42	<u>71.4%</u>
6		Full screen.	32/42	<u>76.2%</u>
7		Return screen to window size (smaller view).	28/42	<u>66.7%</u>
8		Previous artwork to the left.	35/42	<u>83.3%</u>
10		Play animation button to circle artwork.	31/42	<u>73.8%</u>
11		Pause animation button to circle artwork.	38/42	<u>90.5%</u>
12		Slider to zoom in and out of image.	28/42	<u>66.7%</u>
13		Magnifying glass - Pan and zoom image.	30/42	<u>71.4%</u>
14		Next artwork to the right	36/42	<u>85.7%</u>
16		Information on artwork or exhibit.	33/42	<u>78.6%</u>
17		Email - contact the exhibitor or gallery	29/42	<u>69.0%</u>
18		Close window button.	34/42	<u>81.0%</u>
19		Navigation arrow buttons	38/42	<u>90.5%</u>
20		Fast jump to location.	25/42	<u>60.0%</u>

NB: Icons with an IRR percentage equal to ISO 3864-2:2016 [38] standard of 66.7% and above are underlined.

8.2. Test Two – ‘mediocre’ icons results (30% to 59% IRR)

Icons which scored an IRR percentage within the 30% to 59% range of the IRT when in context are classed as ‘mediocre’ and in addition fall below the acceptable level of the ISO standard. In total there are four ‘mediocre’ icons (i.e. Icons 7, 12, 20, 21) which is 19.0% of all the icons evaluated in the IRT. The results ‘in context’ show a decrease in the number of icons classed as ‘mediocre’, as eight icons (i.e. Icons 3, 6, 7, 12, 16, 17, 18 and 20) have now moved into a higher band (i.e. they have become more identifiable when the context is known). One icon (Icon 21) moved into this class from the ‘vague’ category. None of the icons in this category

became less identifiable when the context was made clear. It may be significant that all the ‘mediocre’ icons appear to have been designed specifically for this interface. Their unfamiliarity therefore gives scope for misidentification and confusion over their meaning and purpose. The IRR scores for ‘mediocre’ icons are listed in Table 16 and Appendix B.

Table 16: ‘Mediocre’ icons ‘in context’

‘Mediocre’ icons scoring an IRR 30% -59% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score / Max. score	IRR %
5		Help - with navigation of system or interface.	21/42	50.0%
21		Jump to next room.	24/42	57.1%

8.3. Test Two – ‘vague’ icon results (0% to 32%)

Icons which participants scored an IRR percentage within the 0% to 29% range when evaluated in context are classed as ‘vague’. In total, there are only three ‘vague’ icons (i.e. Icon 4, 9 and 15) which is 14.3% of all icons evaluated. The results show that Icon 21 that was vague ‘out of context’ moved up to the ‘mediocre’ class ‘in context’. The results for ‘vague’ icons are listed in Table 17 and Appendix B.

Table 17: ‘Vague’ Icons ‘in context’

‘Vague’ icons scoring an IRR 0% -29% range				
Icon			IRR for IRT	
No.	Image	Meaning	Score / Max. score	IRR %
4		Back to start point of virtual art exhibition.	5/42	11.9%
9		Rotate left (anti-clockwise)	4/42	9.5%
15		Rotate right (clockwise).	3/42	7.1%

The IRT ‘in context’ showed that an awareness of context through seeing the screenshots made a significant difference to the users’ ability to recognize the purpose of the icons. In some cases (e.g. Icon 2) this is quite small (a 2.4% increase in IRR) but in most cases, increases in the IRR of between 10% and 20% are achieved. In five cases (Icons 8, 9, 11, 14) the increase in IRR is between 20% and 30% and Icons 18 and 21 both achieved increases of more than 40%. This demonstrates clearly and practically that context plays an important role in icon recognition. The icons within each classification and the proportion of the total icons that they represent ‘in context’ is shown in Table 18.

The eight icons that were moved into a higher classification through evidence of their context are shown in green in the ‘comments’ column of the table. Significantly, the icons in the ‘vague’ category that performed less well in context (Icons 9 and 15) appear to have been designed especially for this virtual interface. Knowing the context in these cases did not seem to help.

Table 18: A summary of icon classes, IRR range, icon numbers and proportions and results ‘in context’ compared to ‘out of context’

Icon classes, IRR range, icon numbers and results ‘in context’					
Class	IRR range	Context	Icon No.	No. & %	Comments
‘Identifiable’	60% – 100%	Out	1, 2, 8, 10, 11, 13, 14, 19	8/21 = 38.1%	8 icons (in green) up from the ‘mediocre’ class
		In	1, 2, 3, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20	16 /21 = 76.2%	
‘Mediocre’	30% - 59%	Out	3, 5, 6, 7, 12, 16, 17, 18, 20	9/21 = 42.9%	1 icon (in green) up from the ‘vague’ class
		In	5, 21	2/21 = 9.5 %	
‘Vague’	0% - 29%	Out	4, 9, 15, 21	4/21 = 19.0%	The ‘vague’ class lost one icon to ‘mediocre’ when in context
		In	4, 9, 15	3/21 = 14.3 %	

9. Analysis of pre-test questionnaire responses

The respondents’ demographic data and their personal profiles (e.g. academic training, experience of interface use, familiarity with computer devices and applications) were recorded in the Pre-test Questionnaire as shown in Section 4.4. An analysis of the data allows interesting comparisons to be made with the results of Tests One and Two. The average of the overall averages (i.e. an average of the ‘out of context and ‘in context’ total IRR scores) is 57.1% as shown in Appendix A.

Questions 1 & 2. What is your age group? What is your gender?

An analysis of the responses to Question 1 shows that the findings of Gatsou *et al.* [15], that the ability to recognize icons declines consistently with age, appears to be confirmed. There was one additional observation, that the youngest age group was fourth out of the six, with an overall average of 54.4% (see Table 19) and performed lower than the average of overall averages (i.e. 57.1%) although other factors may have influenced this result. An analysis of the responses to Question 2 shows that, when grouped according to the overall average IRR score, the male respondents performed slightly better than the females (64.0% IRR for males, 50.9% for females). Eight male respondents and three females are above the 57.1% average of overall averages (see Appendix A). It may be implied from this that the males at least in this sample are better at icon recognition than the females.

Table 19: Overall average IRR performance by age range

Overall average IRR performance by age range			
Age range	No. in group	Total overall IRR	Overall average.
18 - 25 years	6	326.4	54.4%
26 - 33 years	6	387.0	64.5%
34 - 41 years	3	182.2	60.7%
42 - 49 years	2	116.7	58.4%
50 - 59 years	2	106.0	53.0%
60 - 69 years	2	81.0	40.5%

Question 3. What is the highest academic qualification you have obtained?

It may be assumed that the level of education relates to the user’s ability to discern the meaning of icons. An analysis of the responses based on the overall IRR average supports this assumption, but not strongly. Two of the respondents educated to College level scored above the average and six below. Four respondents educated to Bachelors’ level scored above the average and two below. Three respondents with Masters’ degrees scored above the average and two below. The single respondent educated to Doctoral level scored above the average but not significantly. Therefore, it can be inferred that the user’s educational level may have a small influence on icon recognition.

Question 4. If you are a current or a past student, please state your course title and main area of study?

It could be assumed that users with qualifications in technical or ‘visual’ subjects would be better at recognizing icons. Significantly, all five of the top five respondents had qualifications in either Computing, Information Technology or Film and Music Technology, which tends to confirm this. Their skill could be because they had experience of virtual interfaces. The other qualifications were generally distributed among the sample, although it is noticeable that the bottom three scores (well below the average of overall averages) had qualifications in Art and Design and Business and Economics.

Questions 5 & 6. Which category best describes your occupation? If you are employed, please state your job title.

It was suspected that experience of certain occupations could affect icon recognition. An analysis of the data shows that 28.6% of the participants were students and 57.1% were employed in various job categories, with two people being retired and one person a home maker. An analysis of the responses to this question indicated that being a student gave only a slight benefit, as the Employed category averaged 58.2% IRR and the Student category averaged 59.0% IRR. The two Retired respondents averaged 40.5% IRR (below the average of overall average IRR of 54.7%) but were by no means the lowest scorers, being in 18th and 19th place. The Home-maker respondent averaged 66.7% IRR and was in sixth place.

Question 7. *Have you ever worked as an icon designer or a webmaster?*

It was suspected that either role would probably have included experience or training that would increase the ability to recognize the meaning of icons. Two participants declared that they had worked in these roles (i.e. 9.5% of the sample). Both achieved IRR scores above the average of 57.1% and were in the top five places. From this we can conclude that experience as an icon designer may improve icon recognition. This may have implications for aligning the designers' conceptual model and the users' mental models when creating virtual interfaces. This mental model is influenced by the user's profile, including his or her experience, interests, learning style and preferences which the designer needs to know.

Question 8. *Typically, how often do you use a computer interface with icons and for what purpose?*

It was felt that regular use of icon-driven interfaces may have a bearing on the IRR score, and some purposes may also favour icon recognition. An analysis of the data in Table 7 shows that seven of the 21 respondents used a computer 'Frequently' (i.e. four points) for all the purposes of Leisure, Home, Work and Study. The data does not show that frequent and varied use is necessarily associated with accurate icon recognition, as three of the seven scored below the average of both overall averages of 57.1% IRR.

Question 9. *How would you describe your level of computer skills?*

It was felt by the researchers that the more skilled in computer use the participants felt themselves to be, the more confident they would be in interpreting the meaning of the icons. An analysis of the data shows that all the respondent felt themselves to have some degree of self-assessed computer skills. However, the difference between the groups was less than may have been expected. The 'Advanced' group achieved an average IRR score (i.e. between 'in context' and 'out of context') of 58.8%, the 'Intermediate' group 55.0% and the 'Basic' group 56.8% (both the latter being below the average). Surprisingly, the Basic group had a slightly higher IRR score than the Intermediate group. So, self-determined computer skills appear to make little difference to icon recognition, as the score for the Advanced group is only slightly above the average for the whole sample.

Question 10. *Which of the following devices do you use to access the internet?*

This question required the participants to indicate how many and which of the ten most common devices they used. It could be assumed that familiarity with more devices increased the user's experience of different interfaces and types of icons, which could increase the IRR score. In fact, the IRR scores shows no significant correlation between the number of different devices used and the user's ability to recognize icons in the two tests. Indeed, the participant with the highest IRR score (User 21 with an average of 76.2%) used only three devices, and seven of the ten highest scoring participants used less than five (i.e. half the available number of devices). It should be noted that the icons in the test were taken from the desktop version of the virtual interface, which makes it different to the small hand-held devices.

Question 11: *Which of the following types of computer application have you used and how frequently?*

By this question the researchers sought to ascertain if the number of different applications used and the frequency of their use had any effect on icon recognition. Nine different types of application were specified, and points were allocated for each and for the frequency of use. It is possible to see a definite correlation between the variety and frequency of use of computer applications and the IRR score. Eight of the top ten highest scorers in terms of IRR percentage had 12 or more points on the scale (see Table 9).

10. Comparison of icons in and out of context

The results of Tests One and Two were examined and discussed among the researchers. Their interpretations of the findings for each icon are included in the following comparative sections. The comments on the 'out of context' and 'in context' results are followed by a textual analysis of the 'free-form' notes taken from the Icon Recognition Booklet (see Section 4.2).

10.1. Icon 1 - Start virtual tour

Out of context, Icon 1 scored 78.6% IRR in the test (see Appendix B) and was therefore classed as 'identifiable'. There were 14 'completely correct' and five 'partially correct' responses, with two 'incorrect' responses. Interestingly, both of these gave a 'don't know' response, which is rather surprising as its purpose (i.e. the word 'tour') is stated on the icon.

In context, the IRR score for Icon 1 increased to 83.3%, raising it even higher in the 'identifiable' category with 15 participants identifying its meaning correctly. There were five 'partially correct' responses and only one 'incorrect' response, which registered a 'don't know' verdict.

A textual analysis of the 'free-form' responses showed that 'out of context' many of the respondents identified the icon correctly (the word 'Tour' was clearly seen) but did not appreciate its true function as *starting* the tour. A circular arrow on the icon caused confusion, with 'slideshow', 'presentation' and even 'headphones' (which are sometimes used on 'real' gallery tours for audio commentary) being offered as possible functions. 'In context' the respondents were able to assign a more accurate meaning to the icon by seeing it in its 'natural surroundings'.

10.2. Icon 2 - Previous, pause & next on tour

Out of context, Icon 2 scored 66.7% IRR in the test (see Appendix B) and is therefore classed as an 'identifiable' icon. Seven participants were 'completely correct', 14 participants were 'partially correct' and there were no 'incorrect' answers.

In context, the IRR score increased to 78.6%, raising it slightly in the 'identifiable' class, with twelve participants being 'completely correct' and nine participants giving a 'partially correct' estimate and no 'incorrect' responses.

A textual analysis of the 'free-form' responses shows that 'out of context' participants assigned a meaning to the icon based on symbols with which they are already familiar - audio and/or video controls. The use of these symbols goes back to the introduction of the cassette tape recorder in 1963 by Phillips NV. They have since become almost universal, so most of the participants have 'grown

up with them'. Knowing the context in the 'in context' test enabled many respondents to provide more detailed, more informed responses, which increased the IRR score for the icon.

10.3. Icon 3 – Virtual exhibition/gallery interface information

Out of context, Icon 3 scored 52.4% IRR in the test (shown in Appendix B) and was therefore classed as 'mediocre'. Only two candidates were 'completely correct', 18 were 'partially correct' and one 'incorrect'.

In context, the icon's IRR score increased significantly to 71.4%, and it has moved well into the 'identifiable' category with nine 'completely correct' and 12 'partially correct' responses, and no 'incorrect' estimates of its meaning.

A textual analysis of the free-form comments showed that 'out of context' the majority of respondents realised that the use of the letter 'i' was for providing information but were unsure about its exact purpose. In this application the 'i' is for general information about the gallery interface, although this is not clear from the use of a grey colour for the icon. One of the participants thought it was a notification symbol, even though this is normally an 'exclamation mark' in other applications. Placing it into context (i.e. on the main toolbar in the screenshots) no doubt allowed the users to deduce that it referred to information about the gallery, rather than to a specific exhibit. This shows the importance of the proximity of an icon to its function (i.e. its literal rather than semantic distance) or its position in relation to other objects.

10.4. Icon 4 - Back to start point

Out of context, Icon 4 was the least recognized icon of all 'out of context', scoring 0% IRR (shown in Appendix B) and is therefore classed as 'vague'. All 21 participants were 'incorrect' and out of those only one gave a 'don't know' response.

In context, the meaning of the icon was slightly more recognizable, with an IRR of 11.9%, however is still classed as 'vague'. A single participant provided a 'completely correct' response, three gave 'partially correct' responses and 17 participants were 'incorrect', including two 'don't know' responses.

A textual analysis of the free-form responses showed that 'out of context' some participants confused it with icons having a different function in other software packages. Most respondents confused the icon with a MS Vista® loading or buffering button. The shading of the icon (it appears lighter on the bottom) may have given the impression of rotation, which is a feature of loading symbols. Some thought it was for 'brightness' or 'no internet connection' or made wild guesses (e.g. 'sunshine'). 'In context', many responses show the same incorrect assumptions, with 'loading' and 'brightness' responses being frequent.

10.5. Icon 5 – 'Help' with navigation of system or interface

Out of context, Icon 5 scored 40.5% IRR (see Appendix B) and is therefore definitely 'mediocre'. Four participants recorded 'completely correct' responses, nine gave 'partially correct' responses and eight were 'incorrect' without a 'no response'.

In context, the icon IRR rose to 50.0% and so was still in the 'mediocre' category. There were still four 'completely correct' responses while the number of 'partially correct' responses had

increased to 13, with four 'incorrect' interpretations (without any 'don't know' responses but with one 'no response').

A textual analysis of the free-form comments showed that 'out of context' most participants identified the basic meaning of Icon 5 with the universal symbol for 'Help' as they are already familiar with it in other contexts without recognising its specific meaning in this application. 'In context', that there were now only four incorrect responses indicates that knowing the context had helped some respondents to improve their estimate of its meaning.

10.6. Icon 6 – Full screen

Out of context, the icon achieved a 54.8% IRR (see Appendix B) and is therefore classed as 'mediocre'. Eight participants were 'completely correct' and seven were 'partially correct'. Six responses were 'incorrect' (with no 'don't know' responses) and one participant gave the opposite meaning (i.e. shrink screen).

In context, the IRR for this icon rose to 76.2% making it clearly 'identifiable'. Twelve respondents were 'completely correct' and eight were 'partially correct' in their estimates. The number of 'incorrect' responses was one without any 'don't know' answers, indicating confidence on the part of the respondents.

A textual analysis of the responses 'out of context' showed that many participants thought that the icon was something to do with navigation (due to the use of arrows). Several users thought it was a 'click and drag' or movement control button and one user thought that it resembled an icon used in Google Maps® for a different purpose. Conventional 'screen adjustment' controls often use overlapping large and small rectangles as icons, showing that standards set by the designers of the most popular applications create *de facto* paradigms that users recognize. In context, most respondents recognized that the icon had something to do with expansion or enlargement but did not know the full functionality.

10.7. Icon 7 – Return screen to window size

Out of context, Icon 7 achieved a 50.0% IRR (see Appendix B) and is clearly classed as 'mediocre' (i.e. slightly less than its opposite Icon 6). Seven responses were 'completely correct' and seven were 'partially correct'. Seven were 'incorrect' including one 'don't know', one 'no response' and one 'opposite meaning'.

In context, the IRR for this icon rose markedly to 66.7%, making it 'identifiable' according to the adopted scoring system. Eleven responses were now 'completely correct' and six were 'partially correct' with four 'incorrect' judgements without any 'don't know' responses, indicating confidence if not correctness on the part of the respondents.

A textual analysis of the free-form comments showed that when taken 'out of context' one respondent thought the icon referred to a meeting place or a central point in the virtual gallery. They may have been influenced by the similarity of the sign to the familiar 'assembly point' emergency warning sign (see Figure 9). In context, one of the 'incorrect' respondents thought that the icon meant a return to a point on the virtual tour as the arrows were converging and one thought it enabled the visitor to 'enter the picture'. This suggests that familiarity with common physical signs (in this case the 'assembly point' sign) can create confusion

in the user's mind if icons have a similar appearance but are meant to convey a different meaning.



Figure 9: Confusion between 'return to small window' icon and 'emergency assembly point' sign (ISO, 2011)

10.8. Icon 8 - Previous artwork/exhibit to the left

Out of context, the icon scored 60.0% IRR in the test (see Appendix B) and is therefore just classed as 'identifiable'. There were five 'completely correct' and 15 'partially correct' responses. Only one participant identified it incorrectly and in this case the response was the opposite of the intended meaning by recording 'go forward to visit next page'.

In context, the IRR rose to 85.7%, one of the largest increases, due to context making it clearly 'identifiable'. There were now 16 'completely correct' estimates and four 'partially correct'. One participant still assigned an incorrect meaning but there were no 'opposite meanings' (zero scores).

A textual analysis of the written responses to Test One showed that the participant who had assigned an opposite meaning to the icon 'out of context' is from a culture which conventionally reads from right to left. This demonstrates that similar virtual interfaces may be intentionally universal in their application, but the icons that control their use are necessarily cultural in their interpretation. In context, the same respondent gave an incorrect (but not opposite) answer, showing that knowing the context suggested a change of interpretation that overcame the cultural expectations.

10.9. Icon 9 - Rotate left (anti-clockwise)

Out of context, Icon 9 scored 16.7% IRR (see Appendix B) and is therefore classed as 'vague'. There were two 'completely correct' and three 'partially correct' responses but these were completely outweighed by 15 'incorrect' responses, of which three had 'opposite' meanings (i.e. rotate in a clockwise direction). Interestingly, no-one recorded 'don't know', which shows that the respondents were confident but mistaken in their interpretation.

In context, the IRR dropped to 9.5%, making it even more 'vague' and being the joint lowest score in the test. This low score was created by one 'correct' response, two 'partially correct' responses and 18 'incorrect' responses with seven 'opposite' directions being assumed and three 'don't knows' recorded.

A textual analysis of the free-form comments suggests that many participants identified the icon as initiating a rotation but mistook the direction (perhaps the concepts of 'clockwise' and 'anticlockwise' are less relevant today). Others confused the icon with a 'redo' button (although it was flipped horizontally) or a 'refresh' button, which is like one of the paired arrows from other software packages as shown in Figure 10. In context, several participants assigned meanings that are logically incorrect, such as 'skip forward' and 'go to previous (artwork)' showing their uncertainty. It is also significant that the number of 'incorrect', 'opposite' and 'don't know' responses increased so, knowing the context within which the icons would be used clearly confused

some of the respondents. In context, the fact that Icon 9 and its opposite Icon 15 were in toolbars on the opposite sides of the screen to what would be expected (i.e. left-hand rotation on the right tool bar and vice versa) caused the direction of rotation around the artwork to be mistaken.



Figure 10: Confusion between Artweb.com 'rotate left' and common 'redo/refresh' icons

10.10. Icon 10 - Play animation button

Out of context, the icon scored 66.7% IRR in the test (see Appendix B) and is therefore classed as an 'identifiable' icon. Nine participants were 'completely correct' in their interpretation, with ten 'partially correct' and two 'incorrect'. One of the incorrect responses was because the participant left the answer blank.

In context, the IRR score for this icon (which is 'toggled' with Icon 11) increased slightly to 73.8% and stayed in the upper category, classed as 'identifiable' with 14 'completely correct' and three 'partially correct' estimates, but the number of 'incorrect' responses interestingly increased from two to four.

A textual analysis of the accompanying responses suggested that most of the participants could translate inferences from other sign systems and media objects (e.g. audio or video players) to identify the purpose of the icon. As with Icon 2 (which also originated in the cassette players of the 1960s) an analysis of the free-form comments showed that most of the participants were familiar with its use in domestic audio equipment. In context, the scenario shown was a still image of a sculpture that could be rotated. Two participants gave a new but incorrect meaning to the icon when shown screenshots, confusing the icon's action function with navigation ('go to the right' and 'go to next picture').

10.11. Icon 11 - Pause animation button

Out of context, Icon 11 scored 71.4% IRR in the test (see Appendix B) and is therefore clearly classed as an 'identifiable' icon. There were ten 'completely correct' and ten 'partially correct' responses. Perhaps surprisingly, the one 'incorrect' response had the opposite estimate of its meaning - to start.

In context, the IRR score of Icon 11 (intended to be 'toggled' with icon 10) rose appreciably to 90.5% - one of the highest scores in the test. There were 17 'completely correct' responses and four 'partially correct'. Clearly, placing the icon into context has radically changed the participant's understanding of it.

As with Icon 10 and Icon 2, a textual analysis of the comments following Tests One and Two shows that the participants made similar inferences in evaluating the purpose of the icon. This is another icon that owes its existence to the early tape recorder, representing two tape rollers on a 'reel to reel' tape deck. It is commonly used in domestic sound and video equipment to pause a player temporarily until it is restarted by pressing the play button (i.e. Icon 10). This 'universal' icon's features were unique and is

unlikely to be confused with other icons in the test, although the one user who interpreted it with an ‘opposite’ meaning thought it was a ‘restart’ icon. In context, the participants showed a greater understanding of its meaning by associating it with Icon 10 through their familiarity with domestic equipment.

10.12. Icon 12 – ‘Slider’ to zoom in and out of image

Out of context, Icon 12 scored 40.5% IRR in the first test (see Appendix B) and is therefore clearly classed as ‘mediocre’. Seven participants were ‘completely correct’ in their interpretation of its meaning, three were ‘partially correct’ and 11 were ‘incorrect’.

In context, the IRR for the icon increased noticeably to 66.7% with eleven ‘completely correct’, six ‘partially correct’ and four ‘incorrect’ responses, placing it clearly in the ‘identifiable’ class.

An analysis of the free-form responses to Test One suggests that this icon is ambiguous, as it was misinterpreted by nine participants as a ‘volume control sign’ with a slider to change the sound level. Three participants thought it was a ‘battery life or power level’ indicator rather than a ‘zoom slider’ due to its similarity to the icon used for this function on some popular devices (see Figure 11). In gaming, it is often used as an indication of a player’s energy or power level and in mobile devices it can indicate signal strength, making its use in this context confusing.

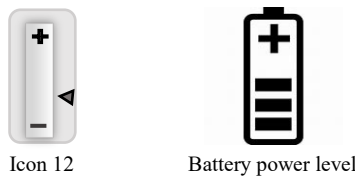


Figure 11: Confusion between Artweb.com ‘zoom control’ and common ‘battery power level’ icons

In context, Icon 12 appears in the right-hand tool bar when viewing a painting with the associated ‘magnifying glass’ (Icon 13). This may have clarified its purpose when seen ‘in context’.

10.13. Icon 13 – ‘Magnifying glass’ to pan and zoom image

Out of context, Icon 13 scored 60.0% IRR (see Appendix B) and is therefore narrowly classed as ‘identifiable’. The ability of the participants to interpret the meaning of this icon is sharply divided. Nine participants were ‘completely correct’ and seven were ‘partially’ correct in their interpretation. On the other hand, there were five ‘incorrect’ estimates, indicating that this a difficult symbol for some users to recognize decisively when out of context.

In context, the IRR score improved markedly to 71.4%, making it solidly ‘identifiable’ and showing that knowledge of its context caused the meaning of the icon to be much clearer to most participants. The increase was produced by 15 ‘completely correct’ and six ‘partially correct’ responses with one ‘incorrect’.

A textual analysis of the free-form entries suggests that this icon is confusing, as the ‘magnifying glass’ part of the sign is ‘solid’ rather than the ring-like or ‘transparent’ device used in other common packages (e.g. Photoshop®). Therefore, ‘out of context’ there were a variety of misconceptions about its meaning. Some of the users saw the icon as a key or a screwdriver symbol, indicating security settings. One saw it as a ‘stop sign’ and three (no doubt

influenced by the common use of a ‘magnifying glass’ in search engines) saw it as a search function. ‘In context’, one Muslim participant misinterpreted it as a ‘Christian cross symbol’ as it was positioned near a painting of the interior of a cathedral. One participant thought it was a ‘search/find symbol’ but instead of looking for information in a search engine it was looking at specific details in the painting.

10.14. Icon 14 - Next artwork to the right

Out of context, Icon 14 scored 60.0% IRR (see Appendix B) and is therefore narrowly classed as ‘identifiable’. There were five ‘completely correct’ responses and 15 ‘partially correct’. Participants appeared to have confidence in their judgement, as there were no ‘don’t know’ responses. The only ‘incorrect’ respondent interpreted the sign as rotation, but in the opposite direction to its intended meaning.

In context, the IRR increased to 85.7% showing that it was clearly ‘identifiable’ by most participants when seen in its surroundings. There were now 16 ‘completely correct’ and four ‘partially correct’ interpretations and the respondent who gave the ‘opposite’ response ‘out of context’ now had the correct direction.

A textual analysis of the responses showed that ‘out of context’ the participant who gave the incorrect ‘opposite’ answer thought the icon’s meaning was to ‘go back’ instead of ‘go forward’, as the person is from a culture which writes from right to left and made the same mistake with Icon 8, thereby emphasising that cultural interpretations should be taken into consideration when designing interfaces. In context, it was shown that some of the participants did not understand the meaning of the icon fully as they stated the direction as ‘go right’ instead of the ‘next artwork on the right’. One participant who gave an ‘incorrect’ answer thought the icon’s meaning was to focus on the right-hand side of the painting itself.

10.15. Icon 15 – Rotate to the right (clockwise)

Out of context, Icon 15 scored 21.4% IRR (see Appendix B) and is therefore clearly classed as ‘vague’. There were four ‘completely correct’ answers and only one ‘partially correct’. An unusually large number of participants (16) gave ‘incorrect’ answers, out of which four gave ‘opposite’ meanings (i.e. rotation in an *anticlockwise* direction). Surprisingly, the respondents showed a high degree of confidence in their understanding of the icon as there were no ‘don’t know’ responses.

In context, this icon had an IRR of 7.1%, showing a notable decline. This was caused by the icon receiving only one ‘completely correct’ response when its context was known. There was again one ‘partially correct’ interpretation, but there were now 19 ‘incorrect’ responses and the number of ‘opposite’ directional interpretations (scored as ‘incorrect’) had increased to seven. Clearly, knowledge of the icon’s context had confused the users!

A textual analysis of the free-form interpretations showed that (as with its opposite Icon 9) many participants identified the purpose of the icon but mistook its direction of rotation ‘out of context’. Also, there was confusion with an ‘undo’ button, a ‘refresh’ button or a ‘return’ button, which use similar symbols (see Figure 12). In context, the position of the icon in a tool bar on the opposite side of the screen to what might have been expected

probably created confusion as to the direction of its rotation (see also Icon 9).

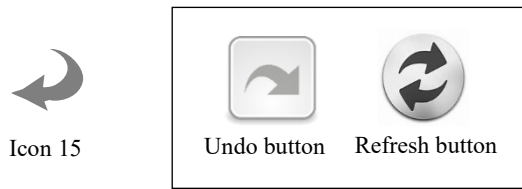


Figure 12: Confusion between Artweb.com 'rotate right' and common 'undo/refresh' icons

10.16. Icon 16 - Information on artwork or exhibit

Out of context, Icon 16 scored 57.1% IRR and was classed as 'mediocre' by a narrow margin. Three participants were 'completely correct' and 18 were 'partially correct'. There were no 'incorrect' responses, indicating only a partial success for this 'universal' icon in communicating its meaning.

In context, like Icon 3 this icon increased its score by a wide margin, achieving 78.6% IRR, making it clearly 'identifiable'. This was largely the result of an increase (to 12) in the number of 'completely correct' responses. There were nine 'partially correct' interpretations and again no 'incorrect' responses.

Textual analysis of the free-form responses showed that 'out of context', all the participants recognized that both Icon 3 (which is grey) and Icon 16 (which is blue) represent 'information' signs but their use tended to elude some of them. This typifies the definition of 'mediocre' icons - the users felt that they knew their meaning but could not work out what to use them for in this interface. 'In context' the IRR score increased considerably, suggesting that association with an artwork helped the participants to identify the purpose of the icon much more accurately. Interface designers should bear this in mind.

10.17. Icon 17 - 'Email' contact the exhibitor or gallery

Out of context, this icon scored 54.8% IRR (see Appendix B) and is therefore classed as 'mediocre', indicating that three participants were 'completely correct' and 17 were 'partially correctly' in their interpretation of its meaning. There was one 'incorrect' response.

In context, Icon 17 increased its IRR score to 69.0%, placing it solidly in the 'identifiable' category. The number of 'completely correct' answers increased to 11 and the number of 'partially correct' responses was now seven. Interestingly, the number of 'incorrect' responses increased to three as context apparently introduced new ambiguity.

Textual analysis of the free-form answers showed that 'in context' most participants identified the basic meaning of Icon 17 with the universal symbol for email. It also shows that some participants could not work out whether the icon was to open an email reader to send or receive an email message and were therefore unable to decide to whom the email was to be sent and about what. This appears to be a case of using a common icon for an unusual purpose, which shows that the design of an icon needs to be aligned with the user's experience and familiarity with similar signs. In context, the email's precise purpose of contacting the exhibitor or gallery about the exhibit became more apparent by its closeness to the exhibit and its association with other icons in

the same tool bar that are used to directly to manipulate the exhibit. Although, with experience of its use in social media applications, some users thought it was a way of posting comments.

10.18. Icon 18 - Close window button

Out of context, Icon 18 scored 42.9% IRR (see Appendix B) and is clearly classed as 'mediocre'. Seven participants identified the meaning of the icon completely correctly and four partially correctly. There were ten 'incorrect' responses from participants who attempted to guess the icon's meaning.

In context, the icon's IRR score almost doubled to 81.0%, moving it well up into the 'identifiable' class. This can be attributed to the increase in 'completely correct' responses to 15, while the 'partially correct' responses remained the same at four and two participants registered an 'incorrect' response, one assuming that it marked an observation point and the other a warning.

This is a clear indication that knowledge of context has enabled many participants to improve their understanding of the meaning of the icon. Textual analysis showed that initially 'out of context' there was a wider interpretation by participants with a number of different meanings for a type of warning sign such as a 'no entry', 'stop sign', 'error sign', 'cancel sign' or 'gallery closed' sign. This represented a mismatch with the participants' expectations based on their experience of the symbol in other applications. In fact, the basic form of the icon (although not necessarily its colour) is commonly used to close pop-up windows in a variety of applications, including MS Word® (see Figure 13). An analysis of the free-form responses showed that 'in context' the icon's position on the corner of a window 'frame' made its purpose clearer to the participants, as this is where they would normally expect to see a 'close window' button. This demonstrates the value of consistency, not just in the appearance of icons but in their position and their association with other parts of the interface.



Figure 13: Similarity between 'close pop-up window' icons (left; 'Artweb.com', right; MS Word®)

10.19. Icon 19 - Navigation arrow button

Out of context, Icon 19 scored one of the highest results with a 78.6% IRR (see Appendix B) and is therefore clearly classed as 'identifiable'. In all, 15 participants interpreted the meaning of the icon completely correctly and three were 'partially correct', while three were 'incorrect'.

In context, the IRR score increased to 90.5%, with 18 participants giving a 'completely correct' response, two 'partially correct' and one 'incorrect', making the icon one of the most identifiable.

Textual analysis of the free-form responses shows that 'in context' some of the participants felt the sign to be like one used in Google Maps® and they therefore interpreted it as a map controller (i.e. for moving a map around a window) rather than a direction control icon (i.e. moving the user's viewpoint). One participant confused this icon with a similar icon often used to enlarge an image or screen. This icon is also familiar to participants

who have experience in playing games which use this type of 3-D navigation tool to move around screens. The context, however, ruled out any association with maps and its position on the screen indicated that it was a navigation tool to most participants.

10.20. Icon 20 - Fast jump to location

Out of context, Icon 20 performed relatively poorly, being placed in the 'mediocre' category with an IRR of 38.1% (see Appendix B). Two participants gave 'completely correct' responses and 12 were 'partially correct' with their estimates. There were seven 'incorrect' answers and out of these, one participant gave a 'don't know' response rather than guessing.

In context, the icon moved up to the 'identifiable' category, although it only just met the criteria with an IRR score of 60%. This was largely because of an increase to eight in the number of 'completely correct' responses. There were nine 'partially correct' estimates and the 'incorrect' responses numbered four. Out of these, the number of 'don't know' answers increased to two and one of these participants had made a wild guess previously 'out of context', which they then discounted.

A textual analysis of the free-form responses showed that many participants mistook this icon 'out of context' for a 'map pin' marking a specific point rather than a navigation aid, based on its similarity to the marker used in Google Maps® and similar applications with which they were familiar. One interpretation given was as a sign for the start of a tutorial. In context, two participants thought it was a marker for the current location and did not know it was a navigation aid to fast jump to another location in the art gallery. Some of the participants stayed with their originally answers 'out of context', although their responses were more descriptive and related to the context. As the meaning is not clear in or out of context, it seems that the icon requires the user to gain experience with the interface, to learn its functionality.

10.21. Icon 21- Jump to next room

Out of context, Icon 21 scored 21.4% IRR (see Appendix B) and is in the 'vague' category. Only one participant gave a 'completely correct' answer, whilst seven gave 'partially correct' estimates. There were 13 'incorrect' responses and out of those two were 'not sure' or had 'no idea', two gave no response and the rest gave a different meaning or a wild guess as to its purpose (e.g. an architectural feature).

In context, the icon's IRR score rose significantly to 57.1%, moving it up a category to the upper end of the 'mediocre' class. There were now eight 'completely correct' and eight 'partially correct' responses, while five respondents gave 'incorrect' estimates of the icon's meaning.

A textual analysis of the free-form answers showed that some participants mistook the icon for a military insignia 'out of context', as a similar icon appears in many computer games that they had played previously. In context, some participants took it for a sign pointing up to the next floor of the art gallery (e.g. a sign for a lift or elevator), rather than for 'jumping' into the next room on the same level. One participant thought it was an end-point in the gallery visit which could be saved, to allow them to return to the same point. This icon probably requires some prior familiarity through learning the interface to know its functionality.

11. Thematic Analysis 1 - findings 'out of context'.

Questions 1 and 2. 'Are any of the icons a) easier, b) harder to recognize out of context?'

The responses to this question suggest that knowledge of context does increase the IRR score but perhaps not as much as previous work [39] would suggest. In some cases, knowing the context made identification more problematic. Icons 9 was felt to be harder to recognize 'in context' by six respondents (28.6% of the sample) while Icon 15 was felt to be more difficult by five respondents (23.8%). These results show that context cannot be relied on to make an icon more understandable, as context can be misleading. With Icons 4, 6, 7, 9, 12, 13, 17 and 18 some respondents who thought that context made identification easier were in fact incorrect in their interpretation. However, knowing the context did enable more accurate recognition in many cases, as was expected. Icons 3, 6, 7, 12, 16, 17, 18 and 20 were felt to be easier to identify through familiarity, and the fact that they moved from the 'mediocre' to the 'identifiable' class bears this out (See Table 19 and Appendix B).

Thematic Analysis 2 - findings 'in context'.

Question 1. 'Do any of the 21 icons change their meaning from what you expected 'in context?'

The textual analysis showed that being seen 'in context' changed the meaning of all the icons except Icon 8. However, the difference between the 'out of context' and 'in context' tests is lower than expected, in two measures. The first measure is the increase or decrease in IRR between the two tests (as shown in Appendix A). The second is the number of respondents seeing a change in the meaning of the icons when taken 'in context'. Overall, the change is minor and the icon with the largest percentage of respondents is Icon 12 with 47.6% (10/21) believing that context changed the icon's meaning. Only two icons were regarded as having changed their meaning by more than a third of the participants (i.e. Icons 12 and 21). As may be expected, Icon 21 had the largest increase in IRR (+40%).

Question 2. 'Are you familiar with any of the icons in other contexts?'

Icon 1 was distinctive and was not confused with other icons. Only one respondent reported seeing something similar on another (un-named) virtual interface. Icon 2 was felt to be like a video, music or media player control by 13 respondents. Interestingly, three thought it was used on YouTube®, but they are not correct. Icon 3 was related to an information function by eight respondents, but none suspected its secondary meaning, which is to give general information about the site. One respondent thought that it was 'greyed out' because it was not active. As the icon cards were displayed in random order, some respondents may have already seen the 'blue' information icon, though none mentioned it. Icon 4 was perceived as a loading/buffering symbol or a brightness control by twelve respondents and as 'settings' (often represented by a 'gear' icon) by four. No-one suspected that it was intended to return the user to the start of the tour. Icon 5 was seen to be a common 'help' icon with confidence by five respondents, and one felt that it served the same function as a in Microsoft® applications. Two respondents offered the extra information that it offered help about the tour.

Icon 6 was felt to be a 'full screen' icon by five respondents and two offered the secondary information that it was like a YouTube® control (not correct). The same respondents felt that Icon 7 was a 'shrink screen' control and again mistakenly attributed it to YouTube®. Icon 8 was likened to the 'go back' icon on websites, etc. by eight respondents. Four identified it correctly as referring to the previous artwork. Icon 9 was felt to be a 'redo' button by four respondents and a rotation control by three more, although the direction confused them. Icon 10 was related with confidence to a media 'play' control by 10 respondents and YouTube® was cited correctly by three, showing the influence of popular social media software.

Icon 11 was identified as a 'pause' control by nine respondents and YouTube® was correctly cited in one case. Icon 12 caused confusion, with respondents seeing it as like icons as diverse as a Wii® controller and an icon from photograph viewing 'apps'. Icon 13 was related to 'map apps' by two respondents and as a 'zoom' control by two more. Icon 14 is the opposite of Icon 8, and five respondents gave it the opposite meaning. Icon 15 (the opposite of Icon 9) was perceived correctly by one respondent as a rotation symbol. Icon 16 was understood as a common 'information' icon by eight respondents, but three of them were confused about whether it was general or specific information.

Icon 17 was viewed with confidence as an 'email' button by thirteen respondents, although no-one deduced its secondary meaning. Icon 18 also caused confusion, some respondents seeing it as a 'no entry' sign. Icon 19 was also seen to resemble icons used in several different common applications. Only two respondents saw it correctly as a navigation control who felt that it was like a control from Google Street View® or an X-box® icon. Icon 20 was recognized as a map pointer as used in Google Maps® by ten respondents with confidence, but most failed to identify it as a 'jump' device. Icon 21 created the most confusion, as five respondents perceived similarities to other interfaces (e.g. Google Street View®) and one felt that it was like the 'collapse' control on a pull-down menu.

Question 3. *Does grouping icons in tool bars make their meaning clearer?*

Seven participants (33.3% of the sample) felt that grouping the icons into tool bars had *not* made their meaning clearer. A typical response was, 'The tool bars do not make any difference...you look (locate) and use the *icon* you require, not the tool bar'. For those respondents who did perceive a positive difference, four significant themes emerge from the analysis:

- **Position** - where a toolbar is placed on the screen (e.g. right or left) suggests navigation in either direction, while the center or top of the screen suggests a more general use;
- **Difference** - icons need to be clearly distinguishable from other icons in the same set;
- **Proximity** - the closeness of a toolbar to an item on the screen (e.g. a painting or a doorway) suggests the meaning of the icon and its intended purpose;
- **Consistency** - the icons in a tool bar should perform functions regularly so that they are learned and understood more easily (e.g. the main tool bar is used more frequently and consistently than the left or right tool bars);

- **Association** - links between groups of icons in a tool bar suggest their use and meaning, which may be transferred over from other applications using similar icons.

The Thematic Analysis shows that an understanding of the meaning of the icons in a tool bar relates strongly to their position in relation to other items on the screen. For the Main Tool Bar ten out of 15 respondents felt correctly that its position (at the top center of every screen) suggested that the icons had a general purpose. For the Left Tool Bar seven out of 12 respondents and the also seven out of 15 respondents for the Right Tool Bar were correct in feeling that the position of the respective tool bars indicated that they operated on the object being viewed (e.g. zoom in or out, rotate left or right, etc.). The other themes were less strongly indicated but with the Right Tool Bar, association (e.g. with other icons in the tool bar) indicated a specific application in five out of 14 responses. This tool bar contained both navigation and information icons and some respondents could not distinguish between them. This appears to reinforce research [40], which suggests that icons need to maintain 'difference'. An icon needs to be clearly distinguished from other icons in the same tool bar and be close semantically to its own function while maintaining as great a semantic distance as possible from the other icons.

For the Main Tool Bar 'consistency' and 'association' were both cited as indications of meaning in two out of 15 responses, whereas 'proximity' and 'association' (e.g. with other icons in the tool bar) were cited in only two out of 12 responses in relation to the Left Tool Bar and in four out of 14 responses to the Right Tool Bar. One respondent offered the comment, 'I think it's a good idea to make the right-hand tool bar look like the left one (only [including] navigation icons) and move the other icons (information ones) in the right tool bar to the top [Main Tool Bar]'. An example of 'association' occurs in the response of one user, who associated Icon 16 'Information' with Icon 17 'Send email', which appear together on the right tool bar, assuming the 'i' symbol led to an address book while the 'envelope' referred to sending the email. Another suggestion was to remove Icon 6 and Icon 7 that vary the image size from the Right Tool Bar and place them on the painting itself, like Icon 18 that closes a window.

12. Discussion of findings from Test One and Test Two

From the results of this study it is possible to make certain observations. Icons that resemble their intended function more closely (i.e. have a close semantic distance) tended to have a higher IRR score both 'out of context' and 'in context'. It can be concluded that this is because less prior learning or familiarity is needed for users to understand their meaning. As computer icons are not 'standardized' as are warning signs through the ISO [26-27] icon designers' adaptation of the same or similar icons for different purposes can create misinterpretation.

Theory suggests that when planning an interface, icon designers have a conceptual model of the way in which the icons will be used [41] based on their training and experience. The users of the interface, on the other hand, will have a mental model of the icons' meaning based on their knowledge, cultural background and familiarity [41]. The importance of matching these models is demonstrated by the confusion caused in the tests by 'familiar' icons whose functions differed from users' expectations. The IRT 'out of context' (Test One) showed that 33.3% of the icons were

clearly identifiable to users (see Table 9). Icon 4 was confused with a 'gear cog' for adjusting system settings. Icons 9 and 15 were too similar to icons used differently in other applications. Icon 21 had been encountered with touch screens for 'swiping', but not for navigating between displays, having an adverse effect on usability through a "lack of conformity with user expectations" [42].

13. Conclusions

The original pilot study suggested that there was a problem with icon recognition, even to expert, qualified computer users [1]. This prompted that further research needed to be done into the phenomenon. The research for this paper is therefore an extended study based on an expanded sample of 21 computer users with different levels of competence, ages, educational attainments and spheres of employment. It is felt that this sample represents typical users of a virtual art gallery.

The extended research project set out to evaluate a set of randomly-chosen icons that carry out the action, information and navigation functions in a virtual gallery interface. A combination of quantitative and qualitative techniques was employed to add depth to the data analysis, while avoiding the use of complex statistics. The study is therefore based on an established method of Icon Recognition Testing (IRT) examining 21 icons from a 'real world' virtual gallery. However, this study is original, as it combines tests in and out of context and draws a comparison between them. An additional innovation is the combination of qualitative textual and thematic analysis (Sections 10 and 11) to establish reasons for the users' interpretation of the icons' meaning. This adds considerably to the contribution of the research. The findings are useful to interface designers and academics alike, by offering advice and by prompting further research. Conclusions for virtual interface design are drawn from the results of the research under the following headings:

13.1. Familiarity helps – people get to know icons

The IT industry is a long way from adopting standards equivalent to those for warning and traffic signs. However, the study shows that familiarity can aid users in recognizing their meaning in different contexts. The consistent use of familiar icons for their expected function is therefore important. When an interface is used regularly (e.g. a word processing package) users gain familiarity with the icons' function, even by its position, without having to decode its meaning. However, individuals tend to visit a virtual gallery relatively infrequently and are less likely to gain familiarity with the icons. The low IRR scores of icons that are 'custom made' for the 'Artweb.com' interface (e.g. Icons 4, 9 and 15) would appear to support this conclusion.

13.2. Abstraction is useful – but should be controlled

This paper begins by discussing concreteness and abstraction in icon design. The research shows that augmenting an icon with text (e.g. Icon 1) assisted the users in understanding its meaning. Therefore, adding more visual detail to the icons (i.e. making them more concrete) may reduce ambiguity. However, it may initially take longer for users to process mentally [16] and could interfere with their enjoyment and detract from the virtual experience. The balance between abstraction and concreteness should be an important consideration for interface designers.

13.3. Icons should be 'audited' regularly

The extended study prompts the recommendation that icon recognition testing should be carried out regularly as a part of an interface design 'audit' to ensure that the icons are continuing to fulfil their intended purpose. The study suggests that after such an audit, icons classed as 'identifiable' should be maintained in their present form. It is further suggested that icons classed as 'mediocre' could be modified economically to be more effective by making them more concrete or sufficiently different from icons used for other functions. However, icons classed as 'vague' should be redesigned completely or replaced, taking into account the recommendations offered in Sections 13.1, 13.2 and 13.3. It is suggested that some icons in this category may be replaced by familiar icons from other software packages that have passed ISO benchmark tests (e.g. MS Word®), subject to legal approval.

13.4. Interface designers need to understand user profiles

The results demonstrate that when designing icons for a virtual interface (in this case a virtual art gallery) it is important that the designer's conceptual model closely matches the users' mental model. Norman [41] explains that the interface designer does not communicate directly with the user, but through the 'system image', which is developed from the designers' own conceptual views and understanding of the nature and purpose of the interface. The users subsequently form a mental model based on their own understanding and interpretation of the system image, influenced by their beliefs, experience and prior knowledge (i.e. their user profile). A match between the conceptual model, the system image and the user profile should result in an enhanced user experience, so virtual interface designers should capture user profiles to adapt the interface to the user's requirements.

14. Limitations of the research

Importantly, this study has its limitations. The IRT focussed on evaluating icons with different functions taken from the same interface (i.e. Artweb.com). The study by Ferreira *et al.* [20] compared icons from different interfaces with the same function. In both Ferreira's and this research, the tests were limited to identifying the icons' meaning using paper-based tests. A more sophisticated and comprehensive icon recognition test could be done with technology that would record more information about the users' intuitive responses and 'thinking time' (e.g. interactive MS PowerPoint with key logging). The tests could be extended so that the participants could compare different virtual interfaces.

15. Suggestions for future research

Many of the virtual gallery interfaces identified in the secondary research currently offer a 'one size fits all' approach to icon design. It is suggested that more needs to be known about the potential for user profiling in virtual interface design, perhaps using an ontology engineering approach. Methods of capturing this profile need to be non-invasive if the user experience is not to be compromised. There is the potential for exploring methods such as gamification as a way of capturing users' profiles and preferences. A variety of frameworks exist to enable designers to do this [43].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

We thank the participants who evaluated the icons and provided detailed interpretations. We also acknowledge the expertise of the designers of the German virtual art gallery website for providing a rich and useful source of enjoyment to art lovers and material for researchers into human-computer interaction.

References

- [1] D.E. Ashe, W.A. Eardley and B.D. Fletcher, "e-Tourism and Culture through Virtual Art Galleries: a Pilot Study of the Usability of an Interface" in 4th International Conference on Information Management (ICIM), Oxford, U.K. 2018. <http://dx.doi.org/10.1109/INFOMAN.2018.8392834>.
- [2] Smithsonian Natural History Museum virtual tour. 2018. <http://naturalhistory.si.edu/panoramas/>.
- [3] Louvre Online Tours. 2018. <http://www.louvre.fr/en/visites-en-ligne>.
- [4] Virtual Tour of the Oxford University Museum of Natural History. 2018. <http://www.chem.ox.ac.uk/oxfordtour/universitymuseum>.
- [5] Virtualfreesites. Virtual Tours of Museums: Exhibits and Points of Special Interest: Virtual Tours. 2018. <http://www.virtualfreesites.com/museums.html>.
- [6] W.A. Eardley, D.E. Ashe and B.D. Fletcher, "An Ontology Engineering Approach to User Profiling for Virtual Tours of Museums and Galleries" *Int. J. Knowl. Eng.*, **2**(2), 85-91, 2016. <http://dx.doi.org/10.18178/ijke.2016.2.2.058>.
- [7] D.J. Mayhew, *Principles and Guidelines in Software User Interface Design*. New Jersey: Prentice Hall, Inc. 1992
- [8] P.B. Anderson, *A semiotic approach to programming*. In: *The computer as medium*. Cambridge, England: Cambridge University Press, 16-67. 1994
- [9] J. Ferreira, P. Barr & J. Noble, "The semiotics of user interface redesign" in *Proceedings of the 6th Australian User Interface Conference (AUIC '05)*, **40** (4) Sidney, Australia, 47-53. 2005.
- [10] R. Yan, "Icon Design Study in Computer Interface." *Pro. Eng.*, **15**, 3134-3138, 2011, <http://dx.doi.org/10.1016/j.proeng.2011.08.588>.
- [11] H.I. Cheng and P.E. Patterson, "Iconic hyperlinks on e-commerce websites" *J. Appl. Ergon.*, **38** (1), 65-69, 2007, <https://doi.org/10.1016/j.apergo.2006.01.007>.
- [12] B. Merdenyan, O. Kocycigit, R. Bidar, O. Cikrikcili and Y.B. Salman, "Icon and User Interface Design for Mobile Banking Applications" in *Proceedings of the 4th International Conference on Advances in Computing and Information Technology (ACITY '14)*, **4**(2), 55-59, 2014, https://dx.doi.org/10.3850/978-981-07-8859-9_18.
- [13] G. Bhutar, R. Poovaiah, D. Katre and S. Karmarkar, "Semiotic Analysis combined with Usability and Ergonomic Testing for Evaluation of Icons in Medical User Interface" in *Proceedings of the 3rd International Conference on Human Computer Interaction (IndiaHCI '11)*, Bangalore, India, 57-67. 2011. <http://dl.acm.org/citation.cfm?id=2407804>.
- [14] P. Barr, J. Noble, and R. Biddle. "Icons R Icons" in *Proceedings of the Fourth Australasian User Interface Conference on User Interfaces*, 18 (AUIC '03), Darlinghurst, Australia: Australian Computer Society, Inc., 25-32. 2003, <https://dl.acm.org/citation.cfm?id=820093>
- [15] C. Gatsou, A. Politis and D. Zevgolis. *The Importance of Mobile Interface Icons on User Interaction*. [Online] *Int. J. Comput. Sci. Appl.*, **9** (3), 92-107. 2012. <http://www.tmrfindia.org/ijcsa/v9i37.pdf>.
- [16] L. Nadin. *Interface design and evaluation - Semiotic implications*. In: H.R. Hartson and D. Hix (Eds) *Advances in human-computer interaction*, **2**. Norwood, NJ, USA: Ablex Publishing Corp. 1988.
- [17] R. Scalisi. "A semiotic communication model for interface design" in *Proceedings of the 1st International Conference on Computational Semiotics for Games and New Media (COSIGN2001)*. Netherlands, Amsterdam. 10-12 Sept. 2001. <http://www.cosignconference.org/conference/2001/>
- [18] S. Isherwood. "Graphics and Semantics: The Relationship between What Is Seen and What Is Meant in Icon Design" in *Proceedings of the 8th International Conference on Engineering Psychology and Cognitive Ergonomics*, Held as part of HCI International, San Diego, California, USA. July 2009. https://doi.org/10.1007/978-3-642-02728-4_21.
- [19] R. Arnheim, *Visual thinking*. London: Faber. 1969.
- [20] J. Ferreira, J. Noble and R. Biddle, "A case for iconic icons" in *Proceedings of the 7th Australasian User Interface Conference (AUIC '06)*, Hobart, Tasmania, Australia: Australian Computer Society, Inc., **50**, 97-100. Jan. 2006. <https://dl.acm.org/citation.cfm?id=1151771>
- [21] Y.B. Salman, H-I. Cheng and P.E Patterson. "Icon and user interface design for emergency medical information systems: A case study" *Int. J. Med. Inf.*, **81**(1), 29-35, 2012. <http://doi.org/10.1016/j.ijmedinf.2011.08.005>.
- [22] S. Schröder and M. Ziefe, "Making a completely icon-based menu in mobile devices to become true: a user-centred design approach for its development" in *Proceedings of the 10th International Conference on human computer interaction with mobile devices and services*, Amsterdam, The Netherlands: ACM, 137-146, Sep. 2008. <http://doi.org/10.1145/1409240.1409256>.
- [23] J. Nielsen and D. Sano, "SunWeb: User interface Design for Sun Microsystem's Internal Web". *Comp. Nets. ISDN Sys.*, **28**(1), 179-188., 1995. [https://doi.org/10.1016/0169-7552\(95\)00109-7](https://doi.org/10.1016/0169-7552(95)00109-7).
- [24] J.J. Foster. *Graphical Symbol: Test methods for judged comprehensibility and for comprehension*. ISO Bulletin, 11-13, December 2001. <http://hablamosjuntos.org/signage/PDF/graphicsymbols0112.pdf>.
- [25] S. Ghayas, S. Sulaiman, M. Khan and J. Jaafar. *The effects of icon characteristics on users' perception*. In *Advances in Visual Informatics*. Switzerland. Springer International Publishing, 652-663. 2013.
- [26] ISO. ISO 9186:2011. *Graphical symbols - Test methods for judged comprehensibility and for comprehension*. International Organization for Standardization. Geneva, Switzerland, 2001.
- [27] ISO. ISO 3864:1984. *Safety colours and safety signs*. International Organization for Standardization. Geneva, Switzerland, 1984
- [28] W.C. Howell and A. H. Fuchs, "Population stereotypy in code design", *J. Org. Behav. Hum. Perf.*, **3**(3), 310-339, 1968, [https://doi.org/10.1016/0030-5073\(68\)90012-3](https://doi.org/10.1016/0030-5073(68)90012-3).
- [29] E. Kaasinen, J. Luoma, M. Penttinen, T. Petäkoski-Hult and R. Södergård. *Key Usability and Ethical Issues in the NAVI programme (KEN)*. 2001. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.4971&rep=rep1&type=pdf>.
- [30] B. Martin and F. Ringham, *Key terms in semiotics*. New York: Continuum International Publishing Group Ltd., 2005.
- [31] P.R. Desai, P.N., Desai., K.D. Ajmera and K. Mehta, K., "A review paper on oculus rift - A virtual reality headset". *Int. J. of Eng. Trend. & Tech. (IJETT)*, **13** (4), 175-179, 2014. <https://doi.org/10.14445/22315381/IJETT-V13P237>.
- [32] D. Biella, D., Sacher, B. Weyers, W. Luther and N. Baloian, "Metaphorical Design of Virtual Museums and Laboratories: First Insights" in *International Conference on Ubiquitous Computing and Ambient Intelligence*, 427-438. Springer, Cham, Dec. 2015.
- [33] Statista.com. *Museum attendance in England by age 2012-2017 UK survey*. 2018. <https://www.statista.com/statistics/418323/museum-gallery-attendance-uk-england-by-age/>.
- [34] M.N. Islam and H. Bouwman, "Towards user-intuitive web interface design and evaluation: a semiotic framework" *Int. J. Hum-Comput. St.*, **86**, 121-137, 2016, <https://doi.org/10.1016/j.ijhcs.2015.10.003>.
- [35] L. Zakowska, "The effect of environmental and design parameters on subjective road safety - a case study in Poland" *Safety Sci.*, **19** (2-3), 227-234, 1995. [https://doi.org/10.1016/0925-7535\(94\)00023-V](https://doi.org/10.1016/0925-7535(94)00023-V).
- [36] E. Duarte, F. Rebelo, J. Teles, and M.S. Wogalter, "Safety sign comprehension by students, adult workers and disabled persons with cerebral palsy" *Safety Sci.*, **62**, 175-186, 2014, <https://doi.org/10.1016/j.ssci.2013.08.007>.
- [37] S. Rosenbaum and J. Bugental, "Measuring the Success of Visual Communication in User Interfaces" *Soc Tech Commun's Technical Communication*, **45** (4), 517-528, Nov. 1998.
- [38] ISO. ISO 3864:2016 *International standard for safety colours and safety signs*. International Organization for Standardization. Geneva, Switzerland, 2016.
- [39] S. McDougall and M. Curry, "More than just a picture: Icon interpretation in context" in *Proceedings of 1st International Workshop on Coping with Complexity*. University of Bath, 73. September. 2004. http://www.academia.edu/download/3251711/Past_Designs_As_a_Way_of_Coping_With_Complexity.pdf#page=73
- [40] J.M. Silvennoinen, T. Kujala and J.P. Jokinen, "Semantic distance as a critical factor in icon design for in-car infotainment systems". *Appl. Ergon.*, **65**, 369-381, 2017. <https://doi.org/10.1016/j.apergo.2017.07.014>.
- [41] D.A. Norman "Some observations on mental models" in D. Gentner and A. Stevens (Eds.) *Mental Models*. Hillsdale, N.J., Erlbaum Associates. 1983.
- [42] D. Oswald, "Dynamic sense-making in use processes of digital products". in *5th International Congress of the International Association of Societies of Design Research (IASDR)*, Tokyo. Aug. 2013. http://www.david-oswald.de/downloads/IASDR2013_oswald-interface-semiotics.pdf.

APPENDIX A

Effect of context on users' icon recognition rate and overall average IRR%											
User	Average IRR Score per User										
	Gender	Age range	Job	Education	Skill Level	Out of context		In context		Change to IRR + or -	Overall Average
						Score/Max. score	Percentage	Score/Max. score	Percentage		
1	Male	26 - 33	Student	Bachelors	Adv.	20/42	47.6%	32/42	76.2%	+ 28.6%	<u>61.9%</u>
2	Male	26 - 33	Student	Bachelors	Adv.	28/42	66.7%	34/42	81.0%	+ 14.3%	<u>73.9%</u>
3	Male	42 - 49	Employed	Doctors	Adv.	25/42	59.5%	29/42	69.0%	+ 9.5%	<u>64.3%</u>
4	Female	60 - 69	Retired	College	Basic	15/42	35.7%	21/42	50.0%	+ 14.3%	42.9%
5	Female	26 - 33	Employed	Bachelors	Inter.	27/42	64.3%	27/42	64.3%	0.0%	<u>64.3%</u>
6	Male	42 - 49	Employed	College	Inter.	19/42	45.2%	25/42	59.5%	+ 14.3%	52.4%
7	Male	18 - 25	Student	College	Inter.	23/42	54.8%	32/42	76.2%	+ 21.4%	<u>65.5%</u>
8	Female	18 - 25	Employed	College	Basic	16/42	38.1%	26/42	61.9%	+ 23.8%	50.0%
9	Female	18 - 25	Employed	College	Basic	18/42	42.9%	23/42	54.8%	+ 11.9%	48.9%
10	Female	18 - 25	Student	Masters	Adv.	16/42	38.1%	25/42	59.5%	+ 21.4%	48.8%
11	Male	26 - 33	Employed	Masters	Adv.	27/42	64.3%	36/42	85.7%	+ 21.4%	<u>75.0%</u>
12	Female	18 - 25	Employed	School	Inter.	17/42	40.5%	31/42	73.8%	+ 33.3%	<u>57.2%</u>
13	Male	34 - 41	Home maker	College	Basic	27/42	64.3%	29/42	69.0%	+ 4.7%	<u>66.7%</u>
14	Female	50 - 59	Employed	Bachelors	Adv.	28/42	66.7%	30/42	71.4%	+ 4.7%	<u>69.1%</u>
15	Female	50 - 59	Employed	Bachelors	Adv.	12/42	28.6%	19/42	45.2%	+ 16.6%	36.9%
16	Female	34 - 41	Employed	Bachelors	Adv.	15/42	35.7%	25/42	59.5%	+ 23.8%	47.6%
17	Female	18 - 25	Employed	College	Basic	16/42	38.1%	31/42	73.8%	+ 35.7%	56.0%
18	Female	60 - 69	Employed	College	Basic	13/42	31.0%	19/42	45.2%	+ 14.2%	38.1%
19	Male	26 - 33	Student	Masters	Inter.	14/42	33.3%	16/42	38.1%	+ 4.8%	35.7%
20	Male	34 - 41	Student	Masters	Adv.	28/42	66.7%	29/42	69.0%	+ 2.3%	<u>67.9%</u>
21	Male	26 - 33	Employed	Masters	Inter.	29/42	69.0%	35/42	83.3%	+ 14.3%	<u>76.2%</u>
Average of overall averages											57.1%

User icon recognition rate % = (Score / Total possible score) * 100.






















Difference in IRR % = In context IRR % - out of context IRR%

* Change in IRR% in red = negative value and in green = positive value.

Average of overall averages (Column 7) = sum of all user overall averages / number of users

Overall averages above the average of overall averages are underlined

APPENDIX B

Icon recognition rate for each icon 'out of context' and 'in context' showing difference + or -							
Icon			Out of context Score/Max score	Out of context IRR %	In context Score/Max. score	In context IRR %	Difference - out of context and in context
No.	Image	Purpose					
1		Action	33/42	<u>78.6%</u>	35/42	<u>83.3%</u>	+ 4.7%
2		Action	28/42	<u>66.7%</u>	33/42	<u>78.6%</u>	+ 11.9%
3		Information	22/42	52.4%	30/42	<u>71.4%</u>	+ 19.0%
4		Navigation	0/42	0%	5/42	11.9%	+ 11.9%
5		Information	17/42	40.5%	21/42	50.0%	+ 9.5%
6		Action	23/42	54.8%	32/42	<u>76.2%</u>	+ 21.4%
7		Action	21/42	50.0%	28/42	<u>66.7%</u>	+ 16.7%
8		Navigation	25/42	60.0%	35/42	<u>83.3%</u>	+ 23.3%
9		Navigation	7/42	16.7%	4/42	9.5%	- 7.2%
10		Action	28/42	<u>66.7%</u>	31/42	<u>73.8%</u>	+ 7.1%
11		Action	30/42	<u>71.4%</u>	38/42	<u>90.5%</u>	+ 19.1%
12		Action	17/42	40.5%	28/42	<u>66.7%</u>	+ 26.2%
13		Action	25/42	60.0%	30/42	<u>71.4%</u>	+ 11.4%
14		Navigation	25/42	60.0%	36/42	<u>85.7%</u>	+ 25.7%
15		Navigation	9/42	21.4%	3/42	7.1%	- 14.3%
16		Information	24/42	57.1%	33/42	<u>78.6%</u>	+ 21.5%
17		Information	23/42	54.8%	29/42	<u>69.0%</u>	+ 14.2%
18		Action	18/42	42.9%	34/42	<u>81.0%</u>	+ 38.1%
19		Navigation	33/42	<u>78.6%</u>	38/42	<u>90.5%</u>	+ 11.9%
20		Navigation	16/42	38.1%	25/42	60.0%	+ 21.9%
21		Navigation	9/42	21.4%	24/42	57.1%	+ 19.0%

Icon recognition rate % = (Score / Total possible score) x 100.
 Difference in IRR % = In context IRR % – out of context IRR%
 * Icon type in **bold** = dominant type of icon in sign
 IRR score underlined would pass ISO 3864-2:2016 level of 66.7%

Identifiable (60% - 100%)	Mediocre (59% - 30%)	Vague (29 - 0%)
------------------------------	-------------------------	--------------------

Emergence of Fun Emotion in Computer Games -An experimental study on fun elements of Hanafuda-

Yuki Takaoka^{*} 1, Takashi Kawakami², Ryosuke Ooe²

¹Division of Engineering, Graduate School of Engineering, Hokkaido University of Science, 006-8585, Japan

²Department of Information and Computer Science, Faculty of Engineering, Hokkaido University of Science, 006-8585, Japan

ARTICLE INFO

Article history:

Received: 28 August, 2018

Accepted: 29 October 2018

Online: 25 November 2018

Keywords:

Hanafuda

Definition of fun

Questionnaire

ABSTRACT

In recent years, research on game AI has expanded, and now it has become possible to construct even AI of complex games. In accordance with this trend, we constructed the AI of the Hanafuda with a certain degree of complexity. Because of applying the method used in other games to the ball game, we could create a computer player with a certain strength. However, some players feel that strong players are not fun. Therefore, we tried to build a computer player that feels interesting. In the previous experiments, the evaluation for the constructed player was not good. In this research, we conducted a questionnaire survey on players of Hanafuda to raise the evaluation of computer players. The result proved that there are some elements of fun in common among the players.

1. Introduction

Researches on game AI have been actively conducted. Especially in recent years the advancement of computers has made AI for games that has been considered difficult. There are “Mafia (also known as “Werewolf” [1])” and “Poker [2]” in what has been reported. Especially in Mafia, it is proposed to use LSTM (Long Short Term Memory) for analysis of games. Also on poker, it has evolved to defeat Texas Hold'em professional player. In this way, it is possible to create a game AI with a high degree of difficulty and advancement of technology is felt. In response to this trend, we have studied “Hanafuda”, which are card games in Japan. The first step was to strengthen the AI of the Hanafuda. Hanafuda is a unique game in Japan, it cannot be said that research is progressing. Therefore, AI which is installed in various Hanafuda game is moderate in ability, never strong. In order to improve AI's ability, we applied UCT (UCB applied to Tree) algorithm to the Hanafuda. As a result, improvement in ability was seen [3].

However, strong AI was not fun AI. Player who battle against AI are sometimes felt boring. Sometimes players do not want to play against AI. Therefore, we aimed to build an AI that made the human player interesting. In order to achieve this goal, definition of “fun” is necessary. Without the definition of “fun”, computer players cannot produce entertainment. In our research, we have done so far, we defined the following two points as “fun”.

1. increase the variance of get or lose scores

2. adjust the final score of one match to near ± 0

These definitions are those that we have devised their own, it was not accurate. A more precise definition is needed when conducting this research. Therefore, in this research, we aim to accurately define the “fun” of the Hanafuda. Specifically, a questionnaire is made to the people who play the Hanafuda, and the result is analyzed.

The structure of this paper is as follows: First, the Hanafuda that this study targeted is described. Secondly, the results of the questionnaire conducted by this research are described. Finally, we analyze the results obtained and describe the policy of future research.

2. Hanafuda

Hanafuda is traditional Japanese card game, and it is the name of the card used in this game. There are 48 cards in this game. Players aim to acquire them according to the rules. Winning or losing is decided with the card taken by the player.

2.1. Game rules

Hanafuda has various rules. The most mainstream among them is “Koi-Koi”. Koi-Koi is a game that makes a combination of specific cards. This game is done by two players. Players scramble for cards with each other. Cards can be taken by matching with cards of the same suit. Figure 1. shows Hanafuda cards and correspondence between the card and the month. The rows of cards belong to the same suit. However, the value of the card is different

^{*}Corresponding Author: Yuki Takaoka, 7-15-4-1, Maeda, Teine, Sapporo, Hokkaido, Email: 9172001@hus.ac.jp

even within the same suit. Each card belongs to one of “Hikari”, “Tane”, “Tan” and “Kasu”. The value of the card affects when making a winning hand. Figure 2. shows a relationship between cards and classification. Other cards belong to “Kasu”.

In the player’s turn, select a card on player’s hand. If there is card of the same suit in the field, the player can acquire them. At this time, if there are two cards of the same suit in the field, select either one to acquire. If there are three cards, acquire all cards. On the other hand, if there is no card of the same suit in the field, player puts the selected card in field. After the selected card has been processed, excavate the top card of deck, and do in the same way. Up to this point is the turn of the player. At this point, if the winning hand is completed, it is selected whether to continue the game. If player continue, the player aims to create a new winning hand. When player do not continue, player receive points from the opponent player according to the winning hand. If the winning hand is not completed, give a turn to the opponent. The above is a rough rule of Koi-Koi. Based on this, the following section explains the flow of Koi-Koi.

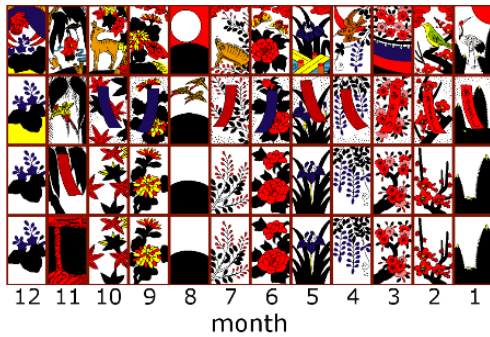


Figure 1. A Hanafuda cards



Figure 2. A relationship between cards and classification

2.2. The flow of Koi-Koi









1. Select the dealer
First, select the dealer. Selection method is random card draw. Each player draws a card; the player who draws a card close to January is a dealer.
2. Deal the cards
The dealer deals cards. Eight cards are dealt to each hand and eight cards are dealt to the field.
3. Game start
A turn is started from the dealer. After that, continue to play until the hand cards runs out or until turn player stops the game on the way. The turn player collects cards according to the rules written in 2.1. If both players run out of cards, give 6 points to the dealer and the next game is started.
4. Game end
It is one game until cards are dealt and either player gains

points. One match is made twelve times in a game. It is the win of the player who has the highest score when one match is over.

2.3. Winning Hans of Koi-Koi

Winning hands of Koi-Koi is as shown in Table 1.

Table 1. Examples of winning hands of Koi-Koi

Name	Combination	Points
Hikari	 (an example)	6~
Inoshikachou		6
Shichigosan		6
Omotesugawara		6
Akatan		6
Aotan		6
Tsukimi-Zake		5
Hanami-Zake		5
Tane	omission	1~
Tan	omission	1~
Kasu	omission	1~

Hikari is determined by the combination of acquired “Hikari” cards. Tane is completed by any five “Tane” cards. Tan is completed by any five “Tan” cards. Kasu is completed by any four “Kasu” cards. Each winning hand, one additional point is awarded for every additional card.

3. Previous study

Previously, we conducted research on “fun”. It was to entertain the players by playing the game according to the definitions of interest we defined. In this research, UCT used for game AI was used. In this chapter, we will describe experiments conducted and UCT which is the method used.

3.1. Monte Carlo method and Monte Carlo Tree Search

Normally, the Monte Carlo method is a method of obtaining results by repeating simulation many times. When this technique is used for a game, it progresses the game at random and judges whether the action is good or bad based on winning or losing. Specifically, play randomly according to the rules of the game, and get win or lose. Next, calculate the expected value of the selected action based on the outcome. This flow is performed for all actions that can be selected at a point in the game, and actions with the highest expected value are selected.

Monte Carlo Tree Search (hereinafter called “MCTS”) is applied to this method for tree search. Characteristics of MCTS are shown in Figure 3. and Figure 4. The MCTS assigns many simulations to the actions that are considered useful (Figure 3.). And when the number of selections of action exceeds a certain value, the tree is expanded (Figure 4.).

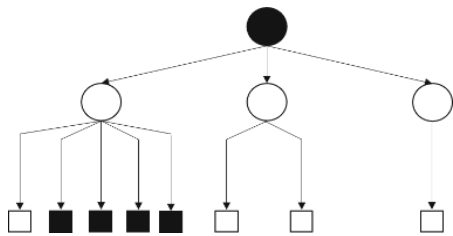


Figure 3. MCTS: many simulations at useful action

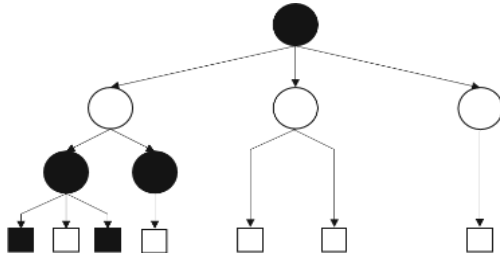


Figure 4. Expansion of tree

The flow of MCTS is as follows.

1. Create a game tree with the current board as the root. In the child node, put behaviors that can be taken at the root node.
2. Proceed the game to the end according to the rules (hereinafter called this act “payout”). At this time, nodes selected a certain number of times are expanded.
3. When winning or losing is confirmed, record it on all the selected nodes.
4. Repeat the specified number of times with 1 to 3 steps as 1 time.

Figure 5. shows these procedures.

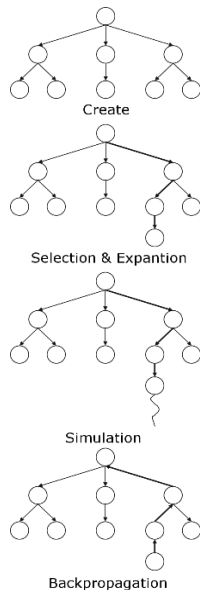


Figure 5. The flow of an MCTS

The advantage of MCTS is, is where it is not necessary design of the evaluation function. By this, it was widely used for games where evaluation of the board surface was difficult. As an example, there is CrazyStone [4] of Go. CrazyStone adopted the MCTS, its ability to win the then Go Tournament at that time.

3.2. UCT

UCT incorporates UCB (Upper Confidence Bound) for tree search. UCB was developed by Auer to solve the Multi-Armed Bandit problem [5]. A commonly used example of the Multi-Armed Bandit problem is a model that gambler plays slot machines. In order to maximize profits, gambler is a matter of thinking which slot machine to play. Gambler uses UCB to solve this problem. The UCB is a value calculated from the play situation of the machine, and by using it, gamblers can make a lot of profit.

UCB is calculated by (1).

$$UCB(i) = \bar{X}_i + c \sqrt{\frac{2 \ln n}{n_i}} \quad (1)$$

\bar{X}_i

UCT handles each child node as a Multi-Armed Bandit problem and performs a tree search. Update the UCB with simulation results and seek the most valuable behavior.

3.3. Research to produce “fun” by UCT

Among the many definitions of fun, we have defined the definition of “fun” in the previous research as follows. [6].

1. increase the variance of get or lose scores
2. adjust the final score of one match to near ±0

The first definition is to avoid becoming a boring match. It tends to be a boring match if there is little exchange of scores. We decided that match will be interesting if player makes active exchange of score, and adopted this definition. The second definition is to improve the impression of the match. Human players tend to worsen the impression of match if they lose a lot to computer players. Conversely, even if a human player wins too much, it is not good and it is necessary to balance. Therefore, we judged that we should adjust the score to around ± 0 without major win or loss. We adopted this definition for the time being because we got a response that suggested this definition to several players and that it is reasonable.

In order to produce the “fun” that we defined, we attempted to change the usage of UCT and realize it. Specifically, change the decision method when selecting a card in hand as follows.

$$Select\ hand = \begin{cases} \min_i |point + UCB(i)| & point > 0 \\ \max_i UCB(i), & otherwise \end{cases} \quad (2)$$

In (2), *point* is the score obtained by the computer player, and *i* is the number of the hand cards. Equation (2) selects a card that loses score if the score of the computer player is plus. For example, the score of the computer player is +7 points. At this time, the computer player selects a card which can be brought closest to ± 0 . In this way we tried to produce “fun”.

However, the experiment result was not good. The comments of the human players who fought against the computer player are shown in Table 2.

Table 2. Human player’s comment

Name	Good point	Bad point
A	<ul style="list-style-type: none"> It is not too strong, but it will not let me win easily. Not a one-sided match, I can enjoy a match to a certain extent. 	<ul style="list-style-type: none"> Computer players sometimes do strange selection.
B	<ul style="list-style-type: none"> I could see it trying to make a game. 	<ul style="list-style-type: none"> The strength was not constant. I felt it was often too weak.

Analyzing these comments, the following can be said.

- It is not so strong that human player cannot win.
- Computer player is trying to entertain them.
- It feels too weak for some people.
- There are scenes in which selection is considered strange.

Accordingly, “fun” defined by us is insufficient, and more precise definition is necessary. In this research, as a first step for accurate definition, it is to investigate the “fun” that the players of the Hanafuda think.

3.4. Other research

Ikeda et al. 's research is an example of research that produces fun. Ikeda et al. researched “computer Go which entertain human players”. For that purpose, Ikeda et al. interviewed experts “What are the elements entertaining Go?” [7]. From the results, researches on elements such as “What is humanness” are under way. Finally, Ikeda et al. said that discovery about “AI like human” was obtained, and they are thinking whether I can create “interesting AI” using this. We decided to conduct research along this trend. As in the case of Ikeda et al.'s research, this research is conducted as a first step to narrow down the elements of fun.

4. Questionnaire survey

In this chapter, the questionnaire survey conducted and the results are described.

4.1. Survey method

We sent a question by e-mail to the person who consented to the investigation and carried out a survey by getting it sent back. The implementation period was from 16th March to 2nd April 2018. There were 76 people who agreed to the survey, 61 of whom returned the mail within the period, and the response rate was 80.3%. Therefore, the total number of respondents is 61 people. In addition, this questionnaire survey is targeted to players who

regularly play Hanafuda, and it is not a questionnaire after collecting participants and making a match.

4.2. Survey content

We made the following items as questionnaire contents:

- Attributes of survey target
 - Age and gender
 - Number of years that they’ve been playing Hanafuda
 - Frequency of playing Hanafuda
- Questions about fun
 - Elements making the Hanafuda interesting
 - Elements making the Hanafuda uninteresting
 - Recommendations for Hanafuda

Question 2 is a form of free description.

5. Questionnaire results

5.1. Attributes of survey target

A. Age and gender

Regarding gender, all respondents were male. The age was wide, ranging from teens to 50’s, with an average age of 31.9 years.

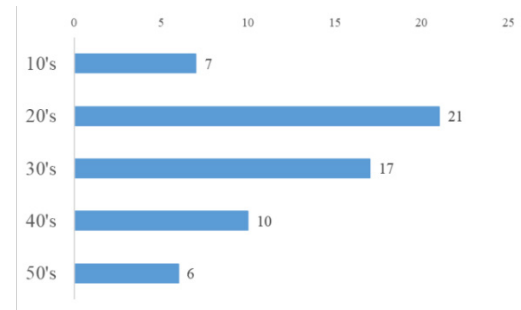


Figure 6. Age of respondent

B. Number of years that they’ve been playing Hanafuda

The minimum was one year and a half, and the maximum was 15 years.

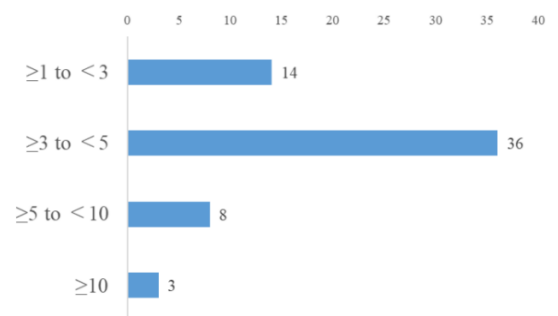


Figure 7. Play number of years

C. Frequency of playing Hanafuda

The most frequent answers were about once a week. The next most frequent answer was about once a month, under it 3 to 4 times a week. In addition, there was an answer that it does not go regularly but does it at a specific time. Incidentally, this answer

includes not only using actual cards, but also those play on electronic devices.

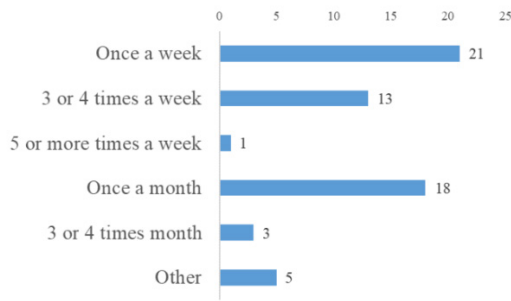
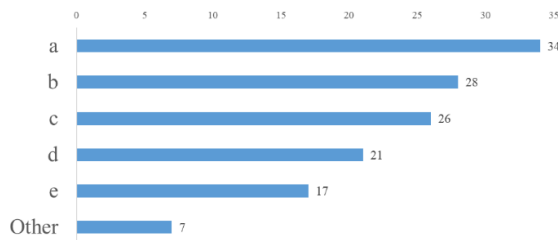
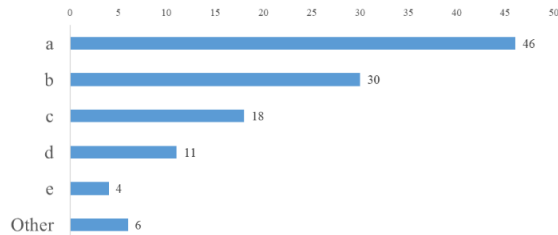


Figure 8. Frequency of playing Hanafuda



- a: Completely victory to opponent
- b: Got a big score
- c: Won by reversing
- d: Dodged opponent's attack
- e: Interfered the opponent

Figure 9. Elements making the Hanafuda interesting



- a: The game ends at an early stage
- b: Could not win the game even once
- c: The score gained by the opponent is low
- d: Drew cards are bad
- e: Dealt cards are bad

Figure 10. Elements making the Hanafuda uninteresting

5.2. Questions about fun

A. Elements making the Hanafuda interesting

We asked the fun elements of the Hanafuda. The results are shown in Figure 6.

Elements with many responses are as follows:

- Completely victory to opponent
- Got a big score
- Won by reversing

These items were the top 3.

B. Elements making the Hanafuda uninteresting

Contrary to 1., we asked the suffering elements of the Hanafuda. The results are shown in Figure 7.

C. Recommendations for Hanafuda

We asked about the recommendations for the Hanafuda. The most frequent content is about “Tsukumi-Zake and Hanami-Zake”. Tsukumi-Zake and Hanami-Zake can get 5 points with two cards. Therefore, the degree of difficulty is low, and the player can easily make a Winning-Hand. It seems that there are many players who view this point as a problem. The second frequent answer was about the rules of Hanafuda. There are various rules in the Hanafuda, but what is mainstream now is “Koi-Koi”. There seems to be some players who wish that not only this but also more diverse rules become common. Other, many recommendations on rules and dissemination were received.

6. Future research policy

Future research will conduct a detailed questionnaire survey based on the obtained answers. In the present situation, only the element of the “fun” of the Hanafuda that each player thinks were obtained, and it is necessary to seek elements that satisfy many players. As a specific procedure, we conduct a questionnaire survey of 2 choices for each element obtained in this research. For example, “Yes / No” answers to the question “Do you feel that “Completely victory to opponent” is interesting?”. Based on the questionnaire survey, we aim to create computer players that many players feel interesting. It is also necessary to ask many players questions that can refine questionnaire questions and analyze the essence of interest. After conducting detailed questionnaires, detailed analysis is also required.

7. Conclusion

In this research, we investigated interesting elements of the game with the goal of creating a computer player that makes human players fun. Among various kinds of games, this research targeted the Hanafuda that are games in Japan. Survey was conducted by sending questions to players who play Hanafuda. By the reply sent back, we got the fun that each player thinks. In future, it is necessary to further investigate and seek more definite definition of fun.

References

- [1] M. Kondoh, et al., “Development of Agent Predicting Werewolf with Deep Learning” in International Symposium on Distributed Computing and Artificial Intelligence, Toledo Spain, 2018. https://doi.org/10.1007/978-3-319-94649-8_3
- [2] M. Moravcik, et al., “Deepstack: Expert-level artificial intelligence in heads-up no-limit poker”, Science, 356(6337), 508-513, 2017. <https://doi.org/10.1126/science.aam6960>
- [3] Y. Takaoka, et al., “A study on strategy acquisition on imperfect information game by UCT search” in the 2017 IEEE/SICE International Symposium on System Integration, Taipei, Taiwan, 2017. <https://doi.org/10.1109/SII.2017.8279334>
- [4] R. Coulom, “Monte-Carlo Tree Search in Crazy Stone” in Game Prog. Workshop, Tokyo, Japan, 2007.
- [5] P. Auer, et al., “Finite-time Analysis of the Multiarmed Bandit Problem”, Machine Learning, 47(2-3), 235-256, 2002. <https://doi.org/10.1023/A:1013689704352>
- [6] Y. Takaoka, et al., “A Fundamental Study of a Computer Player Giving Fun to the Opponent”, Computer Science & Communications, 6(1), 32-41, 2018. <https://doi.org/10.4236/jcc.2018.61004>
- [7] M. Yamanaka et al., “Bad Move Explanation for Teaching Games with a Go Program”, IPSJ SIG Technical Report, 2016-GI-35(5), 1-8, 2016. (Japanese)

Metaheuristics for Solving Facility Location Optimization Problem in Lagos, Nigeria

Chika Yinka-Banjo*, Babatunde Opesemowo

Computer Science, University of Lagos, Akoka, Lagos, 100213, Nigeria

ARTICLE INFO

Article history:

Received: 24 August, 2018

Accepted: 10 November, 2018

Online: 25 November, 2018

Keywords:

Metaheuristics

Facility Location Problem

Optimization

ABSTRACT

Facility location problem is a problem that many organizations still face today because of its increasing constraints and objectives. Decision makers want this problem solved in order to maximize profit and as such, it became a field of interest to many computer scientists over the years. The solution tool used by these scientists; a function of technological advancement, has evolved from the use of classical mathematical approaches to the use of metaheuristics. Some of the metaheuristics used include particle swarm optimization metaheuristics, genetic algorithm metaheuristics and tabu search. The problem considered in this research evolves the study of waste management in Lagos state. How the location of waste evacuation centers could be allocated in order to minimize resources such as transportation cost, facility cost, distance and the capacity of each centers. A mathematical model was developed that serves as a template for the algorithm used to solve the problem. Then particle swarm optimization metaheuristics was used to find the optimal solution in terms of capacity to the problem. Particle swarm optimization minimizes the use of memory and still gives a satisfactory solution. With the result obtained, respective agencies could make good decision as to the best location to build a new facility.

1. Introduction

The study of facility location problem, also known as location analysis, is a branch of operations research and computational geometry concerned with the optimal placement of facilities to minimize resources such as cost, time, distance, etc. In most cases, the cost minimized is transportation cost. There are different areas facility location problem could be considered. These areas may include logistics facilities, warehouses, farm facilities, police stations etc. For this project, we considered a waste management facility known as evacuation centres. Waste generated from the residents at the state are dumped at the evacuation centres where it is recycled. We chose the waste management facility problem because of the way it affect our environment especially the Lagos state government. The state is very keen to the management of waste generated because of her thriving effort to becoming a green state. In this effort, the government has employed private own company known as Visionscape for the management of her waste. We have approximately 21 million people living in Lagos state and quite a large amount of waste are generated in the state daily. When the waste are not well managed based on poor allocation of the evacuation centres this could lead to air pollution of the environment. However, selecting locations out of the blues are not

best measures to obtain a reasonable result. Hence, after the study of major works done on facility location, we embark on finding scientific ways of solving the problem that will give a satisfactory result. This led to the use of metaheuristics.

Metaheuristics is a higher-level procedure or heuristic designed to find, generate, or select a heuristic (partial search algorithm) that may provide a sufficiently good solution to an optimization problem, especially with incomplete or imperfect information or limited computation capacity. Examples of metaheuristics are particle swarm optimization (PSO), hill climbing, simulated annealing, tabu search, genetic algorithm, etc. Figure 1 shows the classification of metaheuristics. Different analysis have been used over the years for these classes to be obtained and presented as a clear cut direction [1].

2. Related Literature

Over the past decades, the study of facility location problem has grown even more popular, not only in the academic literature but also in practice. Facility location problems are often strategic in nature and entail long-term decisions, exposing firms to many uncertainties during the operational lifetime of a facility. When solving such problems, a firm may have to determine the number of facilities to open, their locations, and their capacities. Because

*Chika Yinka-Banjo, Email: cyinkabanjo@unilag.edu.ng

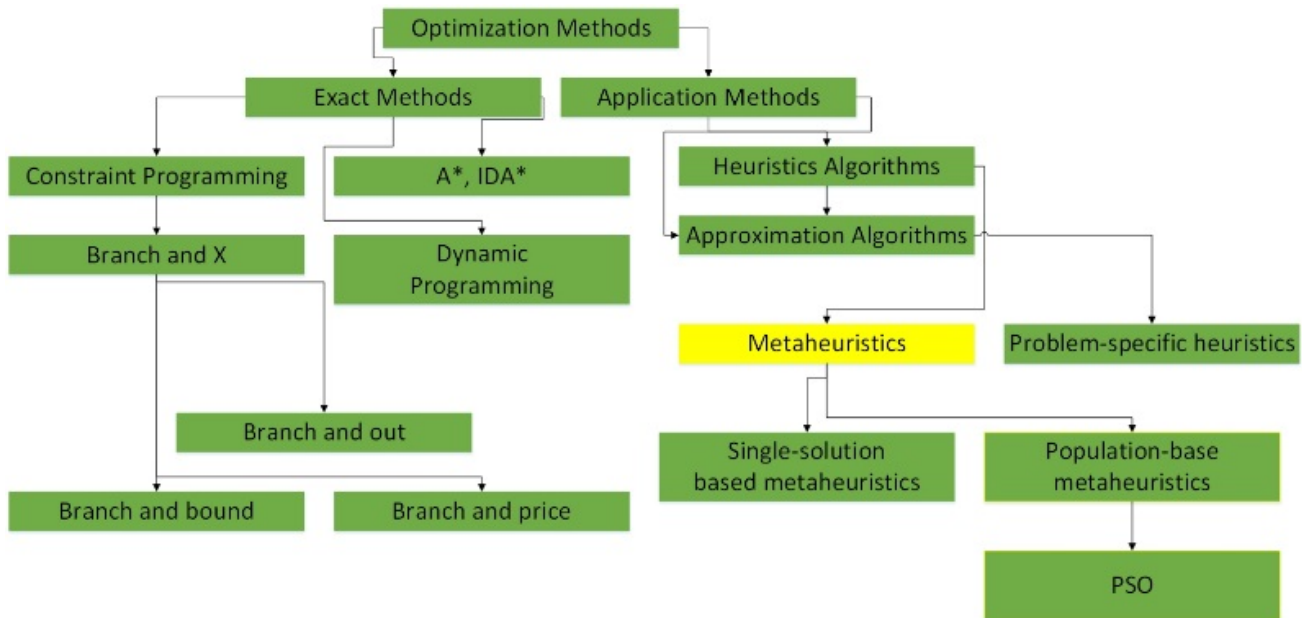


Figure 1: Classification of metaheuristics.

of the high fixed costs incurred in changing a network of facilities, a firm may be limited in the frequency in which it re-examines these strategic decisions. After determining its facility network, the firm is often relegated to making operational decisions such as determining production quantities, service levels, and allocation of supply to demand. Thus, facility location problems often have a two-stage approach to solving them, location followed by evaluation. Facility location problem belongs to the class of a combinatorial optimization problem [2].

The 17th Century marks the beginning of the facility location problem which was considered as a classical mathematical problem [3]. As time went on, it became an interest to many scientists and researchers because of how it relate to real life issues. Also, various aspect of facility location problem have been considered by earlier scientist ranging from customer facility management to logistics management. [4]. However, the problems are majored on customer and warehouse facility. Also, facility location problem is further considered as capacitated [5, 6, 7] and uncapacitated [8]. Capacitated in terms of considering the size of the facility involved or the ability of the facility to manage other entities involved like customer adequately. While uncapacitated doesn't consider the size of the facility, it assumes the facility could always manage the entities involved. Both capacitated and uncapacitated were considered by Bumb and March [9].

Mauricio and Renato presented a multistart heuristic for the uncapacitated facility location problem [10]. The method used combines elements of several other metaheuristics, such as scatter and tabu search (which make heavy use of path-relinking) and genetic algorithms (that deals with the notion of generations) to proffer solutions.

In order to have a good understanding of the various state of facility location problems, Andreas et al. [11] studied continuous location models, network location models, mixed-integer programming models, and applications of facility location problem. In [12], the author proposed particle swarm optimization, simulated annealing and iterated local search metaheuristics for

solving facility location problem that has to do with the health sector. The problem was modelled and the experimental results show that the proposed algorithms reach acceptable performances in a reasonable computation time. In [13], the authors considered forty hospitals and three candidate municipalities in the sub-northeast region of Thailand, and considered multiple factors such as infrastructure, geological and social & environmental factors, and calculating global priority weights using the fuzzy analytical hierarchy process (FAHP). Also, a new multi-objective facility

location problem model which combines FAHP and goal programming (GP), namely the FAHP-GP model, was tested. Their proposed model led to selecting new suitable locations for infectious waste disposal by considering both total cost and final priority weight objectives. The novelty of the proposed model is the simultaneous combination of relevant factors that are difficult to interpret and cost factors which require the allocation of resources.

From the above literature, facility location problems cut across different horizon of research, and few research have provided an overview of the models and algorithms that are applied to the optimization in solving facility location problem that has to do with waste management.

3. Methodology

In this paper, we studied the problem of allocation of new evacuation centres at different potential locations such that it could manage the local government areas assigned to it in relation to the existing potential locations. The major entities are evacuation centres, potential location and local government areas. Evacuation centres are spaces where wastes are dumped and recycled. Whereas, potential locations are the positions where a facility can be allocated. Finally, the waste are generated from the local government areas in a particular city. The problem was well formulated following the principles that governs optimization model. The objective of the optimization model is to minimize the total cost of evacuation sites to be located, chosen among a set of

candidate locations. Such objective ensures not only the reduction of the visual impact due to the presence of collection sites, but also the reduction of the overall cost related to the collection phase. Following below are the indices, parameters and decision variables for the model.

3.1. Definition of model components

INDICES:

i = is the index of each potential location(Town) in the state, $i = 1, 2, \dots, m$ ($m=3$)
 j = is the index of each Evacuation Center, $j = 1, 2, \dots, n$ ($n=2$)
 k = is the index of each Local Government Area (LGA), $k = 1, 2, \dots, t$ ($t=21$)

PARAMETERS:

f_i = is the facility cost (Naira)
 w_i = is the weight of the factor affecting a location
 p_{ik} = is the transportation cost between the potential location and LGA(Naira)
 d_{ik} = is the distance between the potential location and LGA(km)
 s_j = is the size of each evacuation center
 x_k = is the quantity of refuse generated by a LGA

DECISION VARIABLES:

Y_{ik} = is a binary variable where it is 1 if the LGA is served by Potential Location k and 0 if not served
 H_i = is a positive integer value where it is 1 if the Potential Location is selected and 0 if not selected
 O_{ij} = is a binary variable where it is 1 if the Potential Location is opened by selecting Evacuation Center j and 0 if otherwise

$$\min. Z = \sum_{i=1}^m \sum_{j=1}^n f_i O_{ij} + \sum_{i=1}^m \sum_{j=1}^n w_{ij} O_{ij} + \sum_{i=1}^m \sum_{k=1}^t p_{ik} d_{ik} Y_{ik} \quad (1)$$

Subject to:

$$\sum_{i=1}^m Y_{ik} = 1 \quad \forall k \{1, 2, \dots, t\} \quad (2)$$

$$\sum_{j=1}^n s_j O_{ij} \geq \sum_{k=1}^t x_k Y_{ik} \quad \forall i \{1, 2, \dots, m\} \quad (3)$$

$$\sum_{j=1}^n O_{ij} = H_i \quad (4)$$

$$O_{ij} \in \{0, 1\} \quad (5)$$

$$Y_{ik} \in \{0, 1\} \quad (6)$$

$$H_i \in \{0, 1\} \quad (7)$$

Equation (1) is the objective function which is defined to minimize the total cost based on the facility cost, factors affecting the potential location and the transportation cost. Equation (2) ensures that the demand of each LGA k is met. Equation (3) ensures that the sum of service by the LGA does not exceed the capacity of the Evacuation Center. Equation (4) ensures that the selected LGAs must use k -size Evacuation Centers. Constraints (5), (6) and (7) are binary decision variables. From this model, we applied particle swarm optimization technique to proffer an optimized solution on the capacity for each evacuation centers. For the PSO algorithm the following parameters below were used in relation to what was used by Jordanski:

- Initial weight $w = 0.9$
- Number of particles = 40

- Cognitive constant $c1 = 2$
- Social constant $c2 = 2$
- Maximum velocity $v_{max} = \text{capacity_range}[0]$
- Minimum velocity $v_{min} = \text{capacity_range}[1]$

Pseudocode 1: To find the optimal location for the evacuation centers

- Initialize Parameters, Decision Variables and Indices
- For (All number of Locations) {
 - For (All number of LGA) {
 - $\text{ObjCost} \leftarrow \text{FacilityCost} + \text{WeightCost} + \text{TransportationCost}$
 - }
 - $\text{ObjCostRow} [] \leftarrow \text{ObjCost}$
 - }
 - $\text{MinObjCost} \leftarrow \text{sort}(\text{ObjCostRow})$
 - $\text{MinCapacity} \leftarrow \text{Pso}(\text{Capacity})$
 - Return $\text{MinObjCost}, \text{MinCapacity}$

Pseudocode 1 shows the steps involved to solve the facility location problem for the evacuation center. All the parameters, decision variables and indices were initialized. An iteration was done to get the mapping for the potential location to their respective local government area (LGA). This will enable us generate the network analysis for the problem and find the minimum mapping that will guarantee a minimum cost based on all the criteria considered. The single cost for each mapping are placed in an array called ObjCostRow as shown above. It is sorted in an ascending order and pass to MinObjCost . The capacity generated for each evacuation center is passed to the PSO function which returns the minimal capacity for evacuation centers. Lastly, the result are displayed.

4. Application of the proposed methodology

The method was implemented with the PHP programming language and designed with web technologies such as HTML and CSS to make it easy for users to interact with. All the user needs to do is to supply the required parameters needed in the right form and the system validates the parameters provided and process it. Finally, the result is displayed to the user.

Potential Location	Evacuation Center	LGA
<input type="text"/>	<input type="text"/>	<input type="text"/>
Facility Cost	Weight Cost	Capacity Range
<input type="text"/>	<input type="text"/>	<input type="text"/>
Quantity	<input type="text"/>	
Distance	<input type="text"/>	
Transportation Cost	<input type="text"/>	
<input type="button" value="Solve this Problem"/>		

Figure 2: User interface for the parameters.

- **Potential Location:** This field requires the user to enter the number of potential location in consideration. The potential location is the location where the user desire to locate the facility. The value must be an integer.
- **Evacuation Centre:** This field requires the user to enter the number of evacuation center required to be allocated at the potential location. The value must be an integer.
- **LGA:** This field requires the user to enter the Local Government Area in the state where the refuse are coming that are to be managed by the evacuation center. The value must be an integer.
- **Facility Cost:** This field requires the user to enter the cost of building as the case may be at a particular potential location. The values are separated by a comma. The values must be a set of numbers and must correspond to the number of potential location set.
- **Weight Cost:** This field requires the user to enter the cost of the weight generated for each potential location. The weight is generated by considering the factors affecting each potential location. This factors may include geographical factor, closeness to raw materials, electricity, good roads, etc. The values are separated by a comma. Also, the values must correspond to the number of potential location set.
- **Capacity Range:** This field requires the user to enter the capacity search space for the problem. The capacity that is required to be managed for each evacuation center. This allows the user to know the required capacity to be considered for building a particular facility.
- **Quantity:** This field requires the user to enter the amount of waste in kilograms (Kg) generated by a particular LGA per day/month/year. The values could be a set of real numbers and separated by commas.
- **Distance:** This field requires the user to enter the distance from a particular LGA to a potential location in miles. The distance of the set of LGA is separated by commas and separated by a forward slash with respect to a particular potential location.
- **Transportation Cost:** This field requires the user to enter the transportation cost for each LGA to a particular potential location. The set of transportation cost are separated by commas and separated by a forward slash with respect to a particular potential.

We carried out 3 different experiment base on the number of potential locations, evacuation centers and local government areas. We shall be considering the operation of experiment one and give a summary of the three experiment.

Table 1: Parameters for Experiment 1

m	n	t	f_i	w_i	S_j
3	2	21	10,20,30	50,70,40	20,0
x_k	4,3,4,5,2,3,3,4,6,5,6,3,4,5,2,3,3,4,6,5,6				
d_{ik}	2,1,2,3,4,5,6,6,8,8,9,9,1,2,1,2,3,3,3,3,2/ 3,2,3,4,5,1,2,2,9,2,2,3,4,5,2,3,3,4,6,5,6/ 4,3,4,5,2,3,3,4,6,5,6,3,4,5,2,3,3,4,6,5,6				
p_{ik}	2,1,2,3,4,5,6,6,8,8,9,9,1,2,1,2,3,3,3,3,2/ 3,2,3,4,5,1,2,2,9,2,2,3,4,5,2,3,3,4,6,5,6/ 4,3,4,5,2,3,3,4,6,5,6,3,4,5,2,3,3,4,6,5,6				

Table 1 shows that we have 3 potential locations but we want to build only 2 evacuation centers that will minimize general cost (facility cost, transportation cost, and weight cost). Using the arbitrary parameters from Table 1, the objective function was generated following the operations described in Pseudocode 1 is shown in Figure 3.

=====OBJECTIVE FUNCTION=====

OPTIONS/FACILITY	COST	A	B	C
1	967	Close	Open	Open
2	987	Open	Close	Open
3	892	Open	Open	Close

Figure 3: Objective function result for experiment one.

From the objective function shown in Figure 3, we would see that the minimum cost is given by option 3. This option means the best selection would be facility A and B. The network analysis that gives this cost is shown in Figure 4. Hence, facility A would serve LGA (1,2,3,4,5,9, 13,14,15,16,17,18,19,20,21) while facility B serve LGA (6,7,8,10, 11,12).

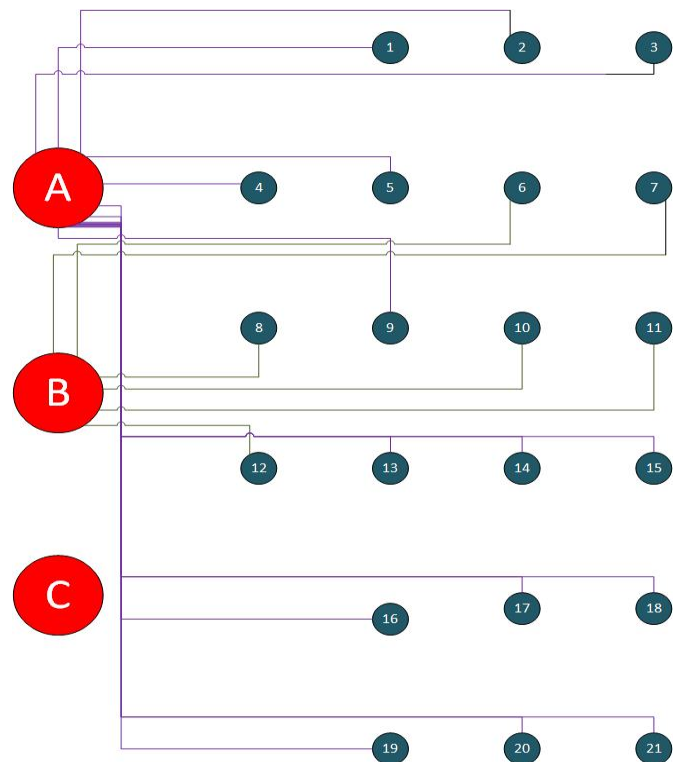


Figure 4: Network analysis for experiment one.

=====CAPACITY=====

FACILITY	CAPACITY
A	64
B	12
C	0

Figure 5: Capacity for experiment one.

Figure 5 shows the capacity for the facilities. That is, the minimum quantity a facility should manage. The same process as explained in Section 4 was carried out for two different scenerios with respect to the potential location, evacuation center and LGA. The time duration was recorded to evaluate the scalability of the solution in relation with time as shown in Figure 6 and 7.

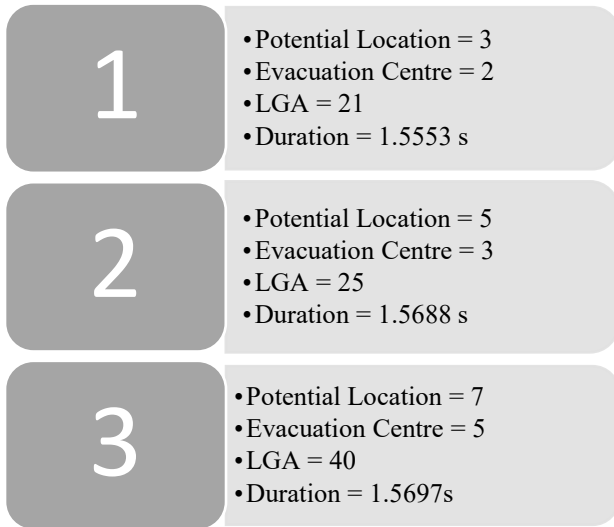


Figure 6: Summary of the experiments.

From Figure 7, we could discover that the time duration compare to the other parameters is significantly small.

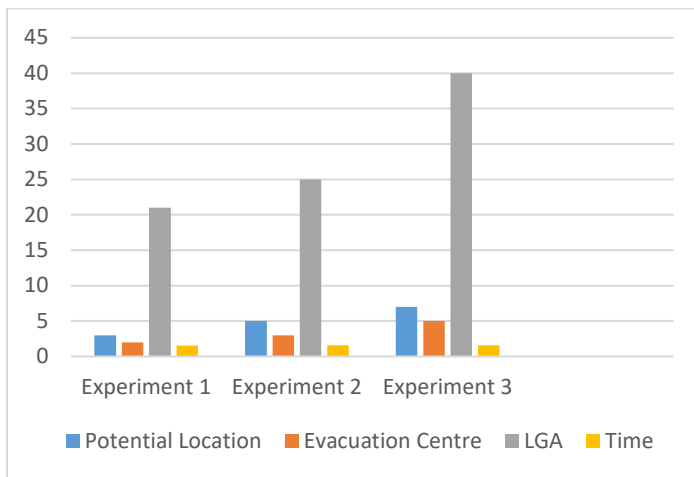


Figure 7: Graphical representation of the experiments.

5. Conclusion

In this project, facility location problem was considered in regards to optimization. The waste management facility serve as the underlying problem dealt with among several problems related to the facility location problem. This problem was chosen because of it adverse effect in the economy of Nigeria especially Lagos state. A redefined PSO was proposed in solving the problem which involves finding the best location to build a set of evacuations canters that would manage the waste gotten from several local government. As most common with facility location problem, the mathematical model to the problem was given which includes the objective functions and the set of constrains involved.

Then a set of instances was used to test the model. The experimental result was gotten that shows the optimal solution to the problem. Hence, we have been able to develop a model that can be used to solve the facility location problem and proffer an optimized solution to the waste management facility location problem. The application developed can also accept other values which requires solution relating to the model worked on making it a flexible solution. Further research should be carried out in other problem areas like the police station facility location, farm facility allocation and any of those that could help in building the nation. Metaheuristics are very good tool in regards to problems that tends to be complex and are unable to be solve within a limited amount of time. However, they might not guarantee optimal solution in some cases. Hence, they must be applied carefully to solve a particular problem.

References

- [1] E. G. Talbi, *Metaheuristics - From Design to Implementation* (John Wiley & Sons, New Jersey, 2009)
- [2] S. Arifin. Location allocation problem using genetic algorithm and simulated annealing. A case study based on school in Enschede (Doctoral dissertation, Master's thesis, the University of Twente, Enschede, the Netherlands, 92p) (2011)
- [3] R. Guner and S. Mehmet. A Discrete Particle Swarm Optimization Algorithm for Uncapacitated Facility Location Problem. *Journal of Artificial Evolution and Applications*, (2008).
- [4] R.H Ballau, Facility Location Decisions. In *Business Logistics: Supply Chain Management* (2004) 550-617
- [5] Ling-YunWua, Xiang-Sun Zhang, Ju-Liang Zhang. Capacitated facility location problem with general setup cost. *Computers & Operations Research*, (2006).
- [6] H. Pirkul and V. Jayaraman. A Multi-Commodity, Multii-Plant, Capacitated Facility Location Problem: Formulation and Efficient Heuristic Solution. Elsevier Science Ltd, (1997).
- [7] Y. Ren. Metaheuristics for multiobjective capacitated location allocation on logistics networks (Doctoral dissertation, Concordia University, 2011).
- [8] R. L. Francis, L. F. McGinnis and J. A. White, *Facility Layout and location: an analytical approach* (Pearson College Division, 1992).
- [9] A. Bumb and W. K. March. A simple dual ascent algorithm for the multilevel facility location problem: Approximation, Randomization, and Combinatorial Optimization. *University of Twente* (2001) 55-63
- [10] G. R. Mauricio and F. W. Renato. A hybrid multistart heuristic for the uncapacitated facility location problem. *European Journal of Operational Research*, (2005).
- [11] A. Klose and A. Drexl. Facility location models for distribution system design. *European Journal of Operational Research*, 162 (2005) 4-29.
- [12] M. Jordanski. Metaheuristic Approaches for Solving Facility Location and Scale Decision Problem with Customer Preference. *IPSI Bgd Internet Research Society*, 13 (2017).
- [13] N. Wichapa, P. Khokhajaikiat. Solving multi-objective facility location problem using the fuzzy analytical hierarchy process and goal programming: a case study on infectious waste disposal centers. Elsevier Science Ltd, (2017).

Contract Price Model Under Active Demand Response

Zhijian Liu, Ni Xiao, Hui Xu*

Kunming University of Science and Technology, Department of Energy and Power Engineering, Electric Power System and Automation, 650500, China

ARTICLE INFO

Article history:

Received: 03 July, 2018

Accepted: 20 November, 2018

Online: 25 November, 2018

Keywords:

Active demand response

Contract price

Electricity consumption mode

Genetic algorithm NSGA-II

ABSTRACT

The power market in China is faced with problems such as the continuous widening of terminal load peak-valley difference and the imbalance between supply and demand. Under this background, this paper to establish the mathematical model of contract price of active demand response. The first is to consider the load component of the user's electricity consumption, whether it can be transferred or can be reduced. Then consider the market incentive conditions and purchase costs, whether the market electricity price difference is enough to attract users to transfer or reduce the load during peak hours. Those can guide users as much as possible Improve the peak consumption pattern, that is, reduce the peak load during peak hours, or use it during off-peak hours.

1. Introduction

Before the reform of the power market, there was an imbalance between supply and demand of the power industry. In addition, due to the lack of coordination and interaction, it is easy to cause power shortage in peak hours, power surplus or insufficient utilization rate of generating units in low periods [1]. And demand response (DR) refers that demand side management how to make full use of market regulation, maintaining the safety stability of the power system operation and the optimal allocation of energy demand side. Simply, demand response is a series of measures to bring the demand side of the power market into the process of feed-related incentive policies [2].

Then, there are some solidification problems in the current price methods in China, and the peak - valley ratio in China is unreasonable. Peak - valley ratio in our country only 2 ~ 3 times, but the peak - valley ratio abroad in the general case for 5 ~ 8 times, even 9 ~ 10 times [3-4]. Enough difference in electricity price can generate effective incentive level for users.

Thus, this paper has established the contract price model under active demand response, which in China are rare research issues about price model.

2. Contract Price Model Under Active Demand Response

2.1. Representation of Contract Price Model

The standpoint of active demand response is to emphasize the enthusiasm and initiative of users to participate in demand

response. The key to achieving this goal is to motivate users in what ways and improve the users' participation for DR's enthusiasm and initiative, which is the core content of this paper.

Based on the above viewpoints, this paper establishes the contract price model under the active demand response. The user purchases electricity from the selling company unilaterally and signs the corresponding electricity contract with the selling company. The contract type is adopted in this paper:

- (1) The price of 24 nodes per day within the trading period has been fixed;
- (2) Regulate the upper limit and lower limit of quantity, and the adjustment method. Namely, if the power consumption of users exceeds the specified range, they will charge DR fees according to the content.

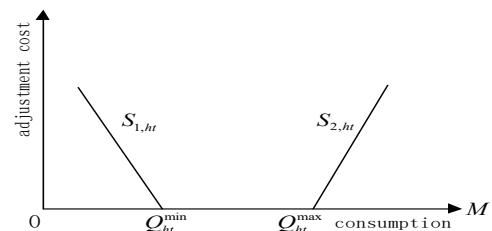


Figure1. Mathematical model of contract mechanism

2.2. Implementation Mechanism of Contract Price Model

In this contract trading mode, when the power consumption of users is not within the contract range, the users will face certain

*Corresponding Author: Hui Xu, Email: 2807511967@qq.com

www.astesj.com

<https://dx.doi.org/10.25046/aj030640>

DR over-limit adjustment fees. Specific rules are shown in figure 1.

As shown in figure 1, Q_{ht}^{\min} and Q_{ht}^{\max} respectively represent the lower and upper limit of the contract quantity. $S_{1,ht}$ and $S_{2,ht}$ represents corresponding over-limit adjustment coefficient. M is a close to the positive infinite natural number. Mark different situations of power consumption transactions with piece-wise functions.

- (1) $0 \sim Q_{ht}^{\min}$: The actual electricity consumption of the user is less than the electricity consumption stipulated in the contract price.
- (2) $Q_{ht}^{\min} \sim Q_{ht}^{\max}$: The actual power consumption of the user is within the scope specified in the contract price.
- (3) $Q_{ht}^{\max} \sim M$: The actual power consumption of the user is greater than that of the contract price.

Therefore, in the specific implementation process of the contract price, the user's electricity expenses during the trading period include: P_{fs} -- the basic electricity fee charged according to the time-sharing electricity price, and P_{ht}^{DR} -- the additional DR fee charged after the user exceeds the power consumption scope.

The total electricity charge is P_{ht} :

$$P_{ht} = \sum_t P_{fs} + \sum_t P_{ht}^{DR}$$

$$P_{fs} = p'_{fs} * Q_t$$

$$P_{ht}^{DR} = \begin{cases} S_{1,ht}(Q_{ht}^{\min} - Q_t) & (0 \leq Q_t \leq Q_{ht}^{\min}) \\ 0 & (Q_{ht}^{\min} \leq Q_t \leq Q_{ht}^{\max}) \\ S_{2,ht}(Q_t - Q_{ht}^{\max}) & (Q_{ht}^{\max} \leq Q_t \leq M) \end{cases}$$

Where: p'_{fs} is the time-sharing electricity price of the time period; Q_t is the power consumption of users; Q_{ht}^{\max} and Q_{ht}^{\min} are respectively the upper limit and lower limit of quantity of the contract quantity.

Therefore, in order to establish the contract price model under the active demand response, the threshold value (Q_{ht}^{\min} , Q_{ht}^{\max}) of the contract quantity must be determined. As well as the DR over-limit adjustment coefficient $S_{1,ht}$ and $S_{2,ht}$, the following chapters will make further research and analysis.

3. Contract Price Model Design

3.1. Electricity Structure of Users

In general, users' electric load can be divided into inelastic partial load, transferable load and reducible load [5-6]. The www.astesj.com

inelastic part load is the user's unchangeable rigid power load. The transferable load refers to the power load that users can transfer from the peak period to the non-peak period. The reducible load is to point to the power load that the user can reduce directly in a certain period of time and won't transfer to other periods of time consumption.

The mathematical relation can be described as:

$$E_t = E_t^{DI} + E_t^{DE}$$

$$E_t^{DE} = \alpha_t E_t^{DE} + \beta_t E_t^{DE}$$

Where: E_t is the total load of the user; E_t^{DI} and E_t^{DE} respectively represent the inelastic part and the elastic part of the electric load; α_t is the transferable load factor and β_t is the reducible load factor.

Under ideal condition, the upper limit quantity of DR contract price model should be inelastic load of peak period. The lower limit quantity shall be the minimum load in the low period, mainly to meet the basic electricity demand of users in the period.

$$Q_{ht}^{\max} = Q_1^{DR} = Q_f^{\max} - \alpha_t E_t^{DE} - \beta_t E_t^{DE} \tag{6}$$

$$Q_{ht}^{\min} = Q_g^{\min} \tag{7}$$

where: Q_1^{DR} is the power consumption limit value adjusted by DR after considering the load characteristics. Q_f^{\max} is the maximum power consumption during peak period of users before DR. Q_g^{\min} is the minimum power consumption during low period of users before DR.

3.2. Incentive Conditions of Electricity Price

In the literature [7] pointed out that, according to the principle of consumer psychology, the user's stimulation have a just noticeable difference (difference threshold). Within the scope of the user basically no reaction or very small, not sensitive period (dead zone); beyond this range user will respond, but the degree of response is related to the degree of stimulus, namely the normal response period (equivalent to the linear zone). The user reaction also has a saturation value for the stimulus, which is the reaction time limit (equivalent to the saturation area).

Usually to simplify the problem, the reaction is represented by a piece wise linear function.

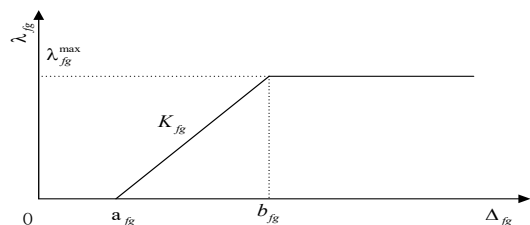


Figure 2. Price incentive and user response curve

The load transfer peak - valley rate of users is expressed by the formula:

$$\lambda_{fg} = \begin{cases} 0 & (0 \leq \Delta_{fg} \leq a_{fg}) \\ K_{fg} (\Delta_{fg} - a_{fg}) & (a_{fg} \leq \Delta_{fg} \leq b_{fg}) \\ \lambda_{fg}^{\max} & (\Delta_{fg} \geq b_{fg}) \end{cases}$$

$$\Delta_{fg} = P_f - P_g$$

$$b_{fg} = \lambda_{fg}^{\max} / K_{fg} + a_{fg}$$

Where: P_f and P_g are respectively the price of electricity during peak and low periods. a_{fg} is the threshold value of the insensitive period (dead zone), which is the first unknown variable. Point $(b_{fg}, \lambda_{fg}^{\max})$ represents the ultimate point of load transfer between peak-valley, λ_{fg}^{\max} is the ultimate load transfer rate and b_{fg} is the threshold value of the reaction saturation area, which is the second unknown variable. K_{fg} is the slope of the user reaction linear region, which is the third unknown variable. The relationship between user response and price incentive can be obtained by fitting these three variables.

Thus, the DR load limit of peak - valley transfer can be further obtained:

$$Q_2^{DR} = Q_f^{\max} - \lambda_{fg}^{\max} Q_f^{\max} \quad (11)$$

Where: Q_2^{DR} is the limit value of consumption adjusted by DR incentive condition; Q_f^{\max} is the maximum consumption during peak period of users before DR.

The initial value in the previous section is further modified in the following three cases:

$$Q_{ht}^{\max}$$

situation	$Q_1^{DR} < Q_2^{DR}$	$Q_1^{DR} = Q_2^{DR}$	$Q_1^{DR} > Q_2^{DR}$
conditions	It satisfies the structure condition, but not the incentive condition	It satisfies the structure condition and the incentive condition	It satisfies the incentive condition, but not the structure condition
Q_{ht}^{\max}	Q_2^{DR}	Q_1^{DR} 或 Q_2^{DR}	Q_1^{DR}

The formula can be expressed as:

$$Q_{ht}^{\max} = \max \{ Q_1^{DR}, Q_2^{DR} \} \quad (12)$$

The purpose of taking their maximum value is to transfer as many elastic loads as possible during the peak period of users under incentive conditions. The corresponding electricity price incentive:

$$\Delta_{fg} = \lambda_{fg} / K_{fg} + a_{fg} \quad (13)$$

Where: Δ_{fg} is the peak - valley price difference.

3.3. Market Electricity Purchase Costs

If it buys a lot of electricity directly from the power generation enterprise, the selling company gets a lower unit price. The company will also receive certain fund subsidies from state policies in response to the national call for "energy conservation and emission reduction".

Therefore, the cost of electricity purchase in the selling company market can be expressed as:

$$\begin{aligned} R^{CB} &= R^{PW} + R^{BT} \\ &= \sum_t \lambda_t^{PW} Q_t^{PW} + \sum_t \lambda_t^{BT} Q_t^{BT} \end{aligned} \quad (14)$$

Where: R^{PW} and R^{BT} are respectively the electricity purchase fees from the power generation enterprise and subsidy fees. λ_t^{PW} and Q_t^{PW} respectively the unit price of electricity and the purchase quantity. λ_t^{BT} and Q_t^{BT} respectively refer to the subsidized unit price and subsidized power quantity.

4. Multi-objective Optimization Model

4.1. Multi-objective Function

From the perspective of selling company, promote the implementation of active demand response purpose is to seek to maximize the profit of the enterprise itself, in market competition to use price leverage and incentive policy of demand response. From the perspective of power users, change the original electricity consumption mode to get lower electricity costs. Therefore, the multi-objective function of the contract price model can be expressed as:

$$\begin{cases} \max \{ \sum_t^{24} P_t^{DR} E_t^{DR} - R^{PW} + R^{BT} \} \\ \min \{ \sum_t^{24} P_t^o Q_t^o - \sum_t^{24} P_t^{DR} E_t^{DR} \} \end{cases} \quad (15)$$

Where: P_t^{DR} and E_t^{DR} are the contract price and electricity consumption. R^{PW} is the total cost of purchasing power from the distribution network. R^{BT} is Subsidies for obtaining state policies. P_t^o and P_t^{DR} are the electricity price before and after DR. Q_t^o and E_t^{DR} are the power consumption before and after DR. In order

to facilitate calculation and comparison, it can be assumed that the total electricity consumption before and after the user participates in the demand response project is basically unchanged.

4.2. The Constraint

1.Total electricity consumption remains constant

In order to facilitate calculation and comparison, it can be assumed that the total electricity consumption before and after the user participates in the demand response project is basically unchanged:

$$\sum_t Q_t = \sum_t Q_t^{DR} \tag{16}$$

2.Contract threshold constraint

Energy threshold should be restricted within the contract price model, in line with the implementation of the contract price model under active demand response in actual situation.

$$0 \leq Q_{ht}^{min} \leq Q_{ht}^{max} \leq \sum_t Q_t$$

$$\begin{cases} Q_{ht}^{max} = \max\{Q_1^{DR}, Q_2^{DR}\} \\ Q_{ht}^{min} = Q_g^{min} \end{cases} \tag{17}$$

3.Price incentive constraints

As is known to 3.2, the implementation of the contract price model requires sufficient incentive level of electricity price. Therefore, the electricity price difference between peak and valley needs to meet the following constraints:

$$\Delta_{fg} = \lambda_{fg} / K_{fg} + a_{fg} \tag{18}$$

4.It is also necessary to ensure that after the implementation of the contract price model, the electricity charges of users will not increase, which is one of the principles of demand respond. And the DR revenue of the implementing company shall not be less than its market purchasing cost.

4.3. Model Solution Flow Chart

Using the method and model proposed in this paper, the multi-objective optimization function based on NSGA-II was solved by using MATLAB R2014a, and its model solution flow chart is as follows.

4.4. The Example Data

In this paper, the load data of 200 residents in a typical day in December 2006 were used as an example to simulate. Before the implementation of the contract price model, the unified price was 0.537 yuan/kw.h. The online price is 0.4 yuan/kw.h. The initial value $Q_{ht}^{max} = Q_1^{DR} = 106.27 * 85\% = 90.32$ kw.h, Q_{ht}^{min} and $= 25.11$ kw.h. At this point, the corresponding incentive condition electricity price difference is $\Delta_{fg} = 0.67$. In order to obtain

better test results, the adjustment coefficient was determined to be $S_{1,ht} = 0.55$, $S_{2,ht} = 5.50$ after multiple tests.

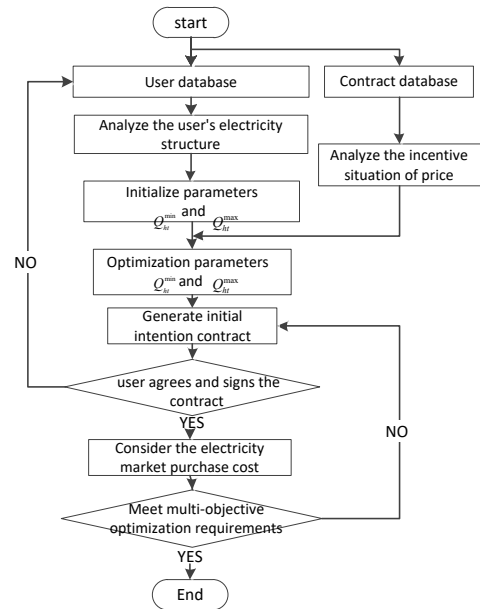


Figure 3 Flow chart of contract price model

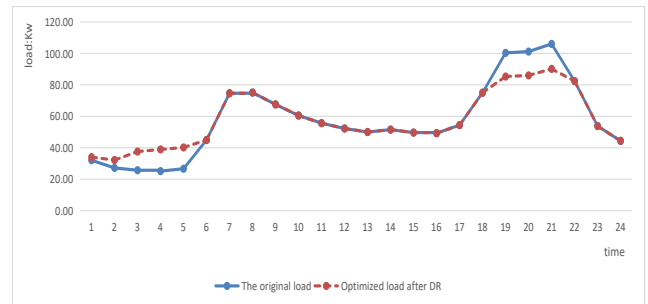


Figure 4 Variation curve of load after DR

5. Case Analysis

5.1. Multi-objective Optimization Results and Analysis

Figure 4 shows a load change after the DR. Obviously, after the contract price model, users' peak-load is down and valley-load is increased. Thus it can be seen that the implementation of the contract price model does has good effect.

Table 2 changes in economic benefits of users

object	state	fee	benefit
users	before DR	743.92	reduction of charges 97.11yuan
	after DR	646.81	

Table 3 changes in economic benefits of company

object	state	cost	fee	profits	benefit
company	before DR	554.13	743.92	189.79	Increase profits 33.59yuan
	after DR	423.43	646.81	223.38	

From table 2 and table 3 can see users and selling company's economic benefit is improved before DR. Users' electric fee decrease after the DR model, which has a good incentive to users. After participating in DR, the profit of the selling company has increased, which satisfies the principle of long-term good and healthy operation of the selling company.

5.2. Taking into Account the Analysis of User Participation

This section continue to delve into in the contract price model after implementing, when users adopt different ways of load transfer, the power system load curve and the user itself and selling company caused by the comprehensive benefit of different effects.

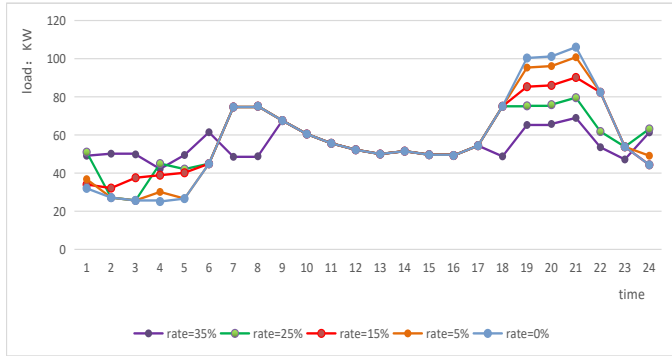


Figure 5 Load change curve under different consumption modes

By comparing the load curve under different power consumption modes in FIG.5, it is obvious that the higher the load transfer rate of the user is, the smoother the load level curve is, the more obvious the peak adjustment effect will be. It is also proved that the contract price model under the active demand response can achieve the effect of "cutting peak and filling valley" and optimizing energy allocation.

Further explore the economic benefits of users and selling company under different power consumption modes, as shown in table 4.8 and 4.9.

Table 4 Benefits of users under different modes

Content	After the implementation of DR model					Original
	One	Two	Three	Four	Five	
Mode	One	Two	Three	Four	Five	empty
Rate	35%	25%	15%	5%	0%	0
Charge	567.4	624.6	656.9	672.0	680.0	743.91
DR	0	0	0	117.5	200.7	empty
Total	567.4	624.6	656.9	789.6	880.7	743.91
Benefit	176.5	119.2	86.92	-45.72	-	0

As can be seen from the above table, when users adopt higher load transfer rate, the benefits of power users become more obvious, and their electricity charges will be reduced. Therefore, it can be obtained that the contract price model can guide and assist users to participate in the actual implementation of DR. According to the contract content, the high fee of DR will urge users to abide by the contract, also can further develop users a sense of long-term good demand response.

Table 5 Benefits of company under different modes

Conten	After the implementation of DR model					Originala
	One	Two	Three	Four	Five	
Mode	One	Two	Three	Four	Five	empty
Rate	35%	25%	15%	5%	10%	0
Cost	423.4	423.4	423.4	423.4	423.4	554.13
Total	567.4	624.6	656.9	789.6	880.7	743.91
Profit	143.9	201.2	233.5	366.1	457.3	189.78

As can be seen from the above table, when users adopt different load transfer rates of electricity consumption, the selling company is in a profit state. When the electricity load transfer rate is lower, sell electricity profits will rise, this is because the default behavior of the users will lead to sell electricity market risk influence a company's the normal operation of the enterprise, so the selling company charge users excess DR fees to compensate. Suggesting that the contract price model under active demand response can effectively guarantee a healthy and orderly operation of electricity market, and the further deepen the reform of the electricity market.

6. Conclusion

In this paper, a mathematical model of contract price under active response demand is established. On the basis of 24 node time TOU, this model introduces the mechanism of contract price, and uses the genetic algorithm NSGA-II to solve the multi-objective optimization. The simulation results show that:

- (1) The model can effectively adjust the electricity consumption mode of users, promote the optimization of electric power market supply and demand balance, improve the stability of power grid operation, and the reasonable allocation of energy resources;
- (2) The model can reflect the interests allocation problem between users and sell electricity company, and further develop the users' initiative sense, embodies the superiority of the electricity market reform.

References

- [1] Yoo T H, Ko W, Rhee C H, et al. The incentive announcement effect of demand response on market power mitigation in the electricity market[J]. Renewable & Sustainable Energy Reviews, 2017, 76: 545-554.
- [2] Sharifi R, Anvari-Moghaddam A, Fathi S H, et al. Economic demand response model in liberalised electricity markets with respect to flexibility of consumers[J]. Iet Generation Transmission & Distribution, 2017, 11(17): 4291-4298.
- [3] Reynders G. Quantifying the impact of building design on the potential of structural thermal storage for active demand response in residential buildings[J], 2015.
- [4] Grifull S R, Welling U, Jacobsen R H. Multi-modal Building Energy Management System for Residential Demand Response[C]. Digital System Design, 2016: 252-259.
- [5] Saleh S A, Aldik A A, Castillo-Guerra E. Distributed energy storage unit-based active demand response for residential loads[C]. IEEE Industry Applications Society Meeting, 2017: 1-9.
- [6] Mokryani G. Active distribution networks planning with integration of demand response[J]. Solar Energy, 2015, 122(3): 1362-1370.
- [7] Yi D, Hui H, Lin Z, et al. Design of Business Model and Market Framework Oriented to Active Demand Response of Power Demand Side[J]. Automation of Electric Power Systems, 2017, 41(14): 2-9 and 189.

Machine Learning Applied to GRBAS Voice Quality Assessment

Zheng Xie^{*1}, Chaitanya Gadepalli², Farideh Jalalinajafabadi³, Barry M.G. Cheetham³, Jarrod J. Homer⁴

¹*School of Engineering, University of Central Lancashire, PR1 2HE, UK*

²*Department of ENT, Salford Royal Hospital Foundation Trust, Salford, M6 8HD, UK*

³*School of Computer Science, University of Manchester, M13 9PL, UK*

⁴*Consultant in Head and Neck Surgery at the Manchester Royal Infirmary, Manchester University Hospitals Foundation Trust, UK*

ARTICLE INFO

Article history:

Received: 24 August, 2018

Accepted: 16 November, 2018

Online: 01 December, 2018

Keywords:

Voice quality assessment
GRBAS, Consistency measures
Cohen Kappa, Fleiss Kappa
Intra-class correlation
Feature detection
Machine learning

ABSTRACT

Voice problems are routinely assessed in hospital voice clinics by speech and language therapists (SLTs) who are highly skilled in making audio-perceptual evaluations of voice quality. The evaluations are often presented numerically in the form of five-dimensional 'GRBAS' scores. Computerised voice quality assessment may be carried out using digital signal processing (DSP) techniques which process recorded segments of a patient's voice to measure certain acoustic features such as periodicity, jitter and shimmer. However, these acoustic features are often not obviously related to GRBAS scores that are widely recognised and understood by clinicians. This paper investigates the use of machine learning (ML) for mapping acoustic feature measurements to more familiar GRBAS scores. The training of the ML algorithms requires accurate and reliable GRBAS assessments of a representative set of voice recordings, together with corresponding acoustic feature measurements. Such 'reference' GRBAS assessments were obtained in this work by engaging a number of highly trained SLTs as raters to independently score each voice recording. Clearly, the consistency of the scoring is of interest, and it is possible to measure this consistency and take it into account when computing the reference scores, thus increasing their accuracy and reliability. The properties of well known techniques for the measurement of consistency, such as intra-class correlation (ICC) and the Cohen and Fleiss Kappas, are studied and compared for the purposes of this paper. Two basic ML techniques, i.e. K-nearest neighbour regression and multiple linear regression were evaluated for producing the required GRBAS scores by computer. Both were found to produce reasonable accuracy according to a repeated cross-validation test.

1. Introduction

Voice problems are a common reason for referrals by primary practices to ear, nose and throat (ENT) departments and voice clinics in hospitals. Such problems may result from voice-strain due to speaking or singing excessively or too loudly, vocal cord inflammation, side-effects of inhaled steroids as used to treat asthma, infections, trauma, neoplasm, neurological disease and many other causes. This paper is an extension of work on voice

quality assessment originally presented in the 10th CISP-BMEI, conference in Shanghai [1]. Speech and language therapists (SLTs) are commonly required to assess the nature of voice quality impairment in patients, by audio-perception. This requires the SLT, trained as a voice quality expert, to listen to and assess the patient's voice while it reproduces, or tries to reproduce, certain standardized vocal maneuvers. In Europe, voice quality assessments are often made according to the perception of five properties of the voice as proposed by Hirano [2]. The five properties are referred to by the acronym 'GRBAS' which stands

^{*}Corresponding Author: Zheng Xie, Email: zxie2@uclan.ac.uk

for 'grade', 'roughness', 'breathiness', 'asthenia' and 'strain'. Each GRBAS property is rated, or scored by assigning an integer 0, 1, 2 or 3. A score of 0 signifies no perceived loss of quality in that property, 1 signifies mild loss of quality, 2 signifies moderate loss and 3 signifies severe loss. The scoring may be considered categorical or ordinal. With categorical scoring the integers 0, 1, 2 and 3 are considered as labels. With ordinal scoring, the integers are considered as being numerical with magnitudes indicating the severity of the perceived quality loss.

Grade (G) quantifies the overall perception of voice quality which will be adversely affected by any abnormality. Roughness (R) measures the perceived effect of uncontrolled irregular variations in the fundamental-frequency and amplitude of vowel segments which should be strongly periodic. Breathiness (B) quantifies the level of sound that arises from turbulent air-flow passing through vocal cords when they are not completely closed. Asthenia (A) measures the perception of weakness or lack of energy in the voice. Strain (S) gives a measure of undue effort needed to produce speech when the speaker is unable to employ the vocal cords normally because of some impairment.

Voice quality evaluation by audio-perception is time-consuming and expensive in its reliance on highly trained SLTs [3]. Also, inter-rater inconsistencies must be anticipated, and have been observed [4] in the audio-perceptual scoring of groups of patients, or their recorded voices, by different clinicians. Intra-rater inconsistencies have also been observed when the same clinician re-assesses the same voice recordings on a subsequent occasion. A lack of consistency in GRBAS assessments can adversely affect the appropriateness of treatment offered to patients, and the monitoring of its effect. A computerised approach to GRBAS assessment could eliminate these inconsistencies.

According to Webb et al. [5], GRBAS is simpler and more reliable than many other perceptual voice evaluation scales, such as Vocal Profile Analysis (VPA) [6] and the 'Buffalo Voice Profile' (BVP) [7], scheme. The 'Consensus Auditory-Perceptual Evaluation of Voice' (CAPE-V) approach, as widely used in North America [8], allows perceptual assessments of overall severity, roughness, breathiness, strain, pitch and loudness to be expressed as percentage scores. It is argued [8] that, compared with GRBAS, the CAPE-V scale better measures the quality of the voice and other aphonic characteristics. Also, CAPE-V assessments are made on a more refined scale. However, GRBAS is widely adopted [9] by practising UK voice clinicians as a basic standard.

No definitive solutions yet exist for performing GRBAS assessments by computer. Some approaches succeeded in establishing reasonable correlation between computerised measurements of acoustic voice features and GRBAS scores, but have not progressed to prototype systems [12]. Viable systems have been proposed, for example [13], but problems of training the required machine learning algorithms remain to be solved. The 'Multi-Dimensional Voice Program' (MDVP) and 'Analysis of Dysphonia in Speech and Voice' (ADSV) are commercial software packages [10] providing a wide range of facilities for acoustic feature analysis. Additionally, ADSV gives an overall assessment of voice dysphonia referred to as the Cepstral/spectral Index of Dysphonia (CSID) [11]. This is calculated from a multiple

regression based on the correlation of results from ADSV analyses with CAPE-V perceptual analyses by trained scorers. The CAPE-V overall measure of dysphonia is closely related to the 'Grade' component of GRBAS, therefore the CSID approach offers a methodology and partial solution to the GRBAS prediction problem. However the commercial nature of the CSID software makes it difficult to study and build on this methodology. Therefore, this paper considers how the results of a GRBAS scoring exercise may be used to produce a set of reference scores for training machine learning algorithms for computerised GRBAS assessment.

For the purposes of this research, a scoring exercise was carried out with the participation of five expert SLT raters, all of whom were trained and experienced in GRBAS scoring and had been working in university teaching hospitals for more than five years. A database of voice recordings from 64 patients was accumulated over a period of about three months by randomly sampling the attendance at a typical voice clinic. This database was augmented by recordings obtained from 38 other volunteers.

The recordings were made in a quiet studio at the Manchester Royal Infirmary (MRI) Hospital. Ethical approval was given by the National Ethics Research committee (09/H1010/65). The KayPentax 4500 CSL ® system and a Shure SM48 ® microphone were used to record the voices with a microphone set at 45 degrees at a distance of 4 cm. The recordings were of sustained vowel sounds and segments of connected speech.

To obtain the required GRBAS scores for each of the subjects (patients and other volunteers), the GRBAS properties of the recordings were assessed independently by the five expert SLT raters with the aid of a 'GRBAS Presentation and Scoring Package (GPSP)' [14]. This application plays out the recorded sound and prompts the rater to enter GRBAS scores. Raters used Sennheiser HD205 ® head-phones to listen to the recorded voice samples. The voice samples are presented in randomised order with a percentage (about 20 %) of randomly selected recordings repeated without warning, as a means of allowing the self-consistency of each rater to be estimated.

Different statistical methods were then employed to measure the intra-rater consistency (self-consistency) and inter-rater consistency of the scoring. Some details of these methods are presented in the next section. The derivation of 'reference' GRBAS scores from the audio-perceptual rater scores is then considered for the purpose of training ML algorithms for computerised GRBAS scoring. The derivation takes into account the inter-rater and intra-rater consistencies of each rater,

Voice quality assessment may be computerised using digital signal processing (DSP) techniques which analyse recorded segments of voice to quantify universally recognised acoustic features such as fundamental frequency, shimmer, jitter and harmonic-to-noise ratio [14]. Such acoustic features are not obviously related to the GRBAS measurements that are widely recognised and understood by clinicians. We therefore investigated the use of machine learning (ML) for mapping these feature measurements to the more familiar GRBAS assessments. Our approach was to derive 'reference scores' for a database of voice recordings from the scores given by expert SLT raters. The reference scores are then used to train a machine-learning

algorithm to predict GRBAS scores from the acoustic feature measurements resulting from the DSP analysis. The effectiveness of these techniques for computerised GRBAS scoring is investigated in Section 13 of this paper.

2. Measurement of Consistency

The properties of a number of well-known statistical methods for measuring rater consistency were considered for this research. The degree of consistency between two raters when they numerically appraise the same phenomena may be measured by a form of correlation. Perhaps the best known form of correlation is Pearson Correlation [15]. However, this measure takes into account only variations about the individual mean score for each rater [16]. Therefore a rater with consistently larger scores than those of another rater can appear perfectly correlated and therefore consistent with that other rater. Pearson correlation has been termed a measure of ‘reliability’ [17] rather than consistency. It is applicable only to ordinal appraisals, and is generally inappropriate for measuring consistency between or among raters [9] where consistency implies agreement. The notion of consistency between two raters can be extended to self-consistency between repeated appraisals of the same phenomena by the same rater (test-retest consistency), and to multi-rater consistency among more than two raters.

An alternative form of correlation is given by the ‘intra-class correlation’ coefficient (ICC) [18] and this may be used successfully as a measure of consistency for rater-pairs. It is also suitable for intra-rater (test-retest) and multi-rater consistency. The scoring must be ordinal. ICC is based on the differences that exist between the scores of each rater and a ‘pooled’ arithmetic mean score that is computed over all the scores given by all the raters. Therefore ICC eliminates the disadvantage of Pearson Correlation that it takes into account only variations about the individual mean score for each rater.

The ‘proportion of agreement’ (P_o), for two raters, is a simple measure of their consistency. It is derived by counting the number of times that the scores agree and dividing by the number, N , of subjects. P_o will always be a number between 0 (signifying no agreement at all) and 1 (for complete agreement). It is primarily for categorical scoring but may also be applied to ordinal scoring where the numerical scores are considered as labels. For ordinal scoring, P_o does not reflect the magnitudes of any differences, and in both cases, P_o is biased by the possibility of agreement by chance. The expectation of P_o will not be zero for purely random scores because some of the scores will inevitably turn out to be equal by chance. With Q different categories or scores evenly distributed over the Q possibilities, the probability of scores being equal by chance would be $1/Q$. Therefore, the expectation of P_o would be $1/Q$ rather than zero for purely random scoring. With $Q = 4$, this expectation would be a bias of 0.25 in the value of P_o . The bias could be even greater with an uneven spread of scores by either rater. The bias may give a false impression of some consistency when there is none, as could occur when the scores are randomly generated without reference to the subjects at all.

The Cohen Kappa is a well known consistency measure originally defined [19] for categorical scoring by two raters. It was later generalised to the weighted Cohen Kappa [20] which is applicable to ordinal (numerical) scoring with the magnitudes of

any disagreements between scores taken into account. The Fleiss Kappa [21] is a slightly different measure of consistency for categorical scoring that may be applied to two or more raters. The significance of Kappa and ICC measurements is often summarised by descriptions [22, 23] that are reproduced in Tables 1 and 2. A corresponding table for the Pearson correlation coefficient may be found in the literature [24].

Table 1: Significance of Kappa Values

Kappa	Consistency
1.0	Perfect
0.8 – 1.0	Almost perfect
0.6 - 0.8	Substantial
0.4 - 0.6	Moderate
0.2 - 0.4	Fair
0 - 0.2	Slight
< 0	Less than chance

Table 2: Significance of ICC Values

ICC	Consistency
0.75 – 1.0	Excellent
0.4 - 0.75	Fair
< 0.4	Poor

3. The Cohen Kappa

The original Cohen Kappa [19] for two raters, A and B say, was defined as follows:

$$Kappa = \frac{P_o - P_e}{1 - P_e} \tag{1}$$

where P_o is the proportion of agreement, as defined above, and P_e is an estimate of the probability of agreement by chance when scores by two raters are random (unrelated to the patients) but distributed across the range of possible scores identically to the actual scores of raters A and B . The estimate P_e is computed as the proportion of subject pairs (i, j) for which the score given by rater A to subject i is equal to the score given by rater B to subject j . This is an estimate of the probability that a randomly chosen ordered pair of subjects (i, j) will have equal scores.

This measure of consistency [19] is primarily for categorical scoring, though it can be applied to ordinal scores considered as labels. In this case, any difference between two scores will be considered equally significant, regardless of its numerical value. Therefore, it will only be of interest whether the scores, or classifications, are the same or different.

The weighted Cohen Kappa [20] measures the consistency of ordinal scoring where numerical differences between scores are considered important. It calculates a ‘cost’ for each actual disagreement and also for each expected ‘by chance’ disagreement. The cost is weighted according to the magnitude of the difference between the unequal scores. To achieve this, equation (1) is re-expressed by equation (2):

$$Kappa = 1 - \frac{1 - P_o}{1 - P_e} = 1 - \frac{D_o}{D_e} \tag{2}$$

where $D_o = 1 - P_o$ is the proportion of actual scores that are not equal and is considered to be the accumulated cost of the disagreements. The quantity $D_e = 1 - P_e$ is now considered to be the accumulated cost of disagreements expected to occur 'by chance' with random scoring distributed identically to the actual scores. Weighting is introduced by expressing D_o and D_e in the form of equations (3) and (4):

$$D_o = \frac{1}{N} \sum_{i=1}^N C(A(i), B(i)) \tag{3}$$

$$D_e = \frac{1}{N^2} \sum_{i=1}^Q \sum_{j=1}^Q A_i B_j C(\alpha(i), \alpha(j)) \tag{4}$$

In equations (3) and (4), $C(a,b)$ is the cost of any difference between scores (or categories) a and b . In equation (4), A_i denotes the number of subjects that rater A scores as $\alpha(i)$ and B_j denotes the number of subjects that rater B scores as $\alpha(j)$. Q is the number of possible scores or scoring categories and these are denoted by $\alpha(1), \alpha(2) \dots \alpha(Q)$. If the cost-function C is defined by equation (5):

$$C(a,b) = \begin{cases} 1: a \neq b \\ 0: a = b \end{cases} \tag{5}$$

the weighted Cohen Kappa [20] becomes identical to the original Cohen Kappa [19] also referred to as the unweighted Cohen Kappa (*UwCK*). If C is defined by equation (6),

$$C(a,b) = |a - b| \tag{6}$$

we obtain the 'linearly weighted Cohen Kappa' (*LwCK*), and defining C by equation (7) produces the 'quadratically weighted Cohen Kappa' (*QwCK*).

$$C(a,b) = (a - b)^2 \tag{7}$$

There are other cost-functions with interesting properties, but the three mentioned above are of special interest. For GRBAS scoring, there are $Q = 4$ possible scores which are $\alpha(1)=0, \alpha(2)=1, \alpha(3)=2$ and $\alpha(4)=3$.

Equation (4) may be re-expressed as equation (8):

$$D_e = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N C(A(i), B(j)) \tag{8}$$

Therefore, from equations (2), (3) and (8), we obtain equation (9) which is a general formula for all 2-rater (pair-wise) forms of Cohen Kappa:

$$Kappa = 1 - \frac{(1/N) \sum_{i=1}^N C(A(i), B(i))}{(1/N^2) \sum_{i=1}^N \sum_{j=1}^N C(A(i), B(j))} \tag{9}$$

The original and weighted Cohen Kappa [19, 20] are applicable when there are two individual raters, A and B say, who both score all the N subjects. The raters are 'fixed' in the sense that rater A is always the same clinician who sees all the subjects; and similarly for rater B. Therefore the individualities and prejudices of each rater can be taken into account when computing P_e , the probability of agreement by chance. For example, if one rater tends to give

scores that are consistently higher than those of the other rater, this bias will be reflected in the value of Cohen Kappa obtained.

4. Other Versions of Kappa

The Fleiss Kappa [21] measures the consistency of two or more categorical raters, and can therefore be a 'multi-rater' consistency measure. Further, the raters are not assumed to be 'fixed' since each subject may be scored by a different pair or set of raters. Therefore, it is no longer appropriate to take into account the different scoring preferences of each rater. If the Fleiss Kappa is used for a pair of fixed raters as for the Cohen Kappas, slightly different measurements of consistency will be obtained.

Assuming that there are n raters and Q scoring categories, Fleiss [21] calculates the proportion p_j of the N subjects that are assigned by raters to category j , as follows:

$$p_j = \frac{1}{N \times n} \sum_{i=1}^N n_{ij} \tag{10}$$

for $j = 1, 2, \dots, Q$, where n_{ij} is the number of raters who score subject i as being in category j . The proportion, P_i , of rater-pairs who agree in their scoring of subject i is given by:

$$P_i = \frac{1}{L} \sum_{j=1}^Q n_{ij} \times (n_{ij} - 1) / 2 \tag{11}$$

where L is the number of different rater-pairs that are possible, i.e. $L = n(n-1)/2$. The proportion of rater-pairs that agree in their assignments, taking into account all raters and all subjects, is now:

$$P_o = \frac{1}{N} \sum_{i=1}^N P_i \tag{12}$$

Fleiss [21] then estimates the probability of agreement 'by chance' as:

$$P_e = \sum_{j=1}^Q p_j^2 \tag{13}$$

Substituting from equations (12) and (13) into the Kappa equation (1) gives the Fleiss Kappa [21] which may be evaluated for two or more raters not assumed to be 'fixed' raters. The resulting equation does not generalise the original Cohen Kappa because equation (13) does not take any account of how the scores by each individual rater are distributed. P_e is now dependent only on the overall distribution of scores taking all raters together. Agreement by chance is therefore redefined for the Fleiss Kappa.

The original Cohen Kappa may be truly generalised [27] to measure the multi-rater consistency of categorical scoring by a group of n 'fixed' raters, where $n \geq 2$. Light [28] and Hubert [29] published different versions for categorical scoring, and Conger [30] extended the version by Light [28] to more than three raters. The generalisation by Hubert [29] redefines D_o and D_e to include all possible rater-pairs as in equation (14):

$$D_o = \frac{1}{L} \sum_{r=1}^n \sum_{s=r+1}^n D_o(r, s) \quad D_e = \frac{1}{L} \sum_{r=1}^n \sum_{s=r+1}^n D_e(r, s) \tag{14}$$

where the expression for $D_o(r,s)$ generalises equation (3) and the expression for $D_e(r,s)$ generalises equation (8) to become the cost of actual disagreement and the expected cost of by chance disagreement between raters r and s . Denoting by $A(i, r)$ the score given by rater r to subject i , we obtain equations (15) and (16):

$$D_o(r,s) = \frac{1}{N} \sum_{i=1}^N C(A(i,r), A(i,s)) \tag{15}$$

$$D_e(r,s) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N C(A(i,r), A(j,s)) \tag{16}$$

where equation (5) defines the cost-function C . Substituting for D_o and D_e from equation (14) into equation (2), with $D_o(r,s)$ and $D_e(r,s)$ defined by equations (15) and (16) gives a formula for the multi-rater Cohen Kappa that is functionally equivalent to that published by Hubert [29]. With C defined by equation (5) it remains unweighted.

The generalisation by Light [28] is different from the Hubert version when $n > 2$. It is given by equation (17):

$$UwCK = 1 - \frac{1}{L} \sum_{r=1}^n \sum_{s=r+1}^n \frac{D_o(r,s)}{D_e(r,s)} \tag{17}$$

Although both generalisations were defined for categorical scoring, they may now be further generalised to weighted ordinal scoring simply by redefining the cost-function C , for example by equation (6) for linear weighting or equation (7) for quadratic weighting. With $n = 2$, both generalisations are identical to the original [19] or weighted [20] Cohen Kappa.

5. Weighted Fleiss Kappa

As explained in [31], the original Fleiss Kappa [21] is given by equation (18) when the cost-function C is as in equation (5).

$$FK = 1 - \frac{\frac{1}{NL} \sum_{i=1}^N \sum_{r=1}^n \sum_{s=r+1}^n C(A(i,r), A(i,s))}{\frac{1}{(Nn)^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{r=1}^n \sum_{s=1}^n C(A(i,r), A(j,s))} \tag{18}$$

The Fleiss Kappa may be generalised to a weighted version for ordinal scoring by redefining cost-function C as for the multi-rater Cohen Kappa. In all cases, the unweighted or weighted Fleiss Kappa is applicable to measuring the consistency of any number of raters including two.

6. Intra-Class Correlation Coefficient (ICC)

In its original form [25], ICC is defined for n raters as follows:

$$ICC = \frac{(1/L) \sum_{i=1}^N \sum_{r=1}^n \sum_{s=r+1}^n (A(i,r) - m)(A(i,s) - m)}{(1/n) \sum_{i=1}^N \sum_{r=1}^n (A(i,r) - m)^2} \tag{19}$$

$$\text{where } m = \frac{1}{nN} \sum_{i=1}^N \sum_{r=1}^n A(i,r) \tag{20}$$

Other versions of ICC have also been proposed [26]. It is known [26] that, for two raters, ICC will be close to quadratically weighted Cohen Kappa ($QwCK$) when the individual mean score for each rater is approximately the same. This property is observed [31] also for multi-rater versions of ICC and $QwCK$. More interestingly, it has been shown [31] that ICC is always exactly equal to quadratically weighted Fleiss Kappa ($QwFK$) regardless of the number of raters and their individual mean scores.

7. Intra-rater Consistency

For the GRBAS rating exercise referred to in Section 1, intra-rater (test-retest) scoring differences were generally small due to the experience and high expertise of the SLT raters. There were some differences of 1, very occasional differences of 2, and no greater differences. The test-retest consistency for the five GRBAS components was measured for all five raters, by unweighted, linearly weighted and quadratically weighted Cohen Kappa ($UwCK$, $LwCK$ and $QwCK$) and ICC . By averaging $UwCK$, $LwCK$ and pair-wise ICC measurements over the five GRBAS components we obtained Table 3 which gives three overall measurements of the test-retest consistency of each rater. $QwCK$ gave a close approximation to ICC , and is not shown in the table. $QwFK$, also not shown, was indistinguishable from ICC . For all forms of Kappa, the P_o and P_e terms were averaged separately. Similarly, the ICC numerators and denominators were averaged separately.

With $UwCK$, any difference in scores incurs the same cost regardless of its magnitude. Small differences cost the same as large differences. This makes $UwCK$ pessimistic for highly consistent raters where most test-retest discrepancies are small. Therefore, the averaged $UwCK$ consistency measurements in Table 3 are pessimistic for our rating exercise.

With $QwCK$, the largest differences in scores incur very high cost due to the quadratic weighting. With ICC , the costs are similar. These high costs are important even when there are few or no large scoring differences because they strongly affect the costs of differences expected to incur 'by chance'. These high 'by chance' costs make both $QwCK$ and ICC optimistic, when compared with $LwCK$, for highly consistent rating with a fairly even distribution of scores. We therefore concluded that $LwCK$ gives the most indicative measure of test-retest consistency for the rating exercise referred to in this paper. A different set of scores may have led to a different conclusion. In Table 3, it may be seen that the self-consistency of raters 1 to 4, as measured by $LwCK$, was considered 'substantial' according to Table 2. The self-consistency of rater 5 was considered 'moderate'. Conclusions can therefore be drawn about the self-consistency of each rater and how this may be expected to vary from rater to rater.

Table 3: Intra-Rater Consistency Averaged over all GRBAS Components

Rater	$UwCK$	$LwCK$	ICC	Consistency ($LwCK$)	Consistency (ICC)
1	0.72	0.77	0.84	Substantial	Excellent
2	0.65	0.76	0.85	Substantial	Excellent
3	0.53	0.64	0.75	Substantial	Excellent
4	0.68	0.73	0.77	Substantial	Excellent
5	0.44	0.60	0.74	Moderate	Fair

Table 4: Intra-rater Consistency Averaged over all 5 raters

Component	UwCK	LwCK	ICC	Consistency (LwCK)	Consistency (ICC)
G	0.64	0.77	0.87	Substantial	Excellent
R	0.57	0.67	0.76	Substantial	Excellent
B	0.55	0.66	0.76	Substantial	Excellent
A	0.68	0.73	0.80	Substantial	Excellent
S	0.59	0.68	0.76	Substantial	Excellent

Table 4 shows *UwCK*, *LwCK* and *ICC* intra-rater consistency measurements for G, R, B, A, and S, averaged over all five raters. According to all the measurements, it appears that test-retest consistency with R, B and S is more difficult to achieve than with G and A.

8. Inter-rater Consistency

Measurements of inter-rater consistency between pairs of raters for any GRBAS component may be obtained using the same forms of Kappa and ICC as were used for intra-rater consistency. Our rating exercise had a group of five raters, therefore ten possible pairs. This means that there are ten pair-wise measurements of inter-rater consistency for each GRBAS component. To reduce the number of measurements, it is convenient to define an 'individualised' inter-rater measurement for each rater. For each GRBAS component, this individualised measurement quantifies the consistency of the rater with the other raters in the group. It is computed for each rater by averaging all the pair-wise inter-rater assessments which involve that rater. Thus an individualised measure of inter-rater consistency is obtained for G, R, B, A and S for each rater. With five raters, the 25 measurements can be reduced to five by averaging the individualised G, R, B, A and S measurements to obtain a single average measure for each rater.

The *UwCK*, *LwCK* and *ICC* individualised inter-rater measurements, averaged over all GRBAS components, are shown in Table 5 for raters 1 to 5. For all raters, the average consistency is 'moderate' according to *LwCK* and 'fair' according to *ICC*. Raters 1, 4 and 5 have almost the same inter-rater consistency, rater 2 has slightly lower consistency and rater 2 is the least consistent when compared with the other raters.

Table 5: Individualised Inter-rater Consistency averaged over all GRBAS Components

Rater	UwCK	LwCK	ICC	Consistency (LwCK)	Consistency (ICC)
1	0.47	0.59	0.70	Moderate	Fair
2	0.40	0.52	0.60	Moderate	Fair
3	0.45	0.57	0.67	Moderate	Fair
4	0.48	0.60	0.71	Moderate	Fair
5	0.47	0.59	0.70	Moderate	Fair

9. Multi-rater Consistency

The multi-rater consistency according to the unweighted Fleiss Kappa (FK), the generalised Cohen Kappa (with linear weighting) and *ICC*, computed for the group of five raters, are shown in Table 6 for each GRBAS component. The values of *UwCK* were indistinguishable from FK to the precision shown in the table. Similarly for the values of *QwCK* and *ICC*. Quadratically weighted *FK*, also not shown, would be exactly equal to *ICC*.

Table 6: Multi-rater Consistency by Fleiss Kappa, Cohen Kappa and *ICC*

	FK	LwCK	ICC	Consistency (LwCK)	Consistency (ICC)
G	0.56	0.71	0.83	Substantial	Excellent
R	0.44	0.57	0.68	Moderate	Fair
B	0.43	0.58	0.71	Moderate	Fair
A	0.38	0.46	0.55	Moderate	Fair
S	0.44	0.54	0.65	Moderate	Fair

In contrast to Table 5 which allows us to compare the overall consistency of raters, Table 6 allows us to compare the difficulty of achieving group consistency for each GRBAS component. It is clear that some GRBAS components are more difficult to score consistently than others. According to *ICC*, group consistency is 'excellent' for Grade and 'fair' for R, B, A and S. *LwCK* gives 'substantial' for Grade and 'fair' for the others. The *FK* and *UwCK* measurements are more pessimistic due to their assumption that the scoring is categorical. According to all measurements of multi-rater consistency, the consistency is highest for highest, followed by Breathiness, Roughness, Strain and Asthenia.

It should be mentioned that the classifications given by Tables 1 and 2 serve only as a rough guide to interpreting the values of Kappa and ICC obtained. However, they are widely used despite the fact that it seems inappropriate to use Table 2 for quadratically weighted Kappa in view of its closeness to ICC. In particular, the category 'Fair' in Tables 2 and 3 refers to quite different ranges which may be misleading if Table 2 were used for *QwCK*. Therefore, it is appropriate to refer to Table 3 for both *ICC* and quadratically weighted Kappa.

10. Reference GRBAS Scores

The feasibility of performing automatic GRBAS scoring by computer was investigated by training machine learning (ML) algorithms for mapping acoustic feature measurements to the familiar GRBAS scale. For the training, a set of accurate and reliable GRBAS scores was required for each of the *N* subjects in our database. We refer to these as 'reference' GRBAS scores. A technique for deriving these reference scores from the scores of a group of audio-perceptual raters, such as that described in Section 1, was therefore devised. The measurements of inter-rater and intra-rater consistency, obtained as described above, is taken into account as a means of optimising the accuracy and reliability of the reference scores.

Given the 'Grade' scores $A(i, r)$ for subject *i*, with rater-index *r* in the range 1 to 5, we first computed weighted average pair-wise scores $G_{rs}(i)$ by equation (21), for all possible rater-pairs (*r,s*). The weighting is by the *LwCK* intra-rater consistency measurements in Table 3 referred to now as w_1, w_2, w_3, w_4, w_5 for raters 1 to 5 respectively.

$$G_{rs}(i) = \frac{w_r A(i, r) + w_s A(i, s)}{w_r + w_s} \tag{21}$$

The 'Grade' reference score for subject *i* is then obtained as a weighted average of the $G_{rs}(i)$ values over all possible rater-pairs, i.e.:

$$G_{ref}(i) = \frac{1}{L} \sum_{r=1}^n \sum_{s=r+1}^n w(r,s) G_{rs}(i) \quad (22)$$

where $L = n(n-1)/2$ with $n=5$. The weights $w(r,s)$ are the pair-wise inter-rater $LwCK$ measurements for Grade. This procedure is performed for all subjects for Grade, and then repeated for the other GRBAS dimensions. The weighting de-emphasises scores from less self-consistent raters in favour of more self-consistent ones. It also de-emphasises the scores from raters who are less consistent with other raters.

11. Voice Quality Assessment by Computer

Considerable published research, including [12] and [13], has not yet established a definitive methodology for GRBAS assessment by computer. An overall CAPE-V assessment of dysphonia, CSID [11], available commercially, is strongly related to 'Grade', but it does not independently assess the other GRBAS and CAPE-V components [8]. Computerised voice quality assessment may be carried out using digital signal processing (DSP) to analyse segments of voice to produce mathematical functions such as the autocorrelation function, fast Fourier Transform and cepstrum. From such functions, acoustic features such as the aperiodicity index (*API*), fundamental frequency (F_0), harmonic-to-noise ratio (*HNR*), jitter, shimmer, cepstral peak prominence (*CPP*), low-to-high spectral ratio (*LH*) and others may be derived. However, these features are not obviously related to GRBAS assessments of voice quality.

Perceived voice quality is strongly dependent on the short term periodicity of the vowels and the nature of the fluctuations in this periodicity. To measure short-term periodicity, and how this varies over a spoken vowel, speech must be segmented into frames. The degree of periodicity of each of these frames may be expressed as an aperiodicity index (*API*) which is equal to $1 - p$ where p is the peak value of a suitable form of autocorrelation function. An *API* of zero indicates exact periodicity and its value increases towards 1 with increasing aperiodicity. The *API* is increased by additive noise due to 'breathiness', fundamental frequency or amplitude variation due to 'roughness' in the operation of the vocal cords, and other acoustic features.

A sustained vowel without obvious impairment will generally have strong short-term periodicity for the duration of the segment, though the fundamental frequency (F_0) and loudness may vary due to natural characteristics of the voice and controlled intonation. By monitoring how the degree of short term periodicity changes over a passage of natural connected speech, vowels may be differentiated from consonants, thus allowing the acoustic feature measurements to concentrate on the vowels.

Jitter is rapid and uncontrolled variation of F_0 and shimmer is rapid and uncontrolled variation of amplitude. Both these acoustic features can be indicative of roughness in GRBAS assessments. They will affect grade also. There are many ways of defining jitter and shimmer as provided by the Praat software package [32]. The *HNR* may be derived from the autocorrelation function and can be indicative of breathiness in GRBAS assessments since the 'noise' is often due to turbulent airflow. Low-to-high spectral ratio (*LH*) measurements are made by calculating and comparing, in the frequency-domain, the energy

below and above a certain cut-off frequency, such as 1.5 kHz or 4.0 kHz. The required filtering may be achieved either by digital filters or an FFT. A high value of *LH* with cut-off frequency 1.5 kHz can be indicative of asthenia [36] and strain [37] due to imperfectly functioning vocal cords damping the spectral energy of formants above 1.5 kHz. *LH* measurements with a cut-off frequency of 4.0 kHz are useful for detecting breathiness and voicing since the spectral energy of voiced speech (vowels) is mostly below 4.0 kHz. *CPP* is widely used as an alternative to *API* and *HNR* as a means of assessing the degree of short term periodicity.

As in [34], well known DSP techniques were employed [14, 35] to recognise vowels and measure the acoustic features mentioned above, and several others. Frame-to-frame variations in these features over time were also measured. Published DSP algorithms and commercial and academic computer software are available for making these measurements from digitised voice recordings [32, 33]. Twenty acoustic features were identified by Jalalinajafabadi [14] as being relevant to GRBAS scores. They were measured by a combination of DSP algorithms specially written in MATLAB and commercial software provided by MDVP and ADSV [10, 11]. For the MATLAB algorithms, the speech recordings were sampled at $F_s = 44.1$ kHz, and divided into sequences of 75% overlapping 23.22 ms frames of 1024 samples. MDVP and ADSV use a slightly different sampling rate and framing. Many of the features were strongly correlated and their usefulness was far from uniform. Therefore, some experiments with feature selection were performed. The usefulness of each possible sub-set of features for predicting each GRBAS component was estimated by a combination of correlation measurements, to reduce the dimensionality of the task, and then a form of direct search. The use of Principle Component Analysis' (PCA) would have reduced the computation, but this was not a critical factor.

Section 14 will evaluate the performance of MLR and KNNR (with and without feature selection) and perceptual analysis against the 'reference GRBAS scores'.

12. Machine Learning Algorithms

We analysed the recordings of sustained vowels obtained from the $N = 102$ subjects mentioned in Section 1. For each recording, acoustic feature measurements were obtained as explained in Section 11. A total of $m = 20$ feature measurements were obtained as detailed in [14]. An $N \times m$ matrix X of feature measurements was defined for each of the five GRBAS components. These matrices became the input to the machine learning (ML) algorithm along with the $N \times 1$ vector \underline{Y} of reference GRBAS scores derived as explained in Section 10. The ML algorithm was designed to learn to predict, as closely as possible, the reference GRBAS scores supplied for each subject. The prediction must be made from the information provided by the m acoustic feature measurements supplied for each voice segment. Two simple ML approaches were compared [14, 35]: K-nearest neighbour regression (KNNR) and multiple linear regression (MLR).

With KNNR, the ML information consists of a matrix X and vector \underline{Y} for each GRBAS component. Supplying the ML algorithm with these arrays is all that is required of the training process. K is an integer that defines the way the KNNR approach

predicts a score for a new subject from measurements of its m acoustic features. The prediction is based on the known scores for K other subjects chosen according to the ‘distance’ of their measured acoustic features from those of the new subject. The concept of distance can be defined in various ways such as the Euclidean distance which we adopted. The distance between the new subject and each of the N database subjects is calculated. Then K subjects are selected as being those that are nearest to the new subject according to their feature measurements. A simple form of KNNR takes the arithmetic mean of the scores of the K nearest neighbours as the result. A preferred alternative form takes a weighted average where the reference scores are weighted according to the proximity of the reference subject to the new subject.

A choice of K must be made, and this may be different for each GRBAS component. The optimal value of K will depend on the number, N , of subjects, the distribution of their scores and the number of acoustic features being taken into account. K is often set equal the square root of N , though investigations can reveal more appropriate values. In this work, Jalalinajafabadi [14] plotted the prediction error against K to obtain a suitable value of K for each GRBAS component. This was done after selecting the most appropriate set of acoustic features for each GRBAS component. The values of K producing the lowest prediction errors were $K=6$ for grade, $K=10$ for roughness, $K=5$ for breathiness and $K = 8$ for strain and asthenia.

The Multiple Linear Regression (MLR) approach computes, for each GRBAS component, a vector $\underline{\beta}$ of K regression coefficients such that

$$\underline{Y} = X.\underline{\beta} + \underline{\varepsilon} \tag{23}$$

where the error-vector $\underline{\varepsilon}$ is minimised in mean square value over all possible choices of $\underline{\beta}$ of dimension K . It may be shown [14] that the required vector $\underline{\beta}$ is given by:

$$\underline{\beta} = X^\#.\underline{Y} \tag{24}$$

where $X^\#$ is the pseudo-inverse of the non-square matrix X . For a subject whose m feature measurements \underline{x} have been obtained, the equation:

$$y = \underline{x}^T.\underline{\beta} \tag{25}$$

produces a scoring estimate y . This will be close to $Y(i)$ for each subject i in our database, and may be expected to produce reasonable GRBAS scores for an unknown subject.

13. Testing and Evaluation

The application developed by Jalalinajafabadi [14] made $m = 20$ voice feature measurements per subject. Feature selection was applied to identify which subset of these m features gave the best result for each GRBAS dimension. It was generally found that, compared with including all 20 feature measurements, better results were obtained with smaller subsets tailored to the GRBAS dimensions. Several computational methods for feature selection were compared [14] in terms of their effectiveness and computational requirements. The results presented here were obtained using a combination of correlation tables (between feature measurements and GRBAS components) and exhaustive

search. The best feature subsets for G, R, B, A and S are generally different, since different feature measurements highlight different aspects of the voice. It was found beneficial to normalise the feature measurements to avoid large magnitudes dominating the prediction process, especially for KNNR.

To evaluate the KNNR and MLR algorithms for mapping acoustic feature measurements to GRBAS scores, 80 subjects were randomly selected for training purposes from the 102 available subjects. The remaining 22 subjects were set aside to be used for testing the mapping algorithms once they had been trained. Twenty ‘trials’ were performed by repeating the training and testing, each time with a different randomisation. The same testing approach was used for both KNNR and MLR. The trained mapping algorithm was used to predict GRBAS scores for the 22 testing subjects from the corresponding acoustic feature measurements. The GRBAS scores thus obtained were compared with the known reference scores. For each trial, a value of ‘root mean squared error’ (RMSE) was computed for each GRBAS component over the 22 testing subjects. These RMSE values were then averaged over the 20 trials. An RMSE of 100% would correspond to an RMS error of 1 in the GRBAS scoring where the averaging is over all 22 testing subjects and all 20 trials.

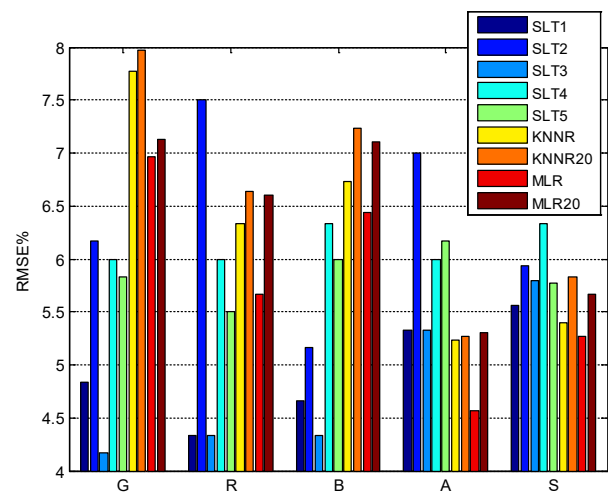


Figure 1: RMSE% for SLT 1-5, KNNR & MLR with feature selection and using all available 20 features (KNNR20 & MLR20).

A comparison of the GRBAS scoring produced by the five SLTs and the KNNR and MLR algorithms is presented in Figure 1. This graph summarises the results of experiments carried out by Jalalinajafabadi [14] with and without feature selection. Measurements obtained without feature selection are labelled KNNR20 and MLR20 since all available 20 features are taken into account. Comparing KNNR (with feature selection) and KNNR20, the feature selection has reduced the prediction error RMSE% by up to about 0.5%. Comparing MLR and MLR20, the reduction due to feature selection is generally greater, i.e. about 1% for Roughness and up to 0.7% for the other components (apart from Grade). With feature selection, the performances of the two machine learning techniques appear quite similar according to the RMSE measurements, though MLR is consistently better than KNNR. For Asthenia and Strain, both KNNR and MLR with feature selection deliver a lower RMSE than was obtained for each of the SLT raters with reference to the corresponding reference-

scores. For Grade, the KNNR and MLR values of RMSE (with feature selection) are both markedly higher than the corresponding values obtained for all the five SLT raters. The worst RMS difference for Grade is about 7.5%. The results for 'Breathiness' are close to those of the two worst performing SLT raters, and the MLR result for 'Roughness' lies between the two best and two worst performing SLT raters. As reported by Jalalinajafabadi [14] and further explained in [1], the RMSE taken over all GRBAS components was found to be marginally lower for KNNR and MLR (both with feature selection) than for each of the five individual SLT raters.

14. Conclusions

Recordings of normal and impaired voices were obtained from randomly selected patients and some other volunteers. These recordings were audio-perceptually assessed by five expert GRBAS raters to obtain a set of GRBAS scores for each recording. Statistical methods for measuring the inter-rater and intra-rater consistency of the scoring were investigated and it was concluded that the linearly weighted Cohen Kappa ($LwCK$) was suitable for this purpose. The measurements suggested that the GRBAS assessments were reasonably consistent. The scores and $LwCK$ consistency measurements were then used to produce a set of 'reference scores' for training machine learning algorithms for mapping acoustic feature measurements to GRBAS scores, and thus performing automatic GRBAS scoring. With the reference scores, and acoustic feature measurements extracted from each of the 102 speech recordings by standard DSP techniques, KNNR and MLR were found to produce comparable automatic GRBAS scoring performances which compared favourably with the scoring by the five SLT raters. Feature selection was applied to determine the best subset of the twenty available acoustic features for each GRBAS dimension.

Conflict of Interest

The authors declare no conflicts of interest.

Acknowledgment

The authors acknowledge the contributions of Ms Frances Ascott, the SLT raters and the participants. We also acknowledge considerable help and advice from Prof. Gavin Brown and Prof. Mikel Lujan in the Computer Science School of Manchester University, UK.

References

- [1] Z. Xie, C. Gadepalli, F. Jalalinajafabadi, B.M.G. Cheetham, J.J. Homer, "Measurement of Rater Consistency and its Application in Voice Quality Assessments", 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Shanghai, China, October, 2017.
- [2] M. Hirano, "Clinical Examination of Voice", New York: Springer, 1981.
- [3] M.S. De Bodt, F.L. Wuyts, P.H. Van de Heyning, C. Croux, "Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality", *Journal of Voice*, 1997, 11(1):74-80.
- [4] C. Sellars, A.E. Stanton, A. McConnachie, C.P. Dunnet, L.M. Chapman, C.E. Bucknall, et al., "Reliability of Perceptions of Voice Quality: evidence from a problem asthma clinic population", *J. Laryngol Otol.*, 2009, pp. 1-9.
- [5] A.L. Webb, P.N. Carding, I.J. Deary, K. MacKenzie, N. Steen & J.A. Wilson, "The reliability of three perceptual evaluation scales for dysphonia", *Eur Arch Otorhinolaryngol*; 261(8):429-34, 2004.
- [6] J. Laver, S. Wirz, J. Mackenzie & S. Hiller, "A perceptual protocol for the analysis of vocal profiles", *Edinburgh University Department of Linguistics Work in Progress*; 14:139-55. 1981.
- [7] D. K. Wilson, "Children's voice problems", *Voice Problems of Children*, 3rd ed., Williams and Wilkins, Philadelphia, PA., 1-15, 1987.
- [8] G.B. Kempster, B.R. Gerratt, K.V. Abbott, J. Barkmeier-Kraemer & R.E. Hillman, "Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol", *American Journal of Speech-Language Pathology*; 18(2):124-32, 2009.
- [9] P. Carding, E. Carlson, R. Epstein, L. Mathieson & C. Shewell, "Formal perceptual evaluation of voice quality in the United Kingdom", *Logopedics Phoniatrics Vocology*, 25(3):133-8, 2000.
- [10] S.N. Awan, and N. Roy, "Toward the development of an objective index of dysphonia severity: a four-factor acoustic model", *Clinical linguistics & phonetics*, 20(1):35-49, 2006.
- [11] S.N. Awan, N. Roy, M.E. Jette, G.S. Meltzner and R.E. Hillman, "Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: comparisons with auditory-perceptual judgements from the CAPE-V", *Clinical Linguistics & Phonetics*, 24(9):742-758, 2010.
- [12] T. Bhuta, L. Patrick & J.D. Garnett, "Perceptual evaluation of voice quality and its correlation with acoustic measurements", *Journal of Voice*, Elsevier, 2004, Vol.18, Issue.3, pp. 299-304.
- [13] F. Villa-Canas, J.R. Orozco-Arroyave, J.D. Arias-Londono et al., "Automatic assessment of voice signals according to the GRBAS scale using modulation spectra, MEL frequency cepstral coefficients and noise parameters", *IEEE Symposium of Image, Signal Processing, and Artificial Vision (STSIVA)*, 2013, pp. 1-5.
- [14] F. Jalalinajafabadi, "Computerised assessment of voice quality", PhD Thesis. 2016, University of Manchester, UK.
- [15] J. Lee Rodgers & W.A. Nicewander, "Thirteen ways to look at the correlation coefficient", *The American Statistician*, 1988, vol. 42(1), pp. 59-66.
- [16] J.F. Bland and D.G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement", *The Lancet*, 1986, 327(8476), pp. 307-310.
- [17] J. Krieman, B.R. Gerratt, G.B. Kempster, A. Erman & G.S. Berke, "Perceptual Evaluation of Voice Quality: Review, Tutorial and a Framework for Future Research", *Journal of Speech and Hearing Research*, Vol. 36, 21-40, 1993, pp 21-40.
- [18] G.G. Koch, "Intraclass correlation coefficient", *Encyclopedia of statistical sciences*, 1982.
- [19] J. Cohen, "A coefficient of agreement for nominal scales", *Educational and Psychosocial Measurement*, 1960, 20, pp. 37-46.
- [20] J. Cohen, "Weighted Kappa: Nominal scale agreement provision for scaled disagreement or partial credit", *Psychological bulletin*, 1968, vol. 70(4), p. 213.
- [21] J.L. Fleiss, "Measuring nominal scale agreement among many raters", *Psychological bulletin*, 1971, vol. 76 no 5, pp. 378-382.
- [22] A.J. Viera and J.M. Garrett, "Understanding inter-observer agreement: the Kappa statistic", *Fam Med*. 2005, vol. 37(5), pp. 360-3.
- [23] J.L. Fleiss, "Design and analysis of clinical experiments", Vol. 73. John Wiley & Sons, 2011.
- [24] J.D. Evans, "Straightforward Statistics for the Behavioral Sciences", Brooks/Cole Publishing Company, 1996.
- [25] E. Rödel & R.A. Fisher, "Statistical Methods for Research Workers", 14. Aufl., Oliver & Boyd, Edinburgh, London 1970. XIII, 362 S., 12 Abb., 74 Tab., 40 s. *Biometrische Zeitschrift*, 1971, vol. 13(6), pp. 429-30.
- [26] J.L. Fleiss and J. Cohen, "The equivalence of weighted Kappa and the intra class correlation coefficient as measures of reliability", *Educational and Psychological Measurement*, 1973, vol. 33, pp. 613-619.
- [27] M.J. Warrens, "Inequalities between Multi-Rater Kappas", *Adv Data Classif*, 2010, vol. 4, pp. 271-286.
- [28] R.J. Light, "Measures of response agreement for qualitative data: some generalisations and alternatives", *Psychol Bull*, 1971, vol. 76, pp. 365-377.
- [29] L. Hubert, "Kappa Revisited", *Psychol Bull* 1977, vol. 84, pp. 289-297.
- [30] A.J. Conger, "Integration and Generalisation of Kappas for Multiple Raters", *Psychol Bull.*, 1980, vol. 88, pp. 322-328.
- [31] Z. Xie, C. Gadepalli, & B.M.G. Cheetham, "A study of chance-corrected agreement coefficients for the measurement of multi-rater consistency", *International Journal of Simulation: Systems, Science & Technology* 19(2), 2018, pp. 10.1-10.9.
- [32] P. Boersma & D. Weenink, "Praat: a system for doing phonetics by computer", *Glott International* (2001) 5:9/10, pp. 341-345.
- [33] O. Amir, M. Wolf & N. Amir, "A clinical comparison between two acoustic analysis softwares: MDVP and Praat", *Biomedical Signal Processing and Control*, 2009, vol.4(3), pp. 202-205.
- [34] S. Hadjitodorov & P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening", *Medical engineering & physics*, 2002, vol. 24(6), pp. 419-29.
- [35] F. Jalalinajafabadi, C. Gadepalli, F. Ascott, J.J. Homer, M. Luján & B.M.G. Cheetham, "Perceptual Evaluation of Voice Quality and its correlation with

acoustic measurement", IEEE European Modeling Symposium (EMS2015), Manchester, 2013, pp. 283-286.

- [36] F. Jalalinajafabadi, C. Gadepalli, M. Ghasempour, F. Ascott, J.J. Homer, M. Lujan & B.M.G. Cheetham, "Objective assessment of asthenia using energy and low-to-high spectral ratio", 2015 12th Int Joint Conf on e-Business and Telecommunications (ICETE), vol. 6, pp. 576-583, Colmar, France, 20-22 July, 2015.
- [37] F. Jalalinajafabadi, C. Gadepalli, M. Ghasempour, M. Lujan, B.M.G. Cheetham & J.J. Homer, "Computerised objective measurement of strain in voiced speech", 2015 37th Annual Int Conf of the IEEE Engineering in Medicine and Biology (EMBC), pp.5589-5592, Milan, Italy, 25-29 Aug 2015.

MRI images Enhancement and Brain Tumor Segmentation

Aye Min^{*1}, Zin Mar Kyu²

¹Digital Image Processing, University of Computer Studies (UCSM), 05013, Myanmar

²Software Department, University of Computer Studies (UCSM), 05013, Myanmar

ARTICLE INFO

Article history:

Received: 14 August, 2018

Accepted: 03 November, 2018

Online: 01 December, 2018

Keywords:

MRI

Fusion based results binding

Adaptive K-means clustering

ABSTRACT

Brain tumor is the abnormal growth of cancerous cells in Brain. The development of automated methods for segmenting brain tumors remains one of the most difficult tasks in medical data processing. Accurate segmentation can improve diagnosis, such as evaluating tumor volume. However, manual segmentation in magnetic resonance data is a laborious task. The main problem to detect brain tumors is less precise to determine the area of the tumor and determine the segmentation accuracy of the tumor. The system proposed the fusion based results binding for MRI image enhancement and combination of adaptive K-means clustering and morphological operation for tumor segmentation. BRATS multimodal images of brain tumor Segmentation Benchmark dataset was used in experiment testing.

1. Introduction

Primary diagnosis of brain tumors is extremely significant, because it can save lives. Accurate segmentation of brain tumors is also important, as it can help medical personnel in the planning of treatment and intervention. Manual segmentation of tumors requires a long period of time, even for a qualified specialist. Fully automated segmentation and quantitative analysis of tumors, therefore, are highly beneficial maintenance. However, it is also very difficult due to the large variety of anatomical structures and low contrast of current imaging techniques that create the distinction between normal and tumor regions. The main objective of our research is to create a trustworthy procedure detection of tumors of a multimodal MRI record based on a controlled machine study the methods using a data set containing MICCAI Brats images with ground truth, provided by human experts.

In this article we propose fusion-based results binding (result fusion) method for image enhancement and combination of adaptive k-means clustering and morphological operation for tumor segmentation and reliable detection system are proposed. In MRI image enhancement, the definition of results fusion method is fusion the filtered results of median filter and wiener filter. We emphasized on the results of median filter and wiener filter. Median filter made the original image to be more sharpening and wiener filter made the image to be more smoothness. Segmentation of brain tumors is extremely

significant, because it can save lives. Accurate segmentation of brain tumors is also important.

By fusion of these results, we got more sharpening and more smoothness image in this research. This is one of contribution for our article. Second is adaptive K-mean clustering is used like as segmentation method in this article. And then, we proposed the usage of opening and closing in morphological operation in this research. Second contribution of our article is the combination of clustering method and modified morphological operation to segment the MRI images. The experimental results will be discussed in the next chapters. This article is implemented with 6 chapters. Chapter 1 is Introduction, chapter 2 is state of the art, chapter 3 is theory background, chapter 4 is material and method, chapter 5 is experiments and chapter 6 is conclusion of the paper. This article is extended version of our paper in PDCAT'17 conference with the title of "MRI images Enhancement and Brain Tumor Segmentation" [1]. In this article, more datasets were tested and presented about the kernel of filters, discussed about the proposed system details.

2. State of The Art

S. Jeevakala and B. Therese described the paper title with "Non Local Means Filter Based Rician Noise Removal of MR Images" in 2016. In this paper, the authors proposed a combination of NLM and stationary wavelet transform (SWT) with adaptive thresholding to remove Rician noise and preserve structural information of edges. The proposed noise elimination algorithm will be useful for the subtle analysis of tissue / organ images [2].

*Aye Min, UCSM, +959789811977 & ayemin@ucsm.edu.mm

M. N. Nobil and M. A. Yousuf proposed the paper title with "A New Method to Remove Noise in Magnetic Resonance and Ultrasound Images". The proposed method is compared with a smoothing, median and midpoint filter using quantitative parameters such as PSNR, SNR, and RMSE. The smoothing filter shows better results, but it is painful because of the blurring effect. In the median filtering technique, it is considered that each pixel calculates the average and all the pixels are replaced by the calculated average. Therefore, the affected pixels are taken into account to calculate the average, and the unaffected pixels are replaced by this calculated average [3]. B. Shinde and AR Dani have announced a "Noise Detection and Removal Filtering Techniques in Medical Images" in 2012. In this experiment, various medical images, such as MRI, cancer, X-ray, brain, etc. All these medical images, after detection of Gaussian noise, use median filtering techniques to remove noise. The results they have achieved are more useful and found useful for general practitioners to easily analyze the patient's symptoms [4].

A. Mihailova et al. (2016) proposed the paper "Comparative Analysis Various Filters for Noise Reduction in MRI Abdominal Images. Gaussian noise is random noise, and has a normal distribution of the probability density function (also known as a Gaussian distribution). Rician noise is not additive noise, but it depends on the data. The median filter performs better than the Gaussian filter. Wiener filter works best, but the most significant results they get from the seismic pulse and especially the wavelet of the homomorphic filter [5]. The next paper is "Propagated Image filtering" and it was presented by J.H.R. Chang et al (2015). In this document, authors proposed a propagation filter as a local filtering operator with the objective of smoothing images while maintaining the context information of the image. Authors also propose technologies when propagation filtering is related to the propagation of beliefs and a greater acceleration of the filtering process is required. In the experiments, the propagation filter was applied to various applications, such as image noise reduction, smoothing, melting, high dynamic range (HDR) compression. Finally, several applications Computer vision and graphics have verified the effectiveness of propagation filters that have proven to be superior to existing image filters in both quantitative and qualitative assessments [6].

The another paper of image fusion is "Image Fusion using NSCT Theory and Wavelet Transform for Medical Diagnosis", authors are P. J. Anju et.al. There are different methods for medical image fusion. NSCT based fusion is further enhanced for better quality by integrating with wavelet fusion. The experimental results are tested and compared with other fusion methods by using the Peak Signal to Noise Ratio (PSNR) and Structured Similarity Index Measure (SSIM) [7]. Md. Sujan et. al. (2016) proposed "A Segmentation-based automated system for the detection of brain tumors". In this work, we proposed threshold processing and a morphological method for detecting a tumor. Authors compared the results of the proposed method with the color segmentation method. They recognized the tumor area by comparing the true positive rate of segmented results. All results are tested with 72 flair sequences of the BRATS dataset [8]. Jyothisna et. al. (2015) proposed "Adaptive K-means clustering for Medical Image Segmentation". A number of the group's researchers focused on improving the clustering process. The proposed method of promoting the adaptive method is that clusters grow without first

selecting the elements that constitute the cluster. It was discovered that it is capable of segmenting the region of different distribution intensity smoothly.

The method was used to achieve a significant process of accelerated research. In this paper, an adaptive clustering algorithm for K-tools is presented and does not depend on seed selection to initialize cluster K. The algorithm is tested for various images and works smoothly, resulting in good data separation and research resulting in the data structure being accelerated remarkably. One can conclude that adaptive to means that it works better, and the speed of K means [9].

Bobotov' et al. (2016) proposed the title with "Segmentation of Brain Tumors from Magnetic Resonance Images using Adaptive Thresholding and Graph Cut Algorithm". They got the result of comparison Graph cut result and result without Graph cut [10]. The following article is described by Edily et al. entitled with "Detection and localization of brain tumors in magnetic resonance". The present inventors propose an automatic frame for the detection and localization of brain tumors capable of detecting and locating brain tumors in magnetic resonance images. The framework for detection and location of brain tumors proposed involves five steps: image acquisition, preprocessing, detection of edges, grouping of modified histograms and morphological manipulation. After morphological manipulation, the tumor appears as a pure white color on a pure black background. This system reached an error rate of 8%. Preliminary results demonstrate how a simple automatic learning classifier with a simple set of image-based features provides high classification accuracy. The preliminary results also demonstrate the effectiveness and efficiency of our 5-step approach to brain tumor detection and detection and extend this framework to detect other types of tumors in other types of medical images and motivate them to be localized [11].

Cabria et al. (2015) proposed "Automated Localization of Brain Tumors in MRI Using Potential-K-means Clustering Algorithm". In this paper, they viewed the intensity of a pixel as equal to its "workload" and employed an unsupervised learning algorithm called potential-K-means that generates a balanced distribution of the pixels into clusters of approximately equal total intensity. A set of 22 images of the FLAIR MRI (axial plane) modality from the BRATS dataset was used [12].

S. Priyanka, Dr. AS Naven kumar proposed the name of the noise elimination document "Noise Removal in Remote Sensing Image Using Kalman Filter Algorithm " in 2016. Remote sensing, in general, using sensors installed on airplanes and space platforms, the authors discussed Gaussian noise and speckle noise (salt and pepper). In this proposed study, authors reduced image noise using the Kalman filter and the Wiener filter. The Kalman filter is suitable for reducing noise while maintaining the basic structure of the image compared to other filters. The Kalman filter shows more filters with improved noise efficiency [13]. The next document is "An interactive graph cut method for brain tumor segmentation ", which is described by N. Birkbeck et al. We have developed a interactive semiautomatic brain tumor segmentation system that incorporates interactive 2D tools and automated 3D Propose Control. The method provided is based on the energy that incorporates the available MRI modalities and the regional

statistics calculated in the normal normalization period. The improvement of the new parameters includes the adjustment of the continuous balance of the adhesion parameters of the operator control and the user interaction in 2D line using Rasso and the brush tool (not including the point and plot click used in the segmentation previous interactive) There is. This improves segmentation control by drastically changing the statistics of the region and limiting the segmentation. Experiments have shown that the proposed tool accelerates segmentation compared to traditional manual segmentation and reduces reproducibility between users and users [14].

S. Priyanka, Dr. AS Naven kumar proposed the name of the noise elimination document "Noise Removal in Remote Sensing Image Using Kalman Filter Algorithm" in 2016. Remote sensing, in general, using sensors installed on airplanes and space platforms, the authors discussed Gaussian noise and speckle noise (salt and pepper). In this proposed study, authors reduced image noise using the Kalman filter and the Wiener filter. The Kalman filter is suitable for reducing noise while maintaining the basic structure of the image compared to other filters. The Kalman filter shows more filters with improved noise efficiency [13]. The next document is "An interactive graph cut method for brain tumor segmentation ", which is described by N. Birkbeck et al. We have developed an interactive semiautomatic brain tumor segmentation system that incorporates interactive 2D tools and automated 3D Propose Control. The method provided is based on the energy that incorporates the available MRI modalities and the regional statistics calculated in the normal normalization period. The improvement of the new parameters includes the adjustment of the continuous balance of the adhesion parameters of the operator control and the user interaction in 2D line using Rasso and the brush tool (not including the point and plot click used in the segmentation previous interactive) There is. This improves segmentation control by drastically changing the statistics of the region and limiting the segmentation. Experiments have shown that the proposed tool accelerates segmentation compared to traditional manual segmentation and reduces reproducibility between users and users [14].

3. Theory Background

3.1. Median Filter

The median filter is a non-linear method used to eliminate noise from the MRI brain images. And it is especially effective for eradicate salt and pepper noise. The median filter works by scrolling the pixel of the image with a pixel, replacing each value with the median value of the neighboring pixels. Pixels are calculated from the first sorting of all pixel values of adjacent patterns in the order, and then replace the pixel when viewed with a half pixel value. The median filter is capable of eliminating noise without degrading the sharpness of the image [5].

$$y[m,n]=\text{median}\{x[i,j],(i,j)\in\omega\} \tag{1}$$

Where ω is a neighborhood defined by the user, centered around location $[m,n]$ in the image. An example of median filter of 3*3 kernel or window size is shown below. We take the original values and order the values to 0, 2, 3, 3, 4, 6, 10, 15 and 97. We find the

medium value and fill this value to the center point. So, centered value 97 is replaced by the medium of all nine values 4.

Unfiltered value

*	*	*
*	4	*
*	*	*

Filtered value

6	2	0
3	97	4
19	3	10

3.2. Wiener Filter

Anti-aliasing is the Wiener (non-linear) filter. This filter simultaneously eliminates noise and blurry integrals. There are two parts of the work: the inverse filter and the noise leveling. The Wiener filters are a class of optimal linear filters, with noisy data because the inputs are the calculation of the difference between the output sequences required of the actual output. Performance supervision can be considered the error of least squares. There is also a Wiener2 filter is an adaptive 2-D noise removal filter. This function works as a filter application. Wiener is a type of image for a linear adaptive filter that adapts to the local variance of the image. Wiener2 has done little to smooth out the great variance. At the small wiener2, more sanding is lit. Therefore, it is often better than linear filtration. In comparison, an adaptive filter is more intuitive than a comparable linear filter, parts of the profiles and high-frequency images. There is no design work, the wiener2 function processes all preliminary calculations, preliminary calculations and filter equipment and the implementation of the general input filter. The Wiener2 filter is more suitable for repairing Gaussian noise.

3.3. Image Fusion

Image fusion is a method of integrating all images of applicable information and balanced similar sources or multiple sources in one merged image without any degradation. The main goal of merging medical imaging is the reliable integration of the observation from different input images into one image, not including any degradation and loss of visual information. There were three main ways of fusion of images - pixel level, feature level and decision level. The pixel level is a low fusion of images, the address of the pixels obtained at the output of the image sensor. Fusion of images at pixel level refers to a mixture of information and synergistic data collected from various image sources to provide an improved type of view. When the merging of images is combined at the pixel level directly into the information layer, the amount of information is greater. Almost all the image algorithms of the merged ones are designed to fall at the pixel level [6].

3.4. Adaptive K-Mean Clustering

Clustering is a major problem in a wide range of fields, such as pattern recognition and artificial vision. The general grouping method is based on the average K. However, it has four main drawbacks. First, it is a late and incorrect time scale. Second, it is often not desirable to wait for the user to identify the number of clusters. Third, it can exacerbate excellent areas. Finally, its performance depends to a large extent on the initial center of the cluster. To overcome the above drawbacks, the grouping algorithm 4 in this document means that K (AKM) is effectively adapted. The AKM to estimate the correct number of clusters and

obtain the initial segmentation of the histogram in the center of the linear norm with the linear norm is composed of a set of data and then a local heuristic improvement to group the K means to avoid the optimal values. Execute the algorithm In addition; the kd tree is used to store data sets to accelerate. AKM has been tested on synthetic and real image data sets.

3.5. Morphological Operation

In morphological operations, binary images and structural elements are often used as input and in combination with use of switches (intersections, conjunctions, inclusions, complements). The processed image object is based on the characteristics of the shape encoded by the input structure element. Mathematical details are described in mathematical morphology. Each pixel of the image, as well as each pixel of the whole image, is compared to a set of elementary pixels. If the group of two elements corresponds to a condition defined by a group operator (for example, when a plurality of pixels of the structural element is a subset of the pixels of the basic image), the pixels below the origin of the structural element are determined Values (for binary images, prerequisites 0 or 1).

Opening. Opening eliminates small objects in the foreground (usually taken as a bright pixel) of images by placing in the background, while closing is eliminating small holes in the foreground and changing the background of small islets in the foreground. These methods can also be used to find specific paths in the image.

$$A \circ B = (A \ominus B) \oplus B \tag{2}$$

Where \ominus and \oplus indicate erosion and dilation, respectively.

Closing. When processing images, closing together with the opening, the basic workhorse of morphological is to eliminate the noise. Opening is eliminating small items, while closing eliminates small holes.

$$A \bullet B = (A \oplus B) \ominus B \tag{3}$$

Where \ominus and \oplus indicate erosion and dilation, respectively.

4. Materials and Methods

4.1. Proposed Method

MRI images are posh by noise such as rice (Rician), Gaussian, salt and pepper. In order to confiscate noise, many noise filtering methods have been proposed. In this system, we propose a fusion method of results for improving MRI image. The way to fuse the results is built by fusion the results of the median filter with the results of the Wiener filter. Direct fusion method is used when merging results. The performance of the medium filter and the winning filter depends on the size of the core or the size of the window. We evaluate and select the best kernel in our study. We examined the choice of kernel size by testing three sets of MRI image data. They are the data set BRATS, DICOM, TCIA. The values of the kernel (3, 5, 7 and 9) or the size of the window are tested and analyzed. Figure (1) shows a part of the sample image of the data set tested for kernel analysis. In MRI images, noise

elimination and kernel value [3 * 3] are very effective in processing time and reconstructed image quality.

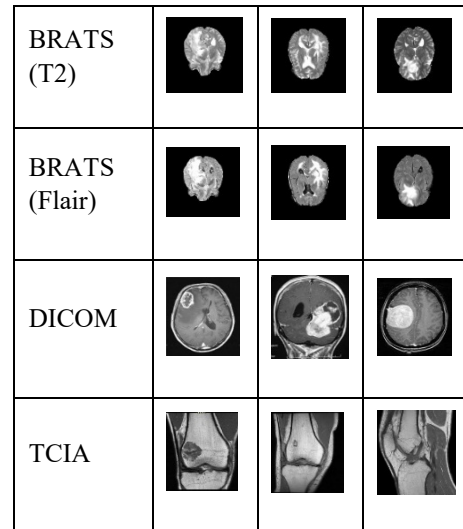


Figure 1: Sample images of datasets

Table 1: Average result of kernel analysis on T2

Kernels	Median Filter		Wiener Filter	
	RMSE	PSNR	RMSE	PSNR
3*3	5.506	33.499	2.677	39.646
5*5	8.946	29.221	3.91	36.35
7*7	11.368	27.107	4.966	34.265
9*9	13.175	25.810	5.905	32.757

Table 2: Average result of kernel analysis on flair

Kernels	Median Filter		Wiener Filter	
	RMSE	PSNR	RMSE	PSNR
3*3	5.567	33.645	2.782	39.557
5*5	8.971	29.483	3.918	36.511
7*7	11.044	27.608	4.648	34.974
9*9	12.623	26.406	5.294	33.822

Table (1), (2), (3) and (4) present the average values of all kernels on T2, Flair, DICOM and TCIA datasets respectively. Allowing to the results of the analysis, the kernel [3 * 3] is ideal for the elimination of noise and the emphasis processing of the

MRI images. Therefore, in this article, we select the median filter and the kernel value [3 * 3] of the Wiener filter.

Table 3: Average result of kernel analysis on DICOM

Kernels	Median Filter		Wiener Filter	
	RMSE	PSNR	RMSE	PSNR
3*3	4.516	35.306	3.888	36.607
5*5	9.965	28.352	6.996	31.428
7*7	15.301	24.580	9.410	28.827
9*9	20.171	22.150	11.324	27.213

Table 4: Average result of kernel analysis on TCIA

Kernels	Median Filter		Wiener Filter	
	RMSE	PSNR	RMSE	PSNR
3*3	8.096	31.086	5.269	34.248
5*5	5.077	34.052	4.637	34.841
7*7	6.672	31.680	5.771	32.940
9*9	7.898	30.214	6.483	31.929

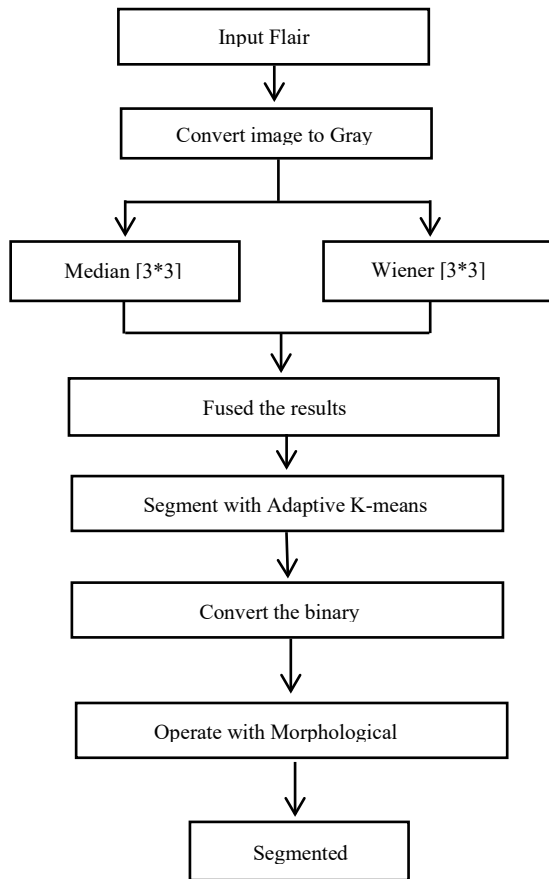


Figure 2: Overview of Proposed System.

Second proposed method is the combination of Adaptive K-means clustering and Morphological operation for MRI images segmentation. The system receives the RGB image and converts the RGB image into a grayscale image. Then, the grayscale image is filtered simultaneously by the medium filter and the Wiener filter. Medium and winning filters use the kernel value [3 * 3] or the size of the window to reduce noise. [3 * 3] The value of the kernel is more powerful and adequate to eliminate noise from MRI images [15]. Therefore, this kernel is used in this research

document. Both filtered results are combined with image fusion. The merged image is segmented using adaptive k-means clustering, after which the segmented image is transformed into a binary image with a threshold of 0.7. The morphological operation re-segments the binary image. In this way, the operation closing and opening are used in order. Closing is [1; 1; 1; 1; 1; 1; 1] value of the kernel and the value of the kernel value Opening is [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]. After applying the morphological manipulation, the system generates images of the tumor segment.

5. Experiments

5.1. Image Enhancement Results

The proposed system is constructed by fusion the result of the Median filter and the result of the Wiener filter. The results of the proposed method are compared with the Medium filter and the Wiener filter. The results of these methods are evaluated using the values of MSE (mean square error) and PSNR (peak signal noise ratio). The table (5) shows the average results of 40 images of Flair and T2. In figure (3), the design of results fusion method is described.

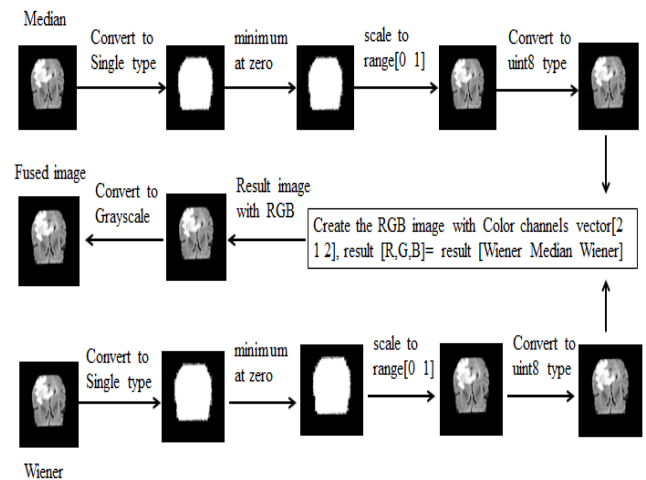


Figure 3: Design of the fusion-base results binding method

Table 5: Average result of MSE and PSNR on 40 flair and 40 T2

Method	MSE		PSNR	
	T2	Flair	T2	Flair
Median filter	28.6604	35.5557	34.1986	33.3925
Wiener filter	31.6925	32.9356	33.4274	33.2784
Proposed Method	21.9649	25.8498	35.1073	34.5308

Table 6: Average result of 72 flair images

Methods	TPR (%)	TNR (%)	PVP (%)	A (%)
AKM	58.47	99.34	87.21	96.16
AKMM (proposed)	85.41	98.90	78.30	98.30
Otsu	55.998	99.39	88.55	95.08
Region growing (RG)	73.64	99.28	86.50	97.42
Particle Swarm Optimization (PSO)	56.10	99.39	88.58	95.13
Interactive Graph Cut (IGC)	42.00	92.64	85.49	86.20

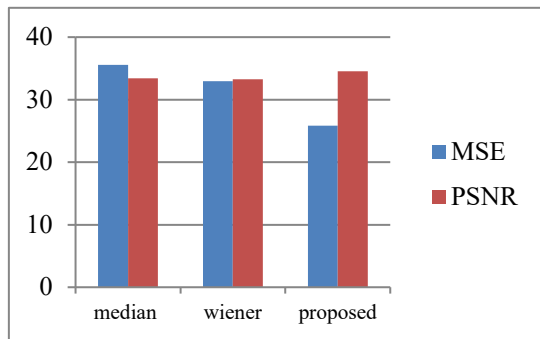


Figure 4: Overview of the Results of MSE and PSNR on T2

5.2. Image Segmentation Results

In this system, the combination of adaptive K-means clustering and morphological approach (AKMM) is proposed for tumor segmentation. Tumor segmentation was tested by two ways to detect advantages and disadvantages of proposed algorithm. First, it was tested with the algorithm which consists only of proposed results fusion method and adaptive K-means clustering (AKM) algorithm. Second, it was tested also with the morphological operation. There are two evaluation methods to use testing accuracy of tumor segmentation. First method includes True positive rate, True negative rate, Predictive value positive and Accuracy. Second method includes Jaccard Similarity index. Resulting tumor segmentation was divided into true positive (TP), true negative (TN), false positive (FP) and false negative (FN) regions. TP represents pixels where tumor was detected and it should be tumor. TN means that tumor was not detected and should not be. FP is when tumor was detected and should not be. Finally if tumor was not detected, but should be, it is FN. Statistical methods were used to evaluate results:

True positive rate – sensitivity (TPR):

$$TPR = TP / (TP + FN) \tag{4}$$

True negative rate- specific (TNR):

$$TNR = TN / (TP + FN) \tag{5}$$

Predictive value positive – precision (PVP):

$$PVP = TP / (TP + FP) \tag{6}$$

Accuracy (A):

$$A = (TP + TN) / (TP + FP + TN + FN) \tag{7}$$

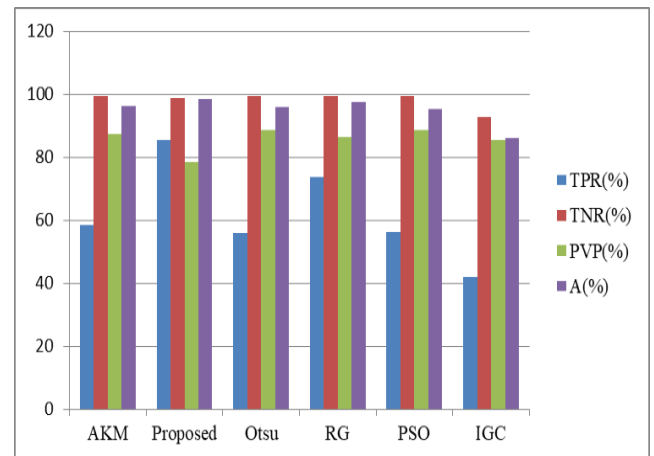


Figure 5: The Results of Comparison Method.

Table 7: Average result of jaccard similarity index in 72 flair images

Methods	Jaccard	Rfn	Rfp
AKM	0.527 (min= 0.174, max=0.7996)	0.9953	1.44
AKMM (proposed)	0.6851 (min= 0.3672, max=0.8427)	0.9958	1.011
Otsu	0.51 (min= 0.0699, max=0.834)	0.9953	2.12
Region growing (RG)	0.647 (min= 0.1124, max=0.87)	0.9954	1.594
Particle Swarm Optimization (PSO)	0.513 (min= 0.0672, max=0.834)	0.9953	2.13
Interactive Graph Cut (IGC)	0.393 (min= 0, max=0.840)	0.9954	2.437

Table 8: Run time duration of comparison methods

Methods	TPR (%)	TNR (%)	PVP (%)	A (%)
AKMM	78.66	99.03	79.81	97.9
PSOM	66.16	99.29	85.99	96.47
OTSUM	66.44	99.29	85.99	96.47

Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets. rfn is ratio of false negative and rfp is ration of false positive.

$$\text{Jaccard}(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (8)$$

$$\text{rfn} = B - |A \cap B| / B \quad (9)$$

$$\text{rfp} = A - |A \cap B| / A \quad (10)$$

Table 9: Performance and comparison analysis of proposed method in BRATS dataset

Performance and comparison analysis of proposed method BRATS brain tumor dataset			
Algorithm	MRI modalities	Approach	True positive rate
Md. Sujan, Nashid Alam(2016)	72-Flair	Thresholding and morphological processing	84.72%
Proposed	72-Flair	Adaptive k-mean clustering and morphological processing	85.41%

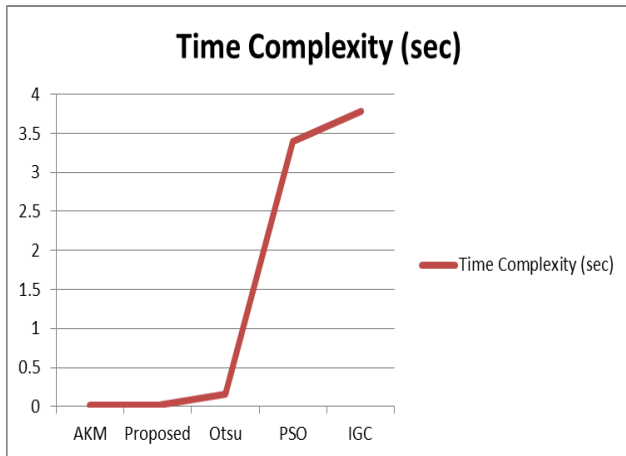


Figure 6: Time Complexity of Comparison Method

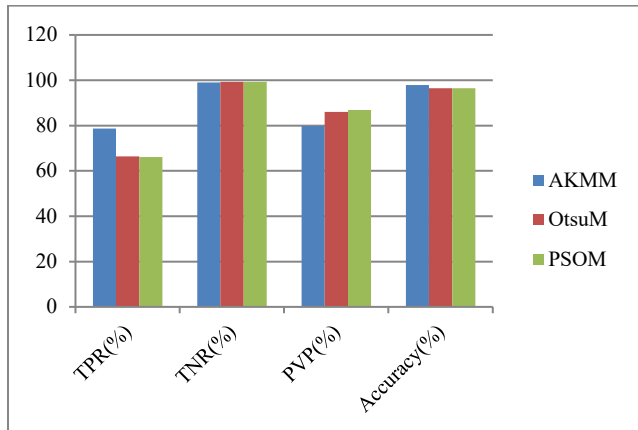


Figure 7: The Averages Results of 110 Flair Images

Table 10: Average results of 110 flair images

Method	Run-time (sec)
AKM	0.0106524
AKMM (proposed)	0.0192652
Otsu	0.15238
PSO	3.39533
IGC	3.77659

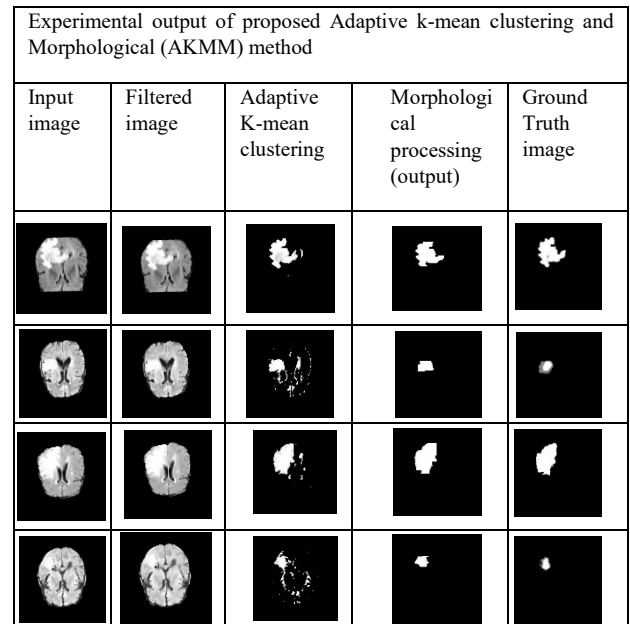


Figure 8: Experimental output of proposed Adaptive k-mean clustering and Morphological (AKMM) method

Table 10 describe about the average experimental results of 110 flair images. In this table, Our proposed method (AKMM) is compared with combination of Otsu and morphological operation and the combination of particle swarm optimization (PSO) and morphological operation. According to the results, our proposed method is more exceeding in Accuracy.

5.3. Research Discussion

In this article, we proposed tow contributions. First is fusion – based results binding to get better result of MRI images enhancement. In this method, we fused the results of Median filter and Wiener filter. At that time, we used more Wiener effect. Thus, our enhancement method image is more smoothness than original and we also save image sharpening from Median filter properties. Among the segmentation methods, almost the segmentation based segment methods are more suitable with median filter and almost the clustering base segment methods are more suitable with wiener filter. Our filtering and enhancement method can be used in both of segmentation. According to experimental results, our enhancement method got better results than Median filter and Wiener filter in MSE and PSNR values. So, we proposed first contribution is better results. And then, we proposed the second contribution.

Second contribution is combination of Adaptive K-means clustering and morphological operation (AKMM). We used proposed filter method in the preprocessing stage. First paper described the results based on 72 flair images. In this article, 110 flair images are tested and described. We also used the Jaccard similarity index in experimental results evaluations. Our proposed method (AKMM) got better results than other comparative base-line methods in accuracy and run time duration. All of the base-line methods are downloaded from matlab file exchange site and re-implemented by our self.

6. Conclusion

In this work, 40 flair and T2 sequences are tested for enhancement of MRI images and 110 flair sequences are tested for tumor segmentation. The fusion based results binding method was proposed for MRI images enhancement and combination of morphological operation and adaptive K-means (AKMM) for tumor segmentation. According to our experimental results, the proposed improvement is superior to Median filter and Wiener filter. Then we test and compare base-line methods such as Otsu's threshold, region growth, particle swarm optimization, and interactive graph cut segmentation with the proposed method (AKMM). The proposed method and all base-line methods are tested on 72 flair images and 110 flair images. The proposed method (AKMM) gained higher accuracy than the basic method compared, and it gets less complex over time.

References

- [1] Aye Min and Zin Mar Kyu, "MRI images Enhancement and Tumor Segmentation for Brain", 18th International Conference on Parallel and Distributed Computing, Applications and Technologies, 0-7695-6330-9/17/31.00©2017IEEE DOI 10.1109/PDCAT.2017.00051
- [2] S. Jeevakala, B. Therese, "Non Local Means Filter Based Rician Noise Removal of MR Images", International Journal of Pure and Applied Mathematics, Volume 109 No. 5 2016.
- [3] M. N. Nobi and M. A. Yousuf, "A New Method to Remove Noise in Magnetic Resonance and Ultrasound Images", Journal of Scientific Research (JSR) Publication, 2011.
- [4] B. Shinde and A.R. Dani, "Noise Detection and Removal Filtering Techniques in Medical Images", International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 4, pp.311-316, July-August 2012.
- [5] A. Mihailova, V. Georgieva, "Comparative Analysis Various Filters for Noise Reduction in MRI Abdominal Images", International Journal "Information Technologies & Knowledge" Volume 10, Number 1, © 2016.
- [6] Rick Chang, J. H., & Frank Wang, Y. C., "Propagated image filtering" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 10-18) 2015.
- [7] P. J. Anju, Dr.D.Loganathan," Image Fusion using NSCT Theory and Wavelet Transform for Medical Diagnosis", International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 7 (3), 2016.
- [8] Md. Sujjan, S.A. Noman, N. Alam, M. J. Islam,"A Segmentation based Automated System for Brain Tumor Detection", International Journal of Computer Applications (0975 – 8887) Volume 153 – No 10, November 2016.
- [9] C. Jyothsna , Dr.G.R.Udupi, "Adaptive K-means Clustering for Medical Image Segmentation",International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Special Issue 31(September, 2015), PP. 15-21
- [10] Z. Bobotov', and W.S. Bene, "Segmentation of Brain Tumors from Magnetic Resonance Images using Adaptive Thresholding and Graph Cut Algorithm", The 20th Central European Seminar on Computer Graphics, Slovakia, Proceedings of CESC 2016.
- [11] Azhari, E. E. M., Hatta, M. M., Htike, Z. Z., & Win, S. L. "Brain tumor detection and localization in magnetic resonance imaging" International Journal of Information Technology Convergence and services (IJITCS), 4(1), 2231-1939 2014.
- [12] I. Cabria, I. Gondra,"Automated Localization of Brain Tumors in MRI Using Potential-K-means Clustering Algorithm", IEEE, 12th Conference on Computer and Robot Vision 2015.
- [13] S. Priyanka, Dr.A.S.N. kumar, " Noise Removal in Remote Sensing Image Using Kalman Filter Algorithm", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016.
- [14] Birkbeck, N., Cobzas, D., Jagersand, M., Murtha, A., & Keszytues, T. "An interactive graph cut method for brain tumor segmentation". In Applications of Computer Vision (WACV), 2009 Workshop on (pp. 1-7). IEEE.
- [15] Aye Min and Zin Mar Kyu, "Kernels analysis in MRI images Noise Removal Methods", 16th International Conference on Computer Application,ICCA(2018).

Semi-Autonomous Robot Control System with an improved 3D Vision Scheme for Search and Rescue Missions. A joint research collaboration between South Africa and Argentina

Jorge Alejandro Kamlofsky^{*1}, Nicol Naidoo², Glen Bright², Maria Lorena Bergamini¹, Jose Zelasco³, Francisco Ansaldo³, Riaan Stopforth²

¹Universidad Abierta Interamericana (UAI), CAETI, C1270AAH, Buenos Aires - Argentina

²University of KwaZulu-Natal (UKZN), Mechatronics and Robotics Research Group (MR²G), 4000, Durban – South Africa

³Universidad de Buenos Aires (UBA), Facultad de Ingeniería, C1063ACV, Buenos Aires - Argentina

ARTICLE INFO

Article history:

Received: 03 July, 2018

Accepted: 26 October, 2018

Online: 01 December, 2018

Keywords:

Search and rescue robots

3D vision

Robotic vision

Robotic navigation

ABSTRACT

Rescue operations require technology to assist the rescue process. The robotic technology in these missions is becoming very important. The important aspects investigated in this study are the integration of a mechatronic system that will allow for a robotic platform with a vision system.

The research collaboration between Argentina and South Africa is discussed, with the correlating research areas that each country investigated. The study permitted the development and advancement of a search and rescue system for different robots (wayfarer and drones) with different vision capabilities. A novel and innovative vision approach is presented.

1. Introduction

This paper is an extension of work originally presented in the 24th edition of the Mechatronics and Machine Vision in Practice (M2VIP) [1].

Search and rescue (SAR) is a general field referring to an emergency response to locate, give aid, medical care and to rescue people. Urban SAR (USAR) is considered a multiple-hazard discipline [2]. Urban environments are often more susceptible to human induced disasters: major industrial and/or transport accidents and/or fires, terrorism, wars, etc. or can have enormous effects in case of extreme natural phenomena like earthquakes, hurricanes, etc.

SAR operations in dangerous, disaster or catastrophe areas can be greatly improved with the assistance of tele-operated robots and/or semi-autonomous robots. Current implementations of mobile robots for SAR operations require human operators to control and guide the robot remotely. Even if human operation can be effective, operators may become stressed and tired rapidly [3]. In addition, current autonomous robots are incapable to work properly by moving in complex and unpredictable scenarios. This challenge could be addressed with a correct balance between the

level of autonomy of the robot and the level of human control over the robot [4].

The first phase in a rescue mission is to identify the target area. The goal is to obtain a precise evaluation of the number and location of victims, detect dangerous situations for rescue personnel or survivors such as gas leaks, live wires, unsafe structures, etc. Semi-autonomous robots carrying correct sensors can obtain data from the field without risks of SAR personnel, and (if connected) can report on line important information. In the other hand, because time is critical during search and rescue operations in catastrophe areas, personnel do not have to waste time waiting for initialization and preparation for robots in the field. The robots must be deployed as soon as possible; if not, the acceptance of the technology by rescue personnel could prove difficult [3].

The mining activity in South Africa has a long list of victims of accidents. This is a case among others, where semi-autonomous mobile robots can access into unstable and toxic areas, in mine shafts, in order to identify dangerous areas, and to save accident victims and rescuers [3]. Argentina had also events in which the use of semi-autonomous robots could have been useful in SAR operations: terrorist attacks, mudslides, and the recent

*Jorge Alejandro Kamlofsky, Email: jorge.kamlofsky@gmail.com

disappearance in the ocean of the military submarine ARA San Juan.

The development and deployment of technology associated with SAR robots will also have a direct benefit to people who live in poor neighbourhoods of South Africa, and Argentina, and in many other circumstances. Fires and mudslides can have a devastating effect in low cost housing developments and the SAR of victims can be improved with technologically advanced semi-autonomous SAR robots. In addition, SAR operations in other disaster cases like, floods [5], mudslides [6], earthquakes, avalanches, fallen buildings, nuclear fusions, explosions, terrorist attacks, forest and/or industrial fires, recovery and deactivation of bombs, etc. can be greatly enhanced with the aid of semi-autonomous robots [7].

SAR personnel use semi-autonomous robots to improve SAR operations. These robots help to recover wounded or trapped people in disaster areas that are dangerous or can pose a threat to human rescuers. SAR operations using semi-autonomous robots can be carried out faster without risk for the SAR personnel. An advantage of using semi-autonomous robots in SAR operations is that they are not susceptible to toxic agents as gases, radiation, acidic or alkaline spills, etc. where toxicity could be dispersed. So, they can access to areas where rescuers cannot reach, or they can do it by risking their lives. If an accident occurs during the operation, there is no injury or loss of human life.

The collapse of the buildings of the Twin Towers in New York (USA) [3], recent catastrophes such as the Fukushima nuclear disaster in 2011 or Nepal earthquake in 2015 showed that SAR or USAR robots can be used to efficiently support rescue teams in finding persons in danger or gathering information more effectively. In Fukushima emergency responders deployed a SAR robot, to check on conditions in the surroundings and allowed workers a safe distance from hazardous radiation [8].

It is proven that semi-autonomous robots are a very useful tool for SAR operations showing the need to investigate how to improve the capacity of SAR robots around the world.

The objective of the joint collaboration of scientific and technological research between South Africa and Argentina is to research, design, develop, assemble and test different systems to improve the functionality of semi-autonomous robots for SAR operations and to improve the quality of life of inhabitants by carrying out rescue enhanced SAR missions in South Africa and Argentina.

The main objective of this paper is to discuss and to analyze the design and development of a semi-autonomous robotic platform that can be implemented in SAR missions in order to improve the results of the operations, as part of the joint research collaboration between South Africa and Argentina. An underlying objective is to remark the importance of the need of research and development of semi-autonomous robotics in order to assist to SAR operations. As an interesting result, in this paper, a simple and promising approach to 3D vision is presented.

The structure of this paper is as follows: Section 2 discusses the methodology of the research collaboration between South Africa and Argentina; Section 3 identifies the basic components (building blocks) of navigation and local control in robots and

discusses the need for a middle-ware service. Section 4 presents the different vision algorithms, techniques and hardware used: the vision system, called here: "The Vision-Ware". Section 5 visualizes the design of a robot control architecture for SAR applications that will be integrated with the middle-ware. Section 6 discusses conclusions and future works.

2. Methodology

The methodology consists of research, design, simulation, development, assembly and testing of different subsystems for semi-autonomous mobile robots in SAR operations. The objective of the research collaboration is to research and develop innovative solutions related to vehicle chassis design, propulsion, navigation, guidance and orientation and 3D vision of the robot.

The methodology will allow the development of new technologies that can optimize the performance of robots in hostile environments so that vehicles can perform the necessary SAR operations reducing risks for human lives.

The research methodology will also focus on the study of new technologies that can be integrated on board of the semi-autonomous robot in order to deliver the necessary first aid equipment and to recover data from field. The methodology includes the development of simulations that will show the viability of subsystems developed and the performance of the robots in different tasks.

The University of KwaZulu-Natal has integrated a drone (Phantom 3) and a prototype platform (Segway RMP400) for vehicle design and propulsion technologies for semi-autonomous mobile robots. Further development and control of this system will be researched and investigated to allow semi-autonomous and autonomous control for SAR purposes.

Argentinean researchers will continue with the investigations carried out in mono and stereo vision, and the integration of cameras, which will help with the location of victims in a disaster environment and navigation and control of the semi-autonomous robotic platform.

3. Background

3.1. Brief and Trends of SAR Robotics

There is a vast literature related development of SAR Robots. Some papers that can show trends are mentioned below.

Actually, because of the most SAR robots are tele-operated, and autonomy robotics do not fit conditions for work and transit in complexity of SAR scenarios, a strong trend is the research and development of semi-autonomous robots, trying to obtain a good balance between the level of autonomy of the robot and human control over the robot [4].

Area mapping and localization for SAR missions research topic since many years, made good progress in terms of performance. A low bandwidth radar-based scanning-technology for mapping is presented in [9]. The radar technology is capable of showing a substantial mapping quality [8].

Swarm robotics consists of a large number of simple and tinny robots working together that perform tasks of greater complexity. It is strongly focused in coordination, cooperation and

collaboration. In particular for swarm robotics applied to SAR operations, optimization algorithms for hazardous environments and laborious tasks were proposed and development [10].

In [11], a case study of IOT application for SAR operations based on cloud is presented. A large-scale deployment of IOT devices in catastrophe scenarios controlled by a cloud based application implementing a “Robot-as-a-Service” scheme (RaaS) giving to SAR operations the advantages of cloud schemes applied to the use of robotic resources: flexibility, cost-efficient, scalability, and virtualization, between others.

3.2. Local Navigation and Control

A very important goal of long term research is the development of a robust navigation area mapping system for mobile robots for unknown and changeable scenarios [12].

The following building blocks of navigation are identified [12, 13]:

- Perception: the robot must obtain field data (raw-data) from their sensors and interpret it and convert it into useful information. Low-level sensing processes extract basic features such as highness or line segments, while high level sensing processes use symbolic models or geometric templates to constrain the sensor interpretation.
- Localization: The robot must identify its position in the field. To achieve this, the robots must have a location support system such as a GPS. If it is not possible, they must recognize the environment in order to know their position on the map. Even more: in several disaster areas, scenarios can vary from saved information. Autonomous vehicles must rely on information obtained from sensor data. They have to be integrated into a unified and consistent mapping model. The positional drift of the sensors due to the robot motion has to be taken into account in the mapping and navigation procedures.
- Cognition: the robot has to decide how to achieve its objectives in the target zone by planning a convenient path (the shortest one in a specific sense).
- Motion Control: In order to move over a planned path, the robot must control and drive its engines connected to its traction system.

In addition to these building blocks, robots must avoid any obstacle, static or dynamic, that may be present on the way to the goal. And if this is not possible, they should be able to plan a new path on the fly. Figure 1 shows a diagram containing the necessary components for the design of semi-autonomous robot system. Regarding to remotely controlled mobile robots, tasks like location, sensing and perception and a part of the cognition module must be controlled locally, in a decentralized manner.

The heterogeneity of the members of a team of robots as well as the information that comes from their various sensors is an advantage that can be used in favor of the success of the mission. For example, in case of an earthquake, a quick picture from a drone can shows the general situation with the possible rescue targets. Then mobile robots like Segway-based ones, more robust and equipped, can be sent to the target area. Meanwhile, some vital

elements can be delivered by drones. Or in a coal mine, a first line of robots can obtain images and gas information in order to be used by rescuers. Once the risky environmental information has been obtained, rescuers will be notified, so they can be prepared [14]. If it is safe for rescuers, they can go ahead and reach the robots and then, operate them continuing with the SAR mission.

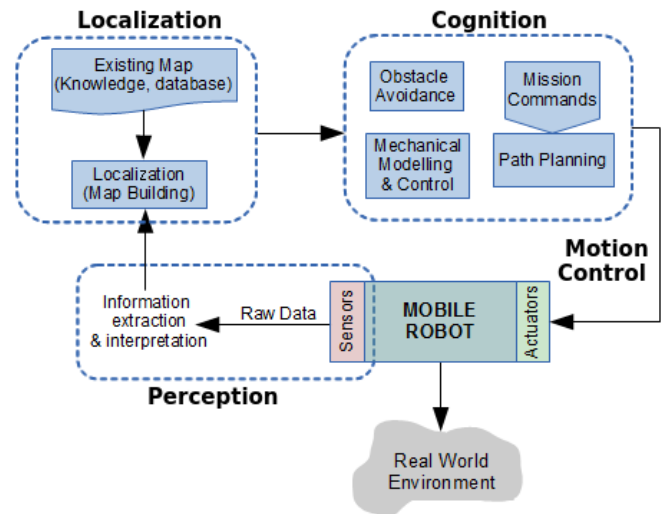


Figure 1. Control structure for autonomous mobile robots

For a collaborative success, the deployed network of robots must use a service software, known as the middle-ware layer.

3.3. Middle-Ware for SAR Robotics

The middle-ware can be seen as the glue that connect everything in the robotic network. It should be designed to allow easy integration of each robot into the network, especially considering that in semi-autonomous robotic implementations for SAR operations, the system configuration time is a critical factor. Another important point that can influence in the direction of the mission is the level of intelligence assigned to the robotic resources for the search process. The middle-ware should provide software interfaces that mask the heterogeneity between the different robots in order to promote cooperation among the robots. This approach allows a more efficient use of the technological resources in order to positively affect the duration of the mission. And considering that time is critical, this will also influence in the final result of the mission.

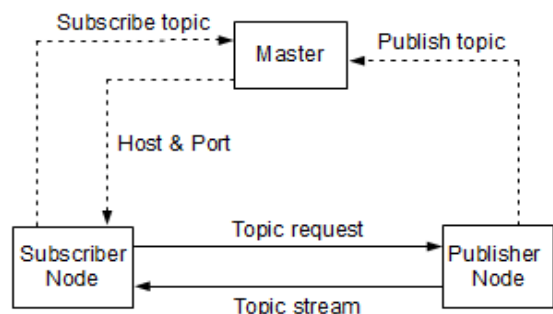


Figure 2. Communication in ROS

The ROS middle-ware consists of nodes, messages, topics and services, and the nodes communicate with each other in a peer-to-peer way (P2P) by publishing messages and subscribing to the

messages posted [15]. An initial event called "naming service" is centralized and based on a master node, which is shown in Figure 2 [15].

A very common and difficult problem when installing a rescue robot in a catastrophe area is the maintenance of a permanent communication. It is expected that communications facilities have been destroyed or inoperative. Therefore, it should be assumed that local communication resources cannot be used, by limiting communications to the native of the system itself [16].

Below is the communication sequence between publisher node and subscriber node:

- Publisher registers a topic (e.g. a laser scan) to the master node informing the point of topic data.
- Subscriber requires the master on how to access to a particular topic.
- Master responds by sending the entry point data, host and port number.
- Subscriber communicates to publisher (host) via TCP or UDP connection requesting for topic data.
- Publisher responds by sending the topic data stream (e.g. laser scan data).

The ROS has been modularly designed. It is organized in packages containing nodes, configuration files, libraries, databases and third party software. So, this packages can be easily integrated into the ROS framework. Exist a wide range of ROS packages available for various robotic implementations such as sensor/actuator drivers, robot path planning and navigation, robot simulation, and others [17].

3.4. Robotic Vision

Computer Vision is a discipline whose main objective is the identification and recognition of objects within digital images, normally acquired by digital cameras. In many applications the identification of moving objects in video captures is required. In addition, in some applications this is required to be achieved in real time. The strategies used for the identification of objects in digital images are based on the analysis of the edges or the study of the whole image or a partial window where the object is probable to be in [18]. The strategies based on edge analysis use techniques to detect lines and edges such as thresholding [19, 20] or by the use of different filters [21] to then extract the characteristics of the shape using digital topology [22, 23] or by recognizing patterns obtained from the border. The strategies based on the analysis of the complete area of the image obtain characteristics of the object contained in the image through training and pattern recognition techniques using neural networks [24], genetic algorithms or using the wavelets transformation [25]. The discretization of the image data, the resolution of the camera, the lack or excess of brightness, the obfuscation, cause lost of clarity or noise in the image, which requires extra treatment [26-28] that in many cases avoid real time requirements.

The robotic vision requires converting the two-dimensional information of the scene obtained from images recorded by cameras into a three-dimensional model in order to recognize objects and places, in order to correctly execute the assigned tasks. Stereoscopy is a technique that imitating human vision, from two images of the

same scene achieves a three-dimensional reconstruction [29]. The computation is achieved after the recognition of homologous points in both images. However, this task requires testing the matching of millions of pixels in millions of possible combinations [30] which compromises the requirements for recognition in real time.

The use of commercial 3D cameras for robotic vision can be expensive and heavy. In the other hand, proprietary technology limits several implementations. The aim of the 3D vision research is to develop a cheap, flexible and open solution with stereo cameras that can generate the 3D reconstruction using stereoscopy in different scenarios.

3.5. About the Research Collaboration between Argentina and South Africa

Within the framework of the Scientific-Technological Cooperation Program between the Ministry of Science, Technology and Productive Innovation of the Argentine Republic (initials in Spanish: MINCYT) and the Department of Science and Technology of the Republic of South Africa (DST), the project called "Semi-Autonomous robots for SAR operations" was selected by the Bilateral Commission MINCYT-DST to be executed in the 2014-2016 triennium under code SA / 13/13.

In charge of the South African part was Dr. Glen Bright of the Mechanics Department of the Kwazulu-Natal University, while in charge of the Argentina part was Dr. José Zelasco of the Mechanics Department of the Faculty of Engineering of the University of Buenos Aires.

Phd students Nicol Naidoo and Jorge Kamlofsky have made progress in their researches. Engineering students from both universities have participated in the developments and attended the presentations of the progress of the project.

Throughout the project various equipment was acquired such as: drones, Segways, cameras and others. Committees of both countries have made presentations about the progress of the project and joint publications were achieved [1, 41, 47, 48]

Although it was not possible to present a functional prototype of rescue robot, the development of a middle-ware to integrate several robotic platforms was made. Several stages of a three-dimensional vision system were developed that prove to be very efficient. Much of this is shown in this work.

Challenges of this experience were the technical developments, the interrelation between the researchers of both countries and the active participation of the students.

4. The Vision-Ware: 3D Vision Algorithms and Hardware Integration

Artificial vision technology is based on a resemblance to human natural vision. Following the analogy with human vision, each eye receives a different luminous stimulus, the distance that separates them produces a parallax angle in the observed object. The image obtained in each retina is integrated and reconstructed in the brain that finally generates the perception of the relief. In order to imitate this reconstruction process, given a pair of images (obtained by cameras) called stereoscopic image, it is necessary to solve the following three stages: Calibration, pairing or put in correspondence of homologous points, and 3D reconstruction.

To obtain the three-dimensional reconstruction of an object present in two images, first, it is necessary to find it in both images. It continues with the pairing or finding of homologous points. This is: to find the location of the points of the object in the first image that correspond to the points of the same object found in the second image. Finally, by means of stereoscopy, the 3D coordinates of those points of the object can be calculated.

In this section, a model that for shape recognition uses just a few points of each object is presented. Therefore, it is attractive to use in real time robotic vision. Experimental data is presented.

4.1. Preparing 3D Models: Digital Camera Calibration

To perform the 3D reconstruction of a scene from multiple images by stereoscopy, these must be calibrated. At this point, the Fundamental matrix is a key concept since with the geometric information available, it allows to obtain the epipolar geometry of the scene from uncalibrated images [49].

If a point P of a three-dimensional referential system is projected on the left image as P_L and on the right one as P_R , (as shown in Figure 3) then the image point direction vectors in the same referential system satisfy the equation $P_L \cdot b \times P_R = 0$, where b is the displacement vector between the cameras: $b = (X_c, Y_c, Z_c)^T$, or the coordinates of right camera in the 3D referential system, if the referential system is posed in the left camera. Since the three vectors are coplanar (epipolar geometry), the mixed product gives zero. By the use of the anti-symmetric matrix B :

$$B = \begin{pmatrix} 0 & -Z_c & Y_c \\ Z_c & 0 & -X_c \\ -Y_c & X_c & 0 \end{pmatrix}$$

The mixed product can be expressed as: $P_L \cdot B \cdot P_R = 0$. W_L and W_R are the coordinates expressed in pixels in the digital images of the points P_L and P_R expressed in length units related by the calibration matrix C as follows:

$$W = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = C \cdot V \quad (1)$$

or:

$$V = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha} & 0 & \frac{-u_0}{\alpha} \\ 0 & \frac{1}{\beta} & \frac{-v_0}{\beta} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = C^{-1} \cdot W \quad (2)$$

where $W = (u, v)^T$ coordinates of the point in images, $(u_0, v_0)^T$ the principal point of the image, α and β focal distances, $V = (x, y)^T$ coordinates of the point in length units.

Because the reference is posed on the left camera, also a rotation of the right camera is considered. Matrix R represents the right camera orientation. Thus:

$$W_L \cdot (C^{-1})^T \cdot B \cdot R \cdot C^{-1} \cdot W_R = 0$$

with:

$$F = (C^{-1})^T \cdot B \cdot R \cdot C^{-1}$$

F is the Fundamental matrix [24, 25, 31] with 9 parameters involved: 4 parameters for the camera calibration, 2 for the rotation and 3 for the base b . It is a square matrix of order 3 and rank 2.

When the calibration is known, only the parameters of B and R are unknown. In this case, it is easy to obtain the essential matrix $E = B \cdot R$ [32] from the Fundamental matrix, since $E = C^T \cdot F \cdot C$. There are several approaches to obtain the 5 involved parameters [33 - 37]. The usual method involves the singular value decomposition of the Essential matrix [38].

Many methods have been proposed in order to calibrate the camera. In [47] is proposed a very simple calibration scheme, which is easy to implement. An assumption made is that it is quite easy to measure the X and Y coordinates of the camera point of view in relation to a 3D referential system when the camera optical axis has a small rotation angle with the Z axis. It is accepted that the error in the Z coordinate of the point of view is less than +/- 1cm. The error in the calibration parameters were obtained and evaluated in [47]. In a later step, assuming that two cameras have a small angle in relation to the Z axis, the calibration parameters error were evaluated and corrected, knowing the distance between two benchmarks in relation to the one calculated by stereoscopy. Regarding the Essential matrix [39], a method for getting the base B and the rotation R , with the solution of linear systems was used [47].

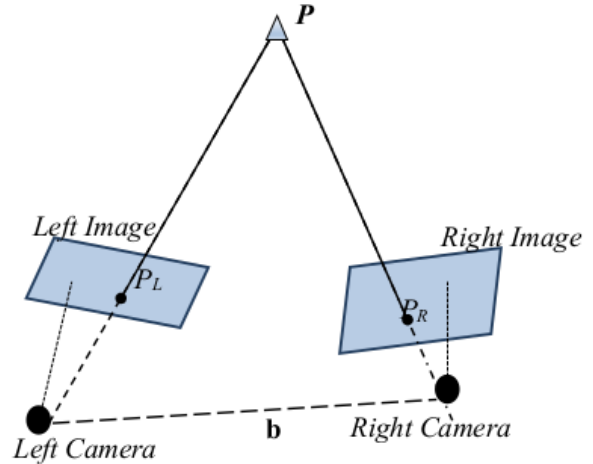


Figure. 3. Epipolar geometry

Calibration parameters are necessary to obtain more precise 3D coordinates of the scene points obtained by stereoscopy.

The interest in recognizing objects within digital images has two main areas of application: obtaining information from images for subsequent human interpretation and processing scenes for autonomous robotic vision [21]. In this last area, for the quick movement and efficient realization of the tasks for which the robot was built, it is especially critical that the processing of the images acquired by the robot, must be done in real time.

The following tasks (detailed below) are necessary to find objects within the images: Image Acquisition, image binarization: separation of objects of interest from the bottom (the complement of objects set), border acquisition: the boundary between objects set and bottom, border simplification: polygonalization, obtaining the curves descriptor pattern, patterns matching. All these tasks have been developed with the premise of realizing as soon as possible, approaching so, to the real time requirements.

Image Acquisition: A Digital Image is a two-dimensional luminous intensity function $f(x, y)$, where x and y denote the spatial coordinates in the image and the value of f at the point (x, y) is proportional to the brightness or gray scale of the image at that point.

To acquire a digital image, an image sensor and the capability to digitize the signal from sensor is required. A sensor could be a monochrome or a color camera [21]. The gray scale of images vary depending the image type. RGB mode is commonly used in programming. In the RGB color images each pixel has three components: Red, Green and Blue, where each component represents the amount of each color in values between 0 and 255 (or one byte) where 0 indicates absence of color and 255 means full color. Thus colors can be represented with 3 bytes or 24 bits. Each point of the digital image is called a pixel. A simple color RGB image has millions of pixels. And each pixel can present color values between 0 and 24 bits.

Image Binarization: In Computer Vision it is convenient to work with binarized images. This is achieved by changing the value of the gray scale by a 1 or a 0. The objective is to separate objects of interest from the background of the image by a process known as "segmentation by thresholding". The simplest way to do this consists of traversing each pixels of the image and assigning a in the gray scale "1" if gray scale conditions are higher than a defined threshold value and "0" otherwise. In this way, a "binarization of the image" is achieved [21]: grouping pixels of objects of interest with one value and pixels of the bottom with the other one.

In other cases, thresholds have to limit color ranges. Thus, multiple thresholds have to be defined: maximum and minimum threshold in each channel: red, green and blue.

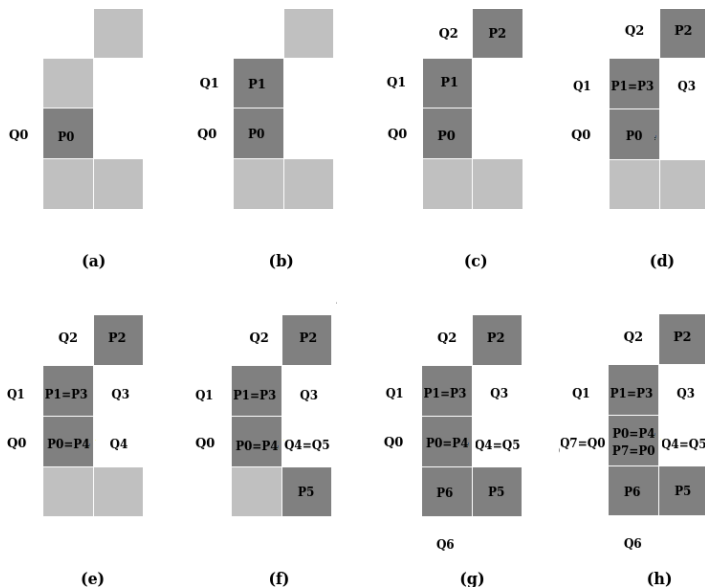


Figure 4. BF8 Algorithm example: (a) Starts in (Q0 , P0). (b) First pair obtained looking 8-neighborhood of P0, going clockwise: (Q1 , P1). Next iterative part illustrated from (c) to (g). Algorithm stops in (h).

The Borders: The objects obtained by binarization of the image are delimited by their borders. A simple algorithm for drawing edges was presented in [22]: BF4 or BF8 depending on whether 4-neighborhoods or 8-neighborhoods are used.

. Figure 4 illustrates algorithm BF8.

BF4 is similar than BF8 but using 4-neighborhoods. With BF4 a thicker edge is obtained, with more points, while with BF8, the edge is thinner, with fewer points. Because of this, the second case is faster.

The edges allow to separate the objects from the background and thus be able to be treated as open sets of the Digital Topology [23]. The shapes of the objects can be characterized from the analysis of their edges.

Border Simplification: By polygonalizing the edges a simplified approximation of the edge curves can be achieved, and thus, an approximate representation of the object hundreds or thousands of times more reduced.

The polygonalization is done iteratively using the convex hull concept [45]: given a set of edge points, those that are inside a convex hull of width ϵ will be eliminated. In this way, the sets of points that are approximately aligned, will be represented by the first and the last which will form two vertexes of a polygon. It will continue with next group of points and then, a new vertex of the polygon is determined till end point is the first point of the process. Figure 5 shows a diagram describing its operation. More details about this process can be found in [46].

The parameter ϵ allows to regulate the compromise relationship between quality and performance: a larger ϵ decreases the quality of the border simplification with greater speed. A small ϵ ensures resolution in the approach. With a larger ϵ , an approximation will be obtained more quickly. It is then left to the user to assume the compromise between resolution and performance through this parameter.

Obtaining of the curves descriptor pattern: From the polygonal curve, a descriptor pattern can be obtained based on the discrete evolution of the curvature parameterized by arc length [42]. It consists of a discrete function $\kappa(\lambda)$ where κ is the curvature (the interior angle) that corresponds to the evolution of the perimeter of the polygon parameterized by the arc length λ . While the evolution of the perimeter λ varies between $[0, 1]$ because it is parameterized by arc length, the curvature accumulates a final value of 2π because it is a simple curve. It is clear that in the points of the polygon that are not vertexes, the curvature is zero. Therefore, only is interesting its analysis in the vertexes. However, from each vertex of each polygon a descriptor pattern can be initiated that will vary by shifting. After determining the orientation of each object [46], a point is defined from where to start each pattern. Thus, each object will be characterized by exactly one pattern.

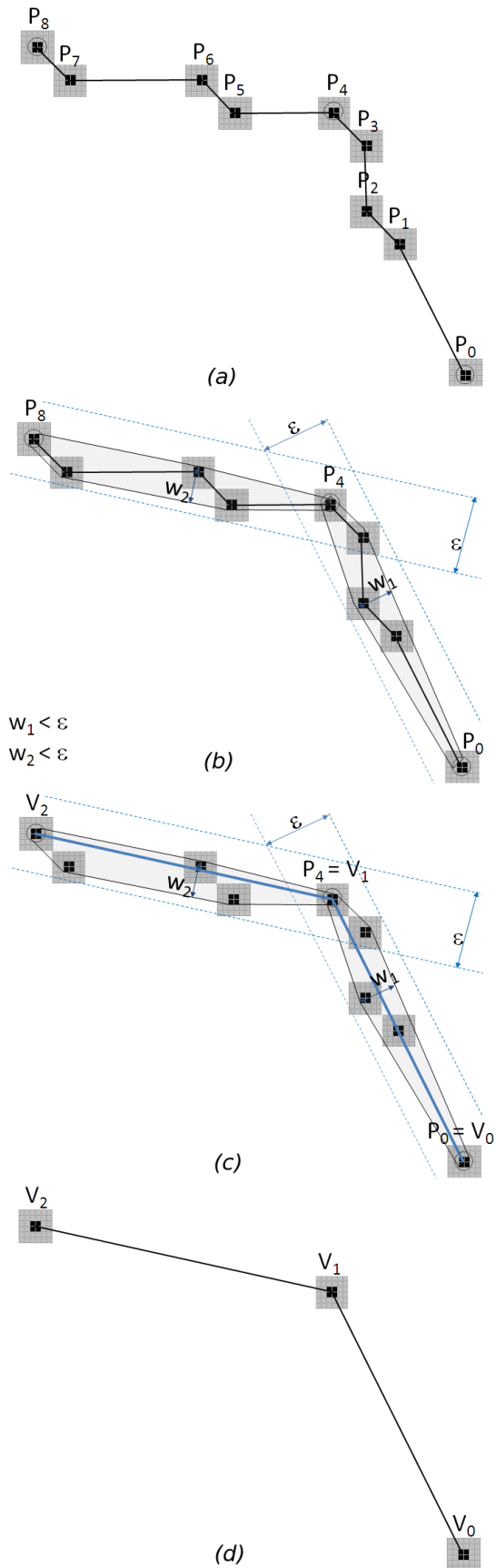


Figure 5: Simplification of curves using the wide convex hull ϵ . (a) A portion of a digital curve composed of 9 points. (b) The convex hulls of width ϵ are established. (c) The vertices of the approximating polygon are defined. (d) Interior points are eliminated and approximation established.

This pattern is invariant to rotations, translations and scaling, that is convenient when it is necessary to find similar objects instead of exactly equal object.

$$d(\kappa_1, \kappa_2) = \int_0^1 (\kappa_1(\lambda) - \kappa_2(\lambda))^2 d\lambda \quad (3)$$

Figure 6 shows a set of patterns corresponding to different objects. In black is shown similar patterns of similar objects.

Because in images normally objects could be in any orientation, these may be subject to transformations (explained in detail in [43]). Since objects are treated as polygons with reduced number of vertices, and polygons can be represented just with its vertices [46], the transformations to all the objects can be performed by a product between the transformation matrix and the matrix formed by the ordered points of the vertices of this polygon.

4.3. Fast Finding of Homologous Points: Preparing 3D models in Real Time

From the recognition of homologous points in both images (called also: “pairing points in images”), the position in the three-dimensional space can be calculated. However, this task requires the pairing of many points, which compromises performance requirements. A model to finding out homologous points was presented in [40] and mentioned in [1]. In this paper, performance improvements is presented. Experimental data is included.

Description of the Process: When an object is found in both images with the method described above, one homologous point is automatically defined (in both curves): the starting point of both patterns.

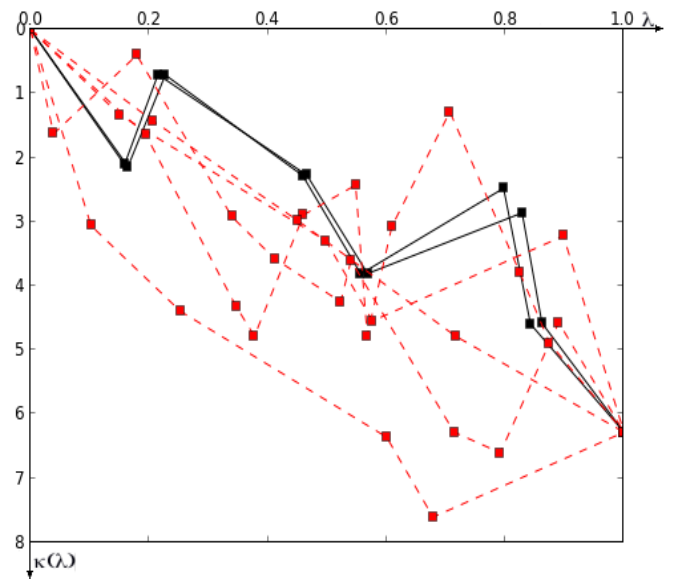


Figure 6: The Curvature Evolution Pattern of various objects.

Thanks to the fact that the curves are parameterized by arc length, and taking into account that the objects are similar, more homologous points can be obtained: this is achieved by traversing the vertices of one of the curves, its counterpart is sought in the

evolution of the normalized perimeter in the second curve, within a tolerance interval. For example, if a tolerance of 5% is performed and for a vertex i of the pattern of the first object with coordinates $(\lambda_{1i}, \kappa_{1i}) = (0.200, 1.000)$, a vertex j in the pattern of the second object is wanted: $(\lambda_{2i}, \kappa_{2i}) / 0.250 > \lambda_{2j} > 0.150$ and $1.050 > \kappa_{2j} > 0.950$. In case of a correspondence, the pair of vertexes (i, j) is added to the set of homologous points.

Performance Improvements: In [1] and in [40] it was shown that for the finding of the homologous points of an object in two images, it took 4.9 seconds. However, contemplating certain restrictions, the same finding could be achieved in times less than 1 second, for the same images.

In the process of objects recognition within digital images, it is intended to improve the performance in the reading of the images (mostly binarized or filtered) that the process provides as parameters.

This operation can be done faster without losing important details under the following assumptions:

- No objects of interest less than 10px wide.
- In a binarized image (black and white), the R, G and B components of the pixel are similar: 0 or 1.
- During the tracking of a moving object, the acceleration and velocity of movement of the object are known.

With these assumptions:

- The horizontal direction of reading of the image is done in discrete 10px jumps instead of being traversed pixel by pixel. If a value of interest is found in one of these jumps, 10 positions are retracted, and the current position is returned by consulting the values of the involved pixels one at a time and extending to the next 10 positions. If this operation was performed in the previous jump, the survey of the previous 10 positions is not carried out.
- If the image is binarized, the R, G and B components of any pixel are similar. Then, instead of requesting the R, G and B components, only one of them is consulted.
- If the speed and acceleration of the object is known, its location can be estimated, so that its search can be done within a reduced window, and not in the whole image.

As a consequence of the above, a great reduction in the time of the reading process of the images is obtained.

4.4. Stereoscopy: Building the 3D Model

Starting with two images, knowing homologous points in both images, by the use of stereoscopy it is possible to perform a 3D re-composition of a scene based on these points. The more homologous points are known, the better the reconstruction will be. Thus, calculation of the 3D coordinates of each of these point (generically named here as M_q) is easily obtained by contributions of information from both images [30]. The spatial coordinates of the generic point $M_q = (x, y, z)^T$ are obtained by using the scene reconstruction equation (Equation 4) applied to each image:

$$\begin{pmatrix} \check{i}.R - U.\check{k}.R \\ \check{j}.R - V.\check{k}.R \end{pmatrix} \cdot M_q = \begin{pmatrix} \check{i}.L - U.\check{k}.L \\ \check{j}.L - V.\check{k}.L \end{pmatrix} \quad (4)$$

Where R is the rotation matrix of the camera, $L = R \cdot S$ with S the position of the camera in the general referential system, and U and V are x and y coordinates in length units of the projection of point M_q in each image. It is important to note that in this equations some parameters calculated in the Calibration step are used. Since M_q is unknown then M_q can be called as X , and (4) can be stated briefly as: $A \cdot X = B$ where A is a 2×3 matrix, and B is a 2×1 matrix. Then, contribution of the left image is: $A_L \cdot X = B_L$ and the contribution of the right image is: $A_R \cdot X = B_R$.

Joining contributions of both images:

$$\begin{pmatrix} A_L \\ A_R \end{pmatrix} \cdot X = \begin{pmatrix} B_L \\ B_R \end{pmatrix} \quad (5)$$

Naming $A = (A_L, A_R)^T$ and $B = (B_L, B_R)^T$, (5) can be presented as:

$$A \cdot X = B \quad (6)$$

Where A is a 4×3 matrix, and B is a 4×1 matrix. Pre-multiplying by A^T in both members of (6), and then by $(A^T \cdot A)^{-1}$, 3D coordinates of the re-construction can be calculated as stated in (7):

$$M_q = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (A^T \cdot A)^{-1} \cdot A^T \cdot B \quad (7)$$

The system is not canonical, that is, the independent term B does not correspond exclusively to the values measured in the images, therefore, even if the error of those measurements was known, the equations could not be weighted and the variance of the elements of the unknown vector could not be calculated.

Once points in spatial coordinates are obtained, the three-dimensional model of the scene can be achieved.



Figure 7: A notebook connected to two regular cameras mounted in a platform with known distance and orientation.

The three dimensional vision system is constructed using two regular cameras. One camera is placed in the origin of the referential system. The other one is located on the x axis, and thus, the referential system is fixed. The distance between both cameras is known, as well as the orientation angles. The greater the distance, the more accurate are the obtained spatial coordinates by stereoscopy. Current cameras can be connected directly to USB ports. And more than one camera can be connected to regular computers, as shown in Figure 7. Cameras are controlled by the vision software. Normally, drivers of regular cameras exist in most

operative system; if not, drivers have to be installed from manufacturer’s site.

The Segway based robot is built to work in harmful environment. Thus, it is recommended to install an industrial computer, which is designed to work under more difficult conditions than regular computers.

Also with small size computers like Raspberry Pi-3 Model-B, 3D vision systems can be built. It is a good option in cases when weight and size availability is limited. In these boards with less than 50 grams and 50cm², 64-bit processor, wireless connectivity and 4 usb ports are included. The base operative system is Raspbian: a specific Linux distribution. If it is not enough, other Linux distributions or Windows 10 for IOT could be installed. These boards can be used in a drone based robot. A different robotic vision system can be constructed with just one camera: images are taken after moving a known distance. The greater the distance between each image, the more accurate are the obtained 3D coordinates. Therefore, precision of 3D re-construction can be controlled by varying the distance between photos, with the height known. Figure 8 shows a Phantom 3 drone with an HD camera.



8: Phantom 3 drone camera with a stabilization mechanism.

In this project is tested the use of regular, IR and thermal cameras.

4.6. The Experiment

Brief Description: The experiment consists of obtaining the homologous points of an object present in two images, and measuring of CPU time consumed in each part of the process repeatedly.

Equipment Used and Settings: For the acquisition of images, two Logitech cameras C170 230mm distant were used. Being conventional cameras, the segmentation is achieved by placing objects on a white background. The algorithm was programmed in Python 2.7. The computer used is an HP Pavillion notebook with AMD A10 processor with 12Mb of Ram and 4 cores. The operating system used is Kali Linux: a Linux distribution based on a Debian kernel. Each image acquired has 640 x 480 pixels.

Values used: The width of convex hulls was $\epsilon = 4$. Tolerance in finding homologous point was 5%.

Results Obtained: The set S of homologous points obtained is:

$$S = \{[(390, 242), (187, 191)], [(417, 248), (212, 198)], [(458, 272), (243, 216)], [(514, 273), (310, 213)], [(511, 354),$$

$$(307, 295)], [(440, 347), (233, 292)], [(436, 333), (225, 277)], [(351, 329), (146, 281)], [(274, 294), (73, 256)], [(250, 265), (46, 221)], [(253, 257), (49, 215)], [(292, 274), (94, 230)], [(341, 277), (140, 227)], [(388, 244), (181, 196)]]$$

Between brackets the homologous points are presented. The first pair corresponds to the first image, while the second pair has its homologous in the second image. Points are expressed in pixels.

In Table, repeated measurements of CPU time consumed in each part of the process, is shown.

From the data showed in Table the average of the time of the process is 0,9199 seconds (less than 1 second). Figure 9 shows images used for the experiment with the homologous points.

Table 1: CPU times (in seconds) of pairing points in images

# test	Model Image (Img 1)			Target Image (Img 2)			Matching	Pairing	Total
	Segment. matrix	Borders	Patterns	Segment. matrix	Borders	Patterns			
1	0,2310	0,0626	0,1121	0,2257	0,0663	0,1015	0,0253	0,0261	0,8506
2	0,2247	0,0578	0,1224	0,2313	0,0589	0,1027	0,0277	0,0237	0,8492
3	0,2435	0,0682	0,1632	0,2356	0,0708	0,1592	0,0262	0,0232	0,9899
4	0,2339	0,0684	0,1689	0,2309	0,0663	0,1632	0,0237	0,0266	0,9819
5	0,2233	0,0725	0,1662	0,2254	0,0668	0,1568	0,0263	0,0260	0,9633
6	0,2254	0,0717	0,1115	0,2553	0,0686	0,1546	0,0309	0,0276	0,9456
7	0,2353	0,0627	0,1129	0,2349	0,0634	0,1085	0,0262	0,0239	0,8678
8	0,2328	0,0656	0,1191	0,2283	0,0674	0,1093	0,0264	0,0226	0,8715
9	0,2286	0,0632	0,1604	0,2397	0,0677	0,1620	0,0200	0,0236	0,9652
10	0,2379	0,0632	0,1139	0,2216	0,0692	0,1612	0,0225	0,0245	0,9140

Previously, the relationship between the level of autonomy and remotely assisted control in semi-autonomous robots for SAR tasks was discussed, and this is presented in this section. A hybrid control architecture is defined: an internal control block for automatic tasks, and a remotely assisted management and control scheme.

The robots have an internal control block that allow efficient time running of automatic tasks as: sensing, moving, obstacle avoidance, local path planning, etc. The robotic system have also a higher level of control called the “Navigation Control Center” (NCC) as shown in Figure 10. The objective of the NCC is to record and update the mapping models with the data acquired from the field and to assign a coverage area to robots and them to determine its navigation algorithm.

The control panel is the part of the NCC (not shown in the NCC block in Figure 10) where robots can be controlled remotely. The operators can request for information or tasks interacting through the graphical user interface (GUI) of the control panel, that will be developed in order to facilitate remote operation. A request made in the NCC is a high level task meanwhile in the local control block of the robot, it is transformed into a specific sequence of tasks just programmed. NCC and local control blocks are connected by a radio link.

The robots of the network will be powered by the ROS middle-ware. It consists of a master - subscriber/publisher nodes as explained in Section III and shown in Figure 2. ROS consist of the following software modules:

- Perception: ROS software drivers consist of packages of software that connect physical sensors with the operative

system. Normally drivers are installed in the ROS in its last version. But in some specific cases drivers have to be installed. Depending of the kind of robot, different sensors will be mounted. While in a basic drone just one camera is carried, in the mobile robot based on a Segway platform a plenty of sensors could be installed: cameras, microphones, temperature, humidity, radiation, gas, etc.

- Localization: It is the module of software that calculates the position of the robot in the field, and dynamically builds the maps of newest areas or updates maps of known areas. The new information is shared with the NCC to optimize SAR process.

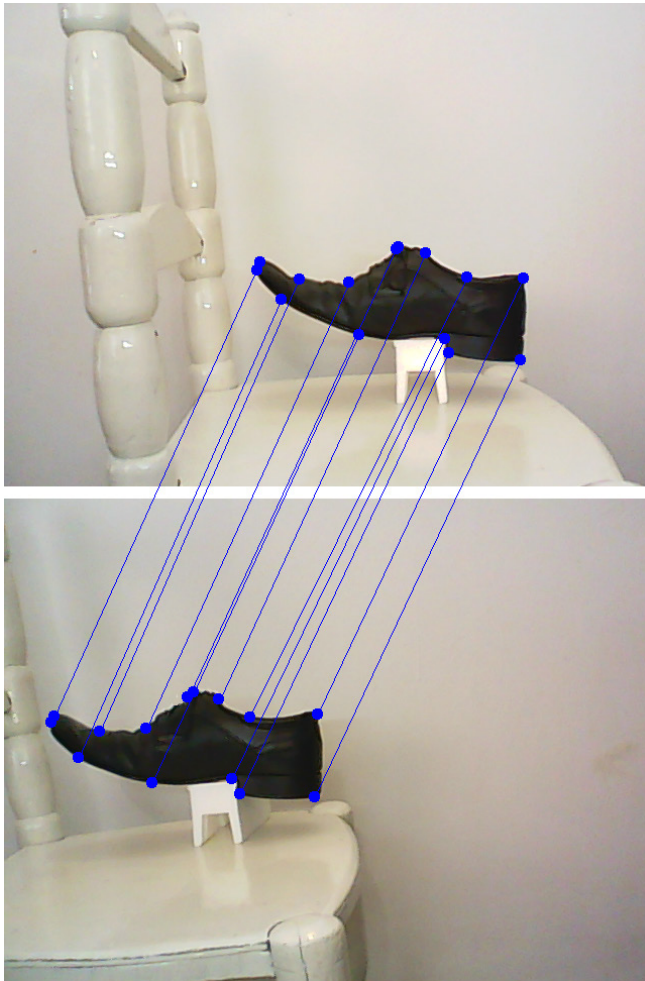


Figure 9: Homologous points of an object in two images.

- Cognition: This module has three important functionalities. The first one is the motors control module: provides electrical control of the motors that make physical actions of the robot such as to catch an object or to move from one site to another. The second functionality is obstacle avoidance: considers safety measures to prevent harm to persons, damages to the environment and to the robot itself. The last one is the local path planning: the objective is to make a surveillance in an area searching for humans that may require help or needs being rescue.
- 3D vision algorithm: this involves the processing of data from images acquired by the stereo cameras. By the 3D vision

algorithm, data (images) are transformed into a higher level of information like an obstacle to be avoided or a possible alive person. This information must be shared with the NCC (so rescue personnel can be sent).

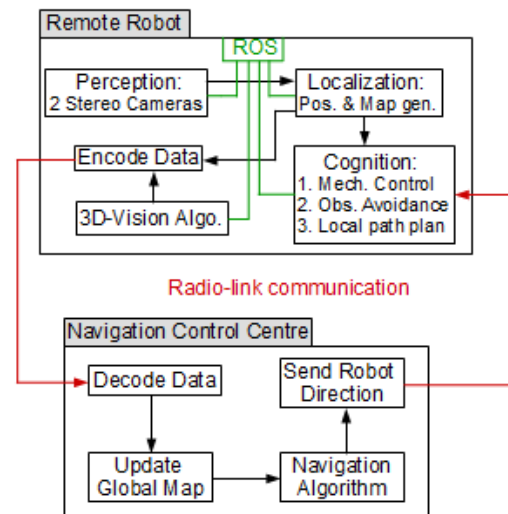


Figure 10. Robot communication and control architecture

A design factor that is a non-trivial problem is the possibility of radio-connectivity problems due to lost of signal strengths caused by interferences of the structures and constructions or because the geography of the site. To solve it can involve an in-depth research analysis, as discussed in [44].

6. Conclusions

In this paper, the need and convenience of the development for its later use in SAR missions of various semi-autonomous robots was introduced and presented.

Collaboration for research between Argentina and South Africa was established, which allowed researchers to make contributions for the design, development, simulations and testing of semi-autonomous robots based on different physical platforms: Segway and drone-based. A hybrid control architecture was presented between the remote robots and the control and navigation center. Interesting improvements in the 3D vision algorithms were presented, which seem to be promising since they would contribute to achieve three-dimensional modeling in real time.

Research results achieved during the binational project related to developments, knowledge exchange and active participation of students from both countries were presented.

Acknowledgement

The research presented in this paper is part of a joint collaboration between the University of KwaZulu-Natal, in South Africa, and the University of Buenos Aires, in Argentina. The

authors wish to thank their respective governments for funding this research and for making the collaboration possible.

References

- [1] N. Naidoo, G. Bright, R. Stopforth, J. Zelasco, F. Ansaldo, M. Bergamini, and J. Kamlofsky, "Semi-autonomous robot control system and with 3D vision scheme for search and rescue missions: A joint research collaboration between South Africa and Argentina." In *Mechatronics and Machine Vision in Practice (M2VIP)*, 24th International Conference on (pp. 1-6). IEEE, 2017.
- [2] G. De Cubber et al., "Search and rescue robots developed by the European Icarus project, 7th Int." In *Workshop on Robotics for Risky Environments*. 2013.
- [3] Greer D., P. McKerrow, and J. Abrantes, "Robots in Urban Search and Rescue Operations", © ARAA, In 2002 Australasian Conference on Robotics and Automation, 2002.
- [4] B. Doroodgar et al., "The search for survivors: Cooperative human-robot interaction in search and rescue environments using semi-autonomous robots." In *Robotics and Automation (ICRA)*, 2010 IEEE International Conference on. IEEE, 2010.
- [5] J. Casper and R. Murphy, "Human-Robot Interaction during the Robot-Assisted Urban Search and Rescue Response at the World Trade Center", *IEEE Transactions on Systems, Man and Cybernetics, Part B*, Vol. 33, No. 3, 2003.
- [6] R. Murphy and S. Stover, "Rescue Robots for Mudslides: A descriptive study of the 2005 La Conchita Mudslide Response", *International Journal of Field Robotics*, Wiley, 2008.
- [7] K. Kleiner, "Better robots could help save disaster victims", 2006.
- [8] S. Brenner, S. Gelfert and H. Rust. "New Approach in 3D Mapping and Localization for Search and Rescue Missions." *CERC2017*: 105-111, 2017.
- [9] P. Fritsche. and B. Wagner, "Scanning Techniques with Low Bandwidth Radar for Robotic Mapping and Localization," in *Informatics in Control, Automation and Robotics (INCINCO 2015)*. IEEE: 321-335, 2015.
- [10] A. Kumar et al., "Search and rescue operations using robotic darwinian particle swarm optimization." In *Advances in Computing, Communications and Informatics (ICACCI)*, 2017 International Conference on. IEEE, 2017.
- [11] C. Mouradian, S. Yangui, and R. Glitho. "Robots as-a-service in cloud computing: search and rescue in large-scale disasters case study." In *15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2018.
- [12] A. Elfes, "Using occupancy grids for mobile robot perception and navigation." *Computer* 6: 46-57, 1989.
- [13] R. Siegwart and I. Nourbakhsh, "Introduction to Autonomous Mobile Robots", MIT Press, Cambridge, Massachusetts, 2004.
- [14] J. Zhao et al. "A search-and-rescue robot system for remotely sensing the underground coal mine environment." *Sensors* 17.10, 2017.
- [15] N. Naidoo, G. Bright, and R. Stopforth, "A Cooperative Mobile Robot Network in ROS for Advanced Manufacturing Environments", in *Proceedings of the International Conference on Competitive Manufacturing (COMA'16)*, 2015.
- [16] S. Shin et al. "Communication system of a segmented rescue robot utilizing socket programming and ROS." *Ubiquitous Robots and Ambient Intelligence (URAI)*, In 14th International Conference on. IEEE, 2017.
- [17] List of ROS packages for Indigo, <http://www.ros.org/browse/list.php>, last accessed 18 July 2017.
- [18] D. Zhang, G. Lu, "Review of shape representation and description techniques". *Patt. Recogn.*, 37 (1),1-19, 2004.
- [19] L. Davis, "A survey of edge detection techniques." *Computer graphics and image processing* 4.3: 248-270, 1975.
- [20] A. Rosenfeld and A. Kak. "Digital picture processing". Vol. 1. Elsevier, 2014.
- [21] R. Gonzalez, R. Woods. "Digital image processing." Addison-Wesley Publishing Company, 1993.
- [22] A. Rozenfeld, "Digital Topology", *The American Mathematical Monthly*, 86(8) pp. 621-630, 1979.
- [23] U. Eckhardt and L. Latecki. *Digital topology*. Inst. für Angewandte Mathematik, 1994.
- [24] H. Rowley, S. Baluja and T. Kanade. "Neural network-based face detection." *IEEE Transactions on pattern analysis and machine intelligence* 20.1: 23-38, 1998.
- [25] A. Fernandez Sarría, "Estudio de técnicas basadas en la transformada Wavelet y optimización de sus parámetros para la clasificación por texturas de imágenes digitales". Diss. Universitat Politècnica de València, 2007.
- [26] J. Weaver, et al. "Filtering noise from images with wavelet transforms." *Magnetic Resonance in Medicine* 21.2: 288-295, 1991.
- [27] L. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms." *Physica D: Nonlinear Phenomena* 60.1: 259-268, 1992.
- [28] T. Nguyen, I. Debléd-Rennesson. "Curvature Estimation in Noisy Curves." In: W. Kropatsch, M. Kampel, A. Hanbury (eds.) *Computer Analysis of Images and Patterns, LNCS*, vol. 4673, pp 474-481, Springer, Heidelberg, 2007.
- [29] H. Moravec. "Robot spatial perception by stereoscopic vision and 3d evidence grids." *Perception*, 1996.
- [30] J. Zelasco et al. "Computer vision in AUVs: automatic roto-rectification of stereo images." *OCEANS 2000 MTS/IEEE Conference and Exhibition*. Vol. 3. IEEE, 2000.
- [31] Q. Luong and O. Faugeras, "Determining the Fundamental matrix with planes: Instability and new algorithms", *Proc. Conf. on Computer Vision and Pattern Recognition*, pp 489-494, 1993.
- [32] Q. Luong and O. Faugeras, "Self-Calibration of a moving camera from Point correspondences and fundamental matrices", *International Journal of Computer Vision*, 22 (3), pp 261-289, 1997.
- [33] H. Longuet-Higgins, "A Computer Algorithm for Reconstructing a Scene from Two Projections", *Nature*, 293 (10), pp 133-135, 1981.
- [34] J. Heikkilä, "Geometric camera calibration using circular control points", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(10), pp. 1066-1077, 2000.
- [35] Z. Zhang. "A flexible new technique for camera calibration". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(11), pp. 1330-1334, 2000.
- [36] Z. Zhang, "Camera calibration with one-dimensional objects" *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(7), pp. 892-899, 2004.
- [37] H. Stewénius, C. Engels and D. Nister, "Recent Developments on Direct Relative Orientation", *ISPRS Journal of Photogrammetry and Remote Sensing* 60, pp 284-294, 2006.
- [38] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2003.
- [39] M. Kalantary and F. Jung, "Estimation Automatique de l'Orientation Relative en Imagerie Terrestre.", *XYZ-AFT*, 114, pp 27-31, 2008.
- [40] J. Kamlofsky and M. Bergamini, "Rápida Obtención de Puntos Homólogos para Vision 3D". In *VI Congreso de Matemática Aplicada, Computacional e Industrial (VI MACI 2017)*, 2017.
- [41] M. Bergamini, F. Ansaldo, G. Bright, and J. Zelasco. "Fundamental Matrix: Digital Camera calibration and Essential Matrix parameters". In *International Journal of Signal Processing*: 120-126, 1 (2016).
- [42] T. Kanade and O. Masatoshi. "A stereo matching algorithm with an adaptive window: Theory and experiment." *IEEE transactions on pattern analysis and machine intelligence* 16.9: 920-932, 1994.
- [43] E. Lengyel. "Matemáticas para Videojuegos en 3D". Second Edition, Cengage Learning, 2011.
- [44] S. Chouhan, D. Pandey and Y. Chul Ho, "CINeMA: Cooperative Intelligent Network Management Architecture for Multi-Robot Rescue System in Disaster Areas", in *Proceedings of the International Conference on Electrical, Electronics, Computer Science, and Mathematics Physical Education and Management*: 51-61, 2014.
- [45] M. De Berg et al. "Computational Geometry", ISBN 3-540-61270-X Springer-Verlag Berlin Heidelberg New York, 1997.
- [46] J. Kamlofsky and M. Bergamini. "Patrón de Evolución Discreta de Curvatura y Concavidad para Reconocimiento de Formas." *CONAHSI*, 2013.
- [47] M. Bergamini, F. Ansaldo, G. Bright and J. Zelasco. "Fundamental Matrix: Digital camera calibration and Essential Matrix parameters". In *16th International Conference on Signal Processing, Computational Geometry and Artificial Vision 2016 proceedings (ISCGAV)*, 2016.
- [48] N. Naidoo et al. "Optimizing search and rescue missions through a cooperative mobile robot network." *Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech)*, IEEE, 2015.
- [49] Q. Luong et al. On determining the fundamental matrix: Analysis of different methods and experimental results. *INRIA*, 1992.

A Practical Approach for Extending DSMLs by Composing their Metamodels

Anas Abouzahra*, Ayoub Sabraoui, Karim Afdel

Laboratory of Computer Systems and Vision LabSIV, Ibn Zohr University, Agadir, Morocco

ARTICLE INFO

Article history:

Received: 27 August, 2018

Accepted: 21 November, 2018

Online: 05 December, 2018

Keywords:

Model Driven Engineering

Domain Specific Modeling

Domain Specific Languages

Model Composition

Software Reuse

Code Generation

Experimental Software

Engineering

ABSTRACT

Domain specific modeling (DSM) has become popular in the software development field during these last years. It allows to design an application using a domain specific modeling language (DSML) and to generate an end-solution software product directly from models. However providing a new DSML is a complex and costly job. This can be reduced by the reuse of existing DSMLs to compose new ones through a metamodel composition approach. This paper provides a composition rules based code generator facility for extending DSMLs. In doing so, it proposes three rules to compose DSMLs by composing their metamodels: reference rule, specialization rule and fusion rule. The results of an exploratory case study on using these rules are depicted. In addition a proof of concept of the code generator facility which generates the necessary infrastructure to quickly build new DSMLs is implemented and applied to the case study. The benefits of our approach are measured relying on three indicators: the reduced development time, the reused software components and the gain on learnability.

1. Introduction

This paper is an extension of work originally presented in 2017 European Conference on Electrical Engineering and Computer Science (ECCS) [1]. The motivation of this paper is to improve the state of the art of quick development of new DSMLs based on existing ones.

Software composition is a fundamental mean for the evolution of complex software systems [2]. While initial approaches were simply focused on textual composition, more efficient approaches take into account syntax and semantics of the software. There was a tendency over the last twenty years towards operation based composition because of its increased expressiveness. In this direction, Model Driven Engineering (MDE) [3,4] was concerned about improving model composition approaches. From early, the researchers have realized that the application of MDE to complex systems will undoubtedly go through the development of smart and agile model composition techniques [5–9].

A use that takes advantage of model composition is to speed up the implementation of new Domain Specific Modeling Languages (DSML). Designing DSMLs is a not an easy job and generally consuming time [10]. This operation can be simplified by reusing existing DSMLs, composing their metamodels, to get new and larger ones [11]. In fact, the definition of a DSML is based on a metamodel and often provides supporting tools as graphical

editors to create and handle models. Therefore, it would be judicious to define the reuse of artifacts at the abstract level; i.e. at metamodels level, then to deduce the projection of this composition at the underneath levels; i.e. the supporting tools.

In the previous work [1], composing metamodels of DSMLs was studied and consequences on their graphical editors were investigated in order to provide a composition of metamodels based approach to extend DSMLs. This work goes further and presents an exploratory study that aims to evaluate the DSML composition approach exposed in [1]. It implements a proof of concept of this approach by developing a code generator facility to make composing graphical editors of DSMLs easier. This prototype provides an automatic code generator which starts from a composition of metamodels of DSMLs, described using composition rules, and generates a layer of code allowing a rapid composition of a new DSML. The gain is measured in terms of development time that can be estimated via the percentage of the generated code. Then, in terms of reused components that can be estimated via the percentage of reused code. As well as in terms of learnability, that can be estimated via the part of kept features and interfaces. These three parameters will be the indicators of evaluation and performance to assess the contribution of this work.

The paper is organized as follows: In Section 2, a series of related works is cited. In Section 3 the problem is stated and the followed methodology is explained. In Section 4 the exploratory

* Anas Abouzahra, Email: abouzahra.anas@gmail.com

study is developed. In Section 5 results are discussed. Finally the Section 6 concludes the paper.

2. Related Works

In this Section a selection of works addressing the composition, extension and reuse of DSMLs in an MDE context is exposed with a brief summary of their features. Moreover, approaches coming from Aspect Oriented Programming (AOP) [12] and Language-oriented programming (LOP) [13] research fields are presented. That is because they have inspired relevant methods for reusing and extending DSMLs. Other approaches that use software product line (SPL) [14] techniques exist [15] but they are minor.

2.1. In MDE

MDE addressed the problem of extending DSMLs by a composition operation. However it varies according to the meaning that each work gives to the composition. This can be a merge operation between models conforming to the same metamodel. As it can be a fusion operation between completely heterogeneous models. It can also be just a resolution of differences between different versions of the same model in order to resolve existing conflicts. Besides, the composition operation can be automatically generated based on mapping calculations or completely custom based on a weaving definition. Furthermore, other approaches provide complementary operations such as checking the consistency of the composition.

Epsilon EML [16] is an Eclipse project which provided a platform for developing substantial and interoperable operations on DSLs among which there is a model composition operation (merging) [17] provided through the Epsilon Merging Language (EML) language [18,19]. EML is applied to compose a number of potentially heterogeneous models. The composition operation is achieved through four steps: comparison, conformance verification, composition and reconciliation. EML was the first language accommodate for model merging and made the case for non-trivial merging of heterogeneous models. However it turned out that it is too verbose for merging homogeneous models. Although, EML is still maintained with significant evolution of its syntax, semantics, capabilities and its underlying platform

AMW [20] is an Eclipse project which proposed a model composition solution (weaving) in parallel of a higher level transformation. In AMW, megamodeling has been introduced to tackle advanced metamodel management, where often the relationships between the metamodels can be considered as composition links. Model weaving has been often used as a solution to compose different DSLs, where the composition is not only the simple gathering of concepts coming from different metamodels, but might also include advanced semantic operators. Unfortunately, and despite all the interest in this tool and its various applications, especially for the traceability of model transformation, the associated eclipse project has been archived. It has not been maintained by the community and no longer by an industrial.

MOMENT [21,22] is a project which aimed to provide a model management platform that furnishes generic operators to handle metamodels described using the Eclipse Modeling Framework (EMF) [23]. In this context, Boronat et al. developed a practical approach for generic model merging. It provides an automate merge operator to merge DSLs artifacts with support for conflict resolution and traceability [24] relying on the QVT Relations

language [25]. This work was applied, and specially proven, to class diagrams integration [26].

Melange [27,28] is a project which treated the modularity and the reuse of DSLs and brought a meta-language for implementing DSLs by composing and specializing existing DSL units. It specifies operators for language assembly, for language extensions and language restrictions. Almost introduced operators by are meant to reuse either the semantics or the metamodel as is, in addition of merging code. Except the inheritance operator which is able to modify the initial definitions in the new DSL. Nevertheless, it is not clearly explained how the extended metamodel modify the original one and which concepts can be overridden.

MetaEdit [29,30] is a graphical workbench which provided a language for creating DSMLs. It introduced the concept of joint/linked modeling constructs to reuse DSLs with code generation facility. In MetaEdit+, the code generation is obtained by the use of a template based on the target language. Consequently, it limits the modularity scope of the generation to the modularity capabilities of the target language. Nevertheless, in MetaEdit+ each created DSL is an addition to, or an extension of, the language workbench itself [31]. This extends the capabilities of reuse to DSLs that are already defined in the workbench.

Other works have treated the problem of model composition and reuse from different angles. Indeed Berg et al. in [32] propose an operational semantics based approach for composing and reusing metamodels and models, by including their operational semantics. Composition is performed relying on a reusable template that permit customizing the metamodel meta-concepts as part of the composition operation. It uses a placeholder mechanism where given meta-concepts of a given metamodel are reused in another metamodel [33]. Schmidt et al. in [34] treated the problem of model composition from a collaborative modeling point of view. They proposed an approach to ease the merging of complex models that are collaboratively developed in teams. This approach aims to furnish collaborative development capabilities in much the same quality as it is provided by version control software or text document merging tools. A recent work in [35] contributed to the same purpose. More, it focused not only on conflicts but also on arbitrary syntactic and semantic consistency issues. Coherent artifacts are merged automatically and only conflicting artifacts are presented to the designer's attention, along with a systematic suggestion of resolution. Otherwise, some works focuses on providing complementary operations to model composition such as checking its consistency. In this direction, Zhang et al. [36] implemented WMCF which is a models composition framework relying on the Alloy language [37,38]. It furnishes a model weaving capability with consistency checking of the resulting composition provided by the Alloy Analyzer.

Besides, generating graphical editors from an abstract definition of a DSL has been addressed by many works. Notably, the *EMF Edit*, the *Graphical Modeling Framework* (GMF) [39,40] and the *Generic Modeling Environment* (GME) [41,42] had brought important contributions. They are mature frameworks based on MDE concepts and furnish tools for defining grammars and generating code for graphical editors.

GMF provides a set of capabilities and runtime infrastructures for generating graphical editors for DSMLs based on their metamodel definitions. Where GME allow decorating a metamodel of a DSML with entities called views. This gathers concepts that will be used to implement models, links between

those concepts and how the concepts will be organized and displayed by the graphical editor. Nevertheless, these frameworks, even if they make the generation of graphical editors of DSLMs much easier, they did not elaborate proper features to support the composition and reuse of DSMLs.

2.2. In AOP

The MDE was much inspired by AOP to deal with the problem of large models for complex systems [43]. The AOP preconizes to design a system by separating the model into different morsels. Each corresponds to a different aspect of the system. This decomposition permits to deal with properties on each aspect before considering the model in its overall. This way we decrease the analysis complexity [44]. However, this requires being able to integrate the morsels of a model with each other's. Thus, AOP has addressed the problem of composition and reuse of models [45].

Hovsepyan et al. [46] elaborated an asymmetric approach to compose artefacts of different DSMLs using an application base model described with UML. This approach was driven by an AOP methodology and was implemented using MDE tools. It introduced the concept of a concern interface which plays the role of a common language between a specific concern and the application base. The composition is then achieved by defining explicitly the syntactic and the semantic relationships between artifacts coming from different concerns.

LARA [47] is a DSL inspired by AOP concepts which brought a novel method for mapping applications to heterogeneous high performance embedded systems. It allows to generate an intermediate aspect representation from a configuration based on different junction points, action models and attributes. This is then given to be processed by the weavers. Pinto et al. [48] has improved *LARA* by furnishing well-defined library interfaces with concrete implementations for each supported target language. This work contributes to make *LARA* aspects more concise and improve their reuse. Moreover, it involve to substantial reductions of job effort when developing weavers for new languages.

MATA [49] is an AOP tool built on the top of IBM Rational Software Modeler. It uses model transformation to define and perform composition operations on aspects of a model. The particularity of *MATA* is that, even if it is inspired by AOP, it did not deal with specific join points. In fact, any model artifact can be considered a join point, and composition is implemented as a special case of model transformation. In addition, critical pair analysis is automatically applied in order to find structural correspondences between various aspects of models. *MATA* was intended to be a generic approach but it was above all proven on UML models (class diagrams, sequence diagrams and state diagrams) [50].

2.3. In LOP

The LOP field is rich in approaches that ease the design and the reuse of DSLMs. Ones of the most presumably technologies to perform it are the projectional language workbenches. In fact, they provide relevant approaches for extending a DSL and often furnish tools to project it on concrete spaces.

TouchRAM [51] provided a rich client tool for flexible software modelling. It enable at developing reusable and scalable design model through a large registry of design basic design models. It takes advantage from model interfaces and aspect oriented model weaving. The conception of a new design model can be obtained by the composition of available design models in the registry. This

work has been improved with *TouchCORE* [52] which furnish new features for model visualization, model editing model assessment and composition traceability.

MPS [53] provided capabilities to define a DSL trough many aspects: abstract aspects (metamodel), sematic aspects (constraints), concrete syntaxes aspects (graphical editors), generators aspects (model transformations) and many others (e.g. behavior, type system, data flow or intentions) [54]. *MPS* furnish two ways to reuse DSLs: the reference and the extension mechanisms. The reference consists to use concepts from a given DSL into another one. The extension allows extending a DSL from another one by creating new concepts that inherit all the properties and behavior of their parents [55].

MetaMod [56] is based on a metametamodel that provides metatools to ease the creation and the reuse of DSLs. Convicted that most of simple DSLs do not require more advanced modularity, *MetaMod* defines the modularity at the value model level provided by the metametamodel itself [57]. Furthermore, having the same modularity mechanisms in many DSLs lead to have robust DSLs, because easier to verify, more fit and easier to learn as well. In addition, this facilitates the reuse of DSLs. However, it limits capabilities of the DSLs if more advanced modularity mechanisms are needed.

Cedalion [58,59] is built on top of Prolog. It provides features for DSLs building with projectional editor trough the description of model aspects such as semantics, structure, projection, and type system definitions from other language workbenches. *Cedalion* proposes a DSL reuse mechanism. However, because of the close link between the structure of a language and its other aspects, this makes the reuse difficult in *Cedalion*. In fact, all language aspects of a DSL need to be reused. Nevertheless, extending a DSL with only additional concepts is thus effortless [60].

Spoofax [61] is a language workbench dedicated to design textual DSLs. The platform provides features for code generation, parsers, type checkers, compilers, interpreters, and other tools for language definitions. *Spoofax* furnishes an API for programmatically composing abstract and concrete syntax of a language. Within *Spoofax*, the management of modularity can be managed directly in target generated language [62].

Xtext [63] is a textual language workbench based on EMF. It provides tools to define textual DSLs. It furnishes a DSL reuse and extending mechanisms. Reuse permits to cross reference concepts between DSLs. Where, extending allows to a DSL to inherit from another one and to override its grammar rules. However, it allows only to completely overriding them. In addition, this extending mechanism limits a DSL to only extend one other. Regarding the dynamic semantics of the DSL, it can be implemented using other languages such as *Xtend* [64].

Monticore [65,66] is a textual language workbench. It provides modularity mechanisms that enable the compositional development of textual DSLs and their supporting tools. Especially, inheritance and embedding mechanisms are proposed. Inheritance allow to extend a language where embedding allow to compose different language fragments. Moreover, a special DSL is proposed for the definition of compositional links between languages.

Other works implemented approaches to provide DSMLs composition and reuse capabilities. Pedro et al. [67] have

contributed with an automatic projection approach from metamodels composition patterns into graphical syntax. In [68] they go further more with a definition of operators to compose DSMLs with a proposal for automatic mapping to graphical syntax. Meyers et al. [69,70] proposes a template based technique for the modular definition and composition of DSMLs, including their abstract syntax, semantics, and concrete syntax (relying on metaDepth [71]).

3. Problem Statement & Methodology

Software engineering is essentially involved in providing textual or graphical languages to describe and set out artefacts of a system; their structures, behaviors and interactions. DSMLs provide capabilities to achieve this and allow designers to handle these artefacts as models. Models are intended to be used by tools. Thus, it should be defined a formal description of their concepts. This well-established set of concepts is called a metamodel. This is the principle of DSMLs design. Accordingly, composing DSMLs is primarily a composition of their metamodels.

The composition of metamodels is special issue of a larger problem in MDE, model composition. The composition of the models is a topic of research in continual but very slow evolution in the MDE. This is partly due to the miss of inspiration of patterns from programming languages [72]. It also never has been the subject to standardization like model transformation.

We can define a composition of models simply with a composition operator \otimes which is a function producing a composed model C by using artifacts of two input models A and B:

$$\otimes : A \times B = C \quad (1)$$

However, model composition can scope various meaning and reach at least three dimensions: abstract syntax, concrete syntax, and semantics [73]. As a DSML is a modeling language we can take inspiration from the composition operation as it is defined in modeling languages; an association of sub languages into one integral language. Where sub artifacts are handled in their original languages and the composed artifact acquire its semantics and its syntax from the composition [74]. In addition, we can draw inspiration from the composition operation as it is defined in programming languages. There is a frank conjunction between the semantic unit (i.e. class) that has a specific interface and the syntactic unit (i.e. file) that is the encapsulation of the implementation [72]. When semantic units are composed, logically de facto the language tool composes the syntactic units. Therefore, a successful approach of composition of DSMLs must deal with the three composition dimensions and maintain the link between abstract syntax (metamodels), concrete syntax and semantics.

In the respect of the aforementioned, this work is an exploratory study whose purpose is to explore means of composing DSMLs by composing their metamodels and studies the projection on their associated graphical editors. Indeed, an appropriate reuse of their syntaxes and graphical editors can be performed. Defining the way that metamodels will be composed implies the way that syntaxes can be merged and editors can be reused. Furthermore, it explores how to automate the composition of graphical editors of the composed DSMLs by implementing a prototype of a composition rules based code generator facility. For that the

proposed exploratory study is segmented into five steps as described in Figure 1.

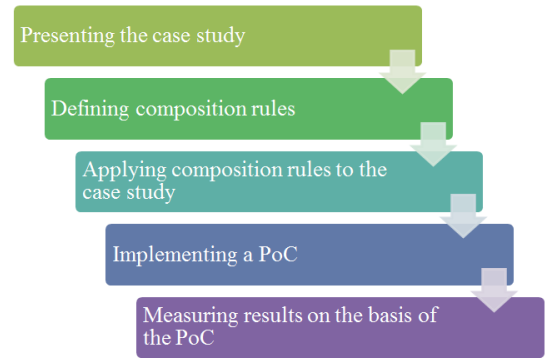


Figure 1. The process of the exploratory study.

A case study is first described. Then the rules for composition of metamodels are defined; on the basis of which DSMLs artifacts can be reused. Later, each is applied to a use case from the case study in order to illustrate it. Next, to prove the concept of this work, an implemented prototype of code generator facility for extending new DSMLs based on the aforementioned composition rules is presented. Finally, three parameters as indicators of evaluation and performance are used to assess the contribution of this study:

- The gain in terms of saved development time that can be estimated via the percentage of the generated code.
- The gain in terms of reused components that can be estimated via the percentage of reused code.
- The gain in terms of learnability, that can be estimated via the part of kept features and interfaces.

These three parameters are measured and discussed in the Section 5.

4. Exploratory Study

4.1. Case Study

Figure 2 represents excerpts of three simple metamodels representing small DSMLs. Each metamodel relies on a different concept.

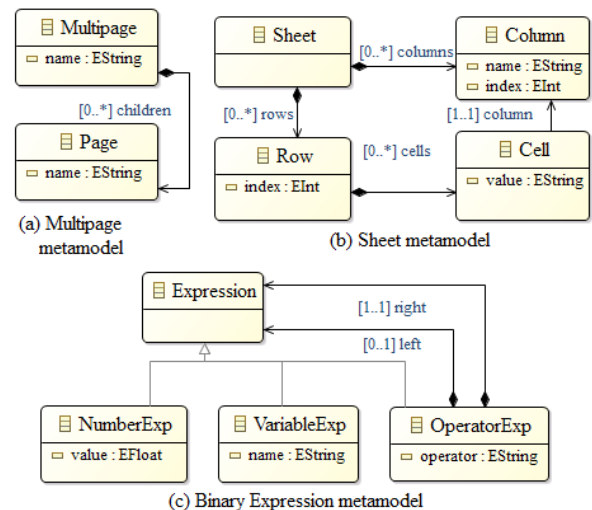


Figure 2. Multipage, Sheet and Expression metamodels.

The first metamodel is the Multipage metamodel ($MM_{MultiPage}$). It can be used to describe a multiple page structure. It allows multiple pages to be contained under a single parent page. According to (a), an instance of the metaclass *MultiPage* may contain one or more children instances of the metaclass *Page*. The second metamodel is the Sheet metamodel (MM_{Sheet}). It can be used to describe a sheet containing a two dimension table. According to (b), an instance of the metaclass *Sheet* contains a collection of instances of the metaclass *Row*. Each of which containing a collection of instances of the metaclass *Cell*. In addition, the instance of *Sheet* defines instances of the metaclass *Column*. Each defined instance of *Cell* is related to one of them. The third metamodel is the Binary Expression metamodel (MM_{Exp}). It can be used to describe a binary expression tree which can contain numbers, variables, and unary or binary operators. According to (c), an instance of the metaclass *Expression* is a tree of nodes which can be instances of three metaclasses: *OperatorExp*, *IntegerExp* and *VariableExp*. The *OperatorExp* instances are contained in the internal nodes of the tree, where instances of *IntegerExp* and *VariableExp* are contained in the leaf nodes. Withal, an *OperatorExp* node may have two children nodes for binary operators (left and right expressions), or one child node for unary operators (only right expression). Each of these DSMLs metamodels relies on a graphical or textual syntax that allows expressing conforming models. Therewith, they are supported by graphical interfaces: editors. Figure 3 shows screen shots of the editors. The graphical syntax of the $MM_{MultiPage}$ DSML (a) expresses a *MultiPage* instance as a multiple tabs window. The children instances of *Page* are embedded as a sequence of tabs in the parent *MultiPage* instance. The associated editor has buttons to add new tabs. The graphical syntax of the MM_{Sheet} DSML (b) expresses an instance of *Sheet* as a two dimensions table with indexed instances of *Row* and named instances of *Column*. Instances of *Cell* are represented by the boxes of the table with their *values* inside. The associated editor permits to extend or reduce the table using a hold and move button. This way, the editor allows creating new instances of *Row* and *Column* or deleting existing ones. *Indexes* of *Row* instances are represented at the left side of the table. Editing *names* of instances of *Column* and *values* of instances of *Cell* is done via the textual edition of the related boxes. The textual syntax of the MM_{Exp} DSML (c) expresses an instance of *Expression* using a mathematical grammar where parentheses represent internal nodes. The associated editor is a textual file editor with syntax highlighting.

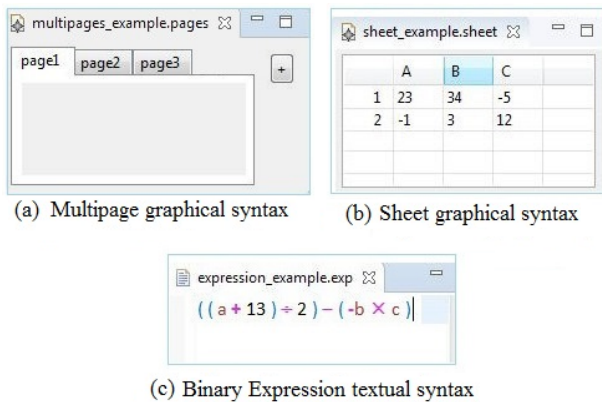


Figure 3. Multipage, Sheet and Expression concrete syntaxes.

This case study is used later to create step by step a new DSML that meets the following requirements reusing DSMLs (a), (b) and (c):

www.astesj.com

- *RQ1*. A sheet cell must be able to contain a binary expression.
- *RQ2*. A binary expression defined inside a cell must be able to reference the value of another cell of the sheet's table.
- *RQ3*. A multiple page must be able to be composed of multiple sheet tabs.

In the following Subsections three rules for composing metamodels are defined. Next, they are illustrated relying on the above requirements. Each time a requirement is fulfilled it uses an application of a defined rule. Besides, an investigation of the reuse of the syntax and graphical artifacts of original DSMLs is realized. It aims to obtain an extended DSMLs based on the performed composition of metamodels expressed by means of the proposed composition rules.

4.2. Composition Rules

To describe a composition rule, the following formalism is used :

$$\otimes \text{Rule: } MM_A \times MM_B (\text{arguments } \dots) = MM_C \quad (2)$$

Where;

- MM_A and MM_B are metamodels to be composed.
- $\otimes \text{Rule}$ is the composition rule.
- MM_C is the composed metamodel.

Reference Rule

A reference rule allows the establishment of discrete connections between instances of a model, conforming to MM_C , defined by concepts coming from MM_A and MM_B . It defines an oriented binary association in MM_C from a metaclass MT_1 of MM_A toward a metaclass MT_2 of MM_B . It is used to connect one instance of MT_1 to many instances of MT_2 . It could be a simple link, an aggregation or a composition. It must specify *multiplicity* to mean how many instances of MT_2 can be referenced from an instance of MT_1 . The $\otimes \text{Ref}$ rule can be defined as follows :

$$\otimes \text{Ref: } MM_A \times MM_B (MT_1, MT_2, [m_1, m_2], c_1, c_2) = MM_C \quad (3)$$

- $[m_1, m_2]$ are integers to express multiplicity with a minimum value m_1 and a maximum value m_2 .
- c_1 is a Boolean value to mean whether the reference expresses a containment association (i.e. aggregation).
- c_2 is a Boolean value to mean whether the reference expresses a container association (i.e. composition).
- MM_C is the composed metamodel.

Specialization Rule

The specialization rule permits to compose metamodels with an inheritance concept similar to the concept of specialization in object oriented programming. It allows to a metaclass MT_1 from a metamodel MM_A to acquire all the properties and behaviors of another metaclass MT_2 from a metamodel MM_B . Thus, attributes,

associations, or methods can be reused. The $\otimes Spe$ rule can be defined as follows :

$$\otimes Spe: MM_A \times MM_B (MT_1, MT_2) = MM_C \quad (4)$$

Fusion Rule

The fusion rule is used to bind metaclasses coming from different metamodels in order to fusion them in the composed metamodel. It allows a metaclass MT_1 from a metamodel MM_A and a metaclass MT_2 from a metamodel MM_B to form a new hybrid metaclass MT_3 in the composed metamodel through over a customized binding. The binding defines the matching between the properties of metaclasses MT_1 and MT_2 : attributes references and methods. In addition it specifies those to keep and those to delete. The $\otimes Fus$ rule can be defined as follows:

$$\otimes Fus: MM_A \times MM_B (MT_1, MT_2, \{bindings\}) = MM_C \quad (5)$$

4.3. Rules Application

A $\otimes Ref$ rule can be applied to fulfill the requirement RQ1. Considering that MT_1 is the *Cell* metaclass as defined in the MM_{Sheet} metamodel and MT_2 is the *Expression* metaclass as defined in the MM_{Exp} metamodel. The $\otimes Ref$ rule can be applied as follows:

$$\otimes Ref_{expression}: MM_{Sheet} \times MM_{Exp} (Cell, Expression, [0, 1], true, true) = MM_{C1} \quad (6)$$

By applying the $\otimes Ref_{expression}$ rule, the composed metamodel MM_{C1} is obtained. MM_{C1} is shown in Figure 4, where the $\otimes Ref_{expression}$ rule is represented with the bold line starting with a lozenge. The designed composition allows an instance of *Cell* to contain an instance of *Expression*. The achieved composition is projected in order to create a new extended graphical editor for the new DSML (d). It is based on the MM_{C1} metamodel and reuses the graphical artifacts of DSMLs (b) and (c). A mock-up of the extended editor of (d) is shown in Figure 4 where the textual editor of the DSML (c) is included in the top of the editor of the DSML (b). According to (d), a sheet's cell is able to contain the value of a binary expression. Therefore, the cell's value is the computed value of an expression which can be edited in the top textual editor.

Similarly, another $\otimes Ref$ rule can be applied to fulfill the requirement RQ2. However, this time the $\otimes Ref$ rule has to define a simple link reference, from the metaclass *VariableExp* toward the metaclass *Cell*. The $\otimes Ref$ rule can be applied as follows :

$$\otimes Ref_{refersTo}: MM_{C1} \times MM_{C1} (VariableExp, Cell, [0, 1], false, false) = MM_{C2} \quad (7)$$

By applying the $\otimes Ref_{refersTo}$ rule, the composed metamodel MM_{C2} is obtained. MM_{C2} is shown in Figure 4, where the $\otimes Ref_{refersTo}$ rule is represented with the bold arrow. The designed composition allows the definition of cross references between cells expression. In this way, an instance of *Expression* contained inside an instance of *Cell* can reference the value of another instance of *Cell* present in the table. Explained otherwise, an instance of *Expression* which is structured as a tree can include in its nodes a

reference to an instance of *Cell* through an instance of *VariableExp*. It is important to observe that this design adapts the *Cell*'s instances to behave as *Expression*'s instances. It is worth mentioning that such pattern, applied with the $\otimes Ref$ rule can be useful to solve situations where it is need to adapt concepts of different metamodels when composing them. The achieved composition is projected in order to create a new extended graphical editor for the new DSML (e). It is based on the MM_{C2} metamodel and reuses the graphical artifacts of the DSML (d). A mock-up of the extended editor of (e) is shown in Figure 4. According to (e), a binary expression defined inside a cell must be able to reference the value of another cell of the sheet's table. Therefore, the cell's value is the computed value of an expression which can use the computed values of expressions defined elsewhere in the sheet.

A $\otimes Spe$ rule can be applied to fulfill the requirement RQ3. Considering that MT_1 is the *Sheet* metaclass as defined in the MM_{C2} metamodel and MT_2 is the *Page* metaclass as defined in the $MM_{Multipage}$ metamodel. The $\otimes Spe$ rule can be applied as follows:

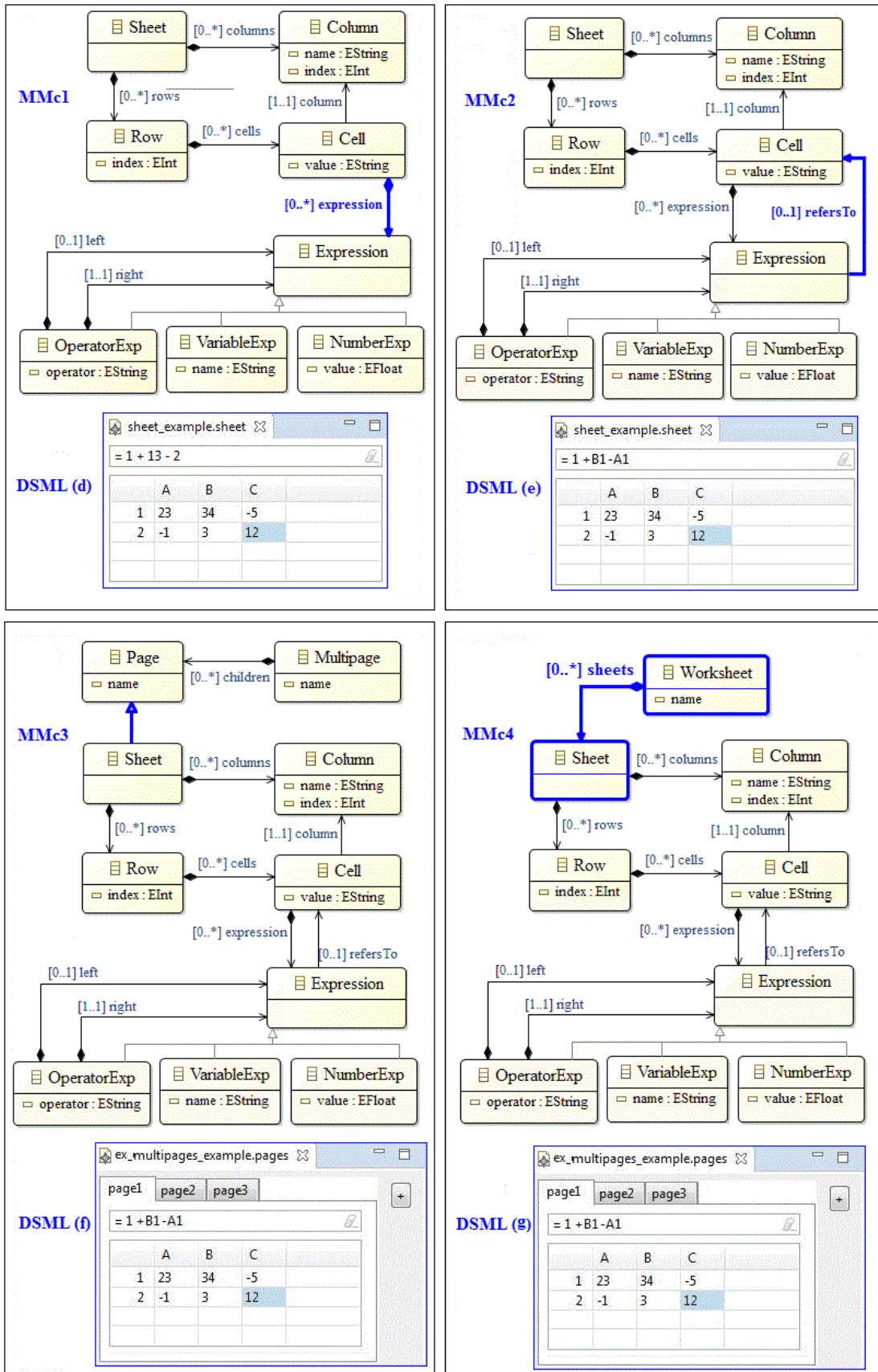
$$\otimes Spe_{page}: MM_{C2} \times MM_{Multipage} (Sheet, Page) = MM_{C3} \quad (8)$$

By applying the $\otimes Spe_{page}$ rule, the composed metamodel MM_{C3} is obtained. MM_{C3} is shown in Figure 4, where the $\otimes Spe_{page}$ rule is represented with the bold arrow. The designed composition allows a sheet to be a kind of page and then to be a candidate to be a tab of the multiple page. In this way an instance of *Multipage* can contain instances of *Sheet*. Therefore, a multiple page is able to be composed of multiple sheet tabs. The achieved composition is projected in order to create a new extended graphical editor for the new DSML (f). It is based on the MM_{C3} metamodel and reuses the graphical artifacts of the DSML (f). A mock-up of the extended editor of (f) is shown in Figure 4. According to (f), it is possible to create multiple tabs of sheets using the means of the *multipage* editor; i.e. the button that creates pages. The graphical interface of a sheet is then embedded in the graphical container provided for a page and behaves autonomously. However a question may arise about the semantic and utility of the metaclass *Page* in MM_{C3} . The answer depends on the understanding of the requirement RQ3. If it requires obtaining a multiple pages editor composed "only" of sheet tabs. It is probably cleaner to merge metaclasses *Page* and *Sheet*. Moreover, it would be clearer to give a new semantic to the metaclass *Multipage* to indicate that it is a multiple sheets tabs. This leads to define a new rule: the fusion rule.

A $\otimes Fus$ rule can fulfill the requirement RQ3 in case it requires obtaining a multiple pages editor composed only of sheet tabs. Considering that MT_1 is the *Sheet* metaclass as defined in the MM_{C3} metamodel and MT_2 is the *Page* metaclass as defined in the $MM_{Multipage}$ metamodel. The $\otimes Fus$ rule can be applied as follows:

$$\otimes Fus_{sheet}: MM_{C2} \times MM_{Multipage} (Sheet, Page, \{Sheet.name, Page.name\}) = MM_{C4} \quad (9)$$

By applying the $\otimes Fus_{sheet}$ rule, the composed metamodel MM_{C4} is obtained. MM_{C4} is shown in Figure 4, where the $\otimes Fus_{sheet}$ rule had the apparent effect to superimpose the metaclasses *Page* and *Sheet* in one metaclass. Additionally, the metaclass *Multipage* can be renamed to *Worksheet* in order to give a better representative name to the container of sheet tabs. The achieved composition projected in order to create the extended graphical editor for the



The applications of composition rules on the case study.

new DSML. It is based on the *MMC4* and leads to the graphical editor of the DSML (g). It very close to the DSML (f) except that the content of the multiple sheets (worksheet) can only be sheet tabs. According to this new DSML, an instance of *Worksheet* can be composed, and only composed, of multiple instances of *Sheet*.

4.4. Proof of Concept

In order to validate the approach exposed in this paper and going further than the theoretical exposition, a Proof of Concept (PoC) was achieved. For this purpose, a prototype of code generator facility based on the aforementioned composition rules was implemented. Then it was applied to our case study. However, before doing so, it is necessary to implement the DSMLs (a) and (b) used in the case study. Next, the metamodels of these DSMLs are composed using one of the rules previously defined. Finally the implemented prototype is used to project the composition onto the graphical editors of DSMLs.

EMF is used to implement the PoC. EMF allows the generation of class architecture that represents metamodel concepts. EMF does not only generate Java classes, but also an associated infrastructure. Thus, one benefits from the persistence of the model in XML Metadata Interchange (XMI) [75] format, but also from a set of tools to handle the model completely independent from the objects it contains. This infrastructure makes it possible to build higher level tools for processing models created with EMF. Within this framework, one of the functionalities is notably the visualization and the edition of models thanks to the EMF Edit framework. Using the capabilities of EMF, the prototype of the PoC aims to generate an overlay layer of code following the EMF code generation. It must provide the necessary infrastructure to make quick building of new DSMLs based on the composition of their metamodels possible.

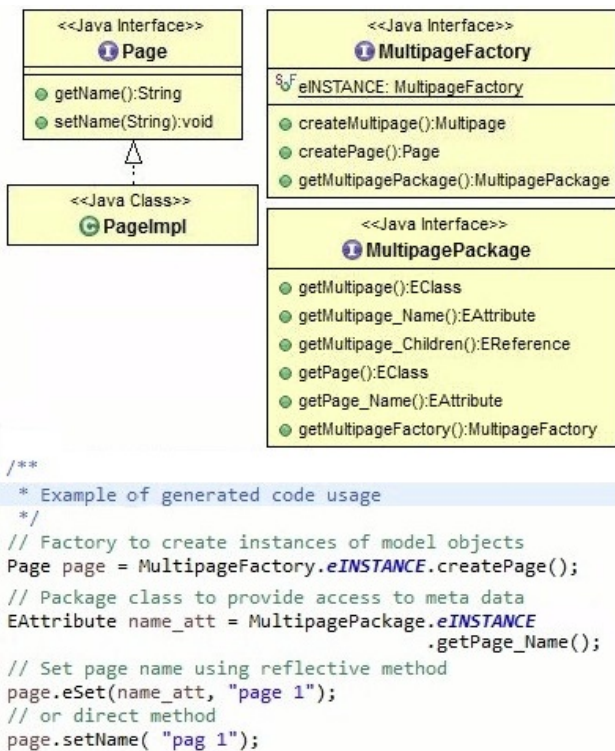


Figure 5. A class diagram representing EMF Impl generated classes for the metaclass Page and an example of code usage.

The first step is obviously to implement metamodels. The PoC use case relies only on the two metamodels *Multipage* and *Sheet*. The EMF essential MOF [76] implementation (Ecore) is used to describe metamodels. As aforementioned, EMF provides a code generator facility to generate Java code from a metamodel described in Ecore. It generates, inter-alia, two based packages: One for model implementation (EMF Impl) and another one for graphical user interface editing (EMF Edit). Figure 5 shows an excerpt of the classes generated for the implementation of the *Multipage* metamodel and an example of code usage for creating and manipulating a *Page* instance.

Furthermore, EMF Edit provides capabilities to build a graphical editor. It enables the visualization of model elements and their command-based editing. Figure 6 shows an overview of the architecture of a graphical editor build using the EMF Edit generated code. This will be needed in order to understand the solution exposed further in the paper since it extends the mechanism of EMF Edit. The generated code includes:

- *ItemProvider(s)*: They are generated for each class of the metamodel. They are used to display model elements in a graphical editor via an Adapter design pattern (a delegation mechanism that makes it possible to "act as if" an object of type A was an object of type B).
- *ItemProviderAdapterFactory*: It is used to group all *ItemProvider* classes and provide a centralized mechanism to request them.
- *ContentProvider(s)* and *LabelProvider(s)*: They are used to provide the display of an item. A *ContentProvider* retrieves the content of an item displayed by a graphical interface where the *LabelProvider* takes care of the visualization (image and text) of the item. The *ContentProvider(s)* and *LabelProvider(s)* can (and usually should) delegate to the same *AdapterFactory* and, therefore, to the same *ItemProvider(s)*.
- *ComposedAdapterFactory*: It is useful in order to stick different adapter factories (for individual models).
- *EditingDomain*: It is an editing command structure, including a set of generic command implementation classes to build editors that fully support, cancel and redo actions.

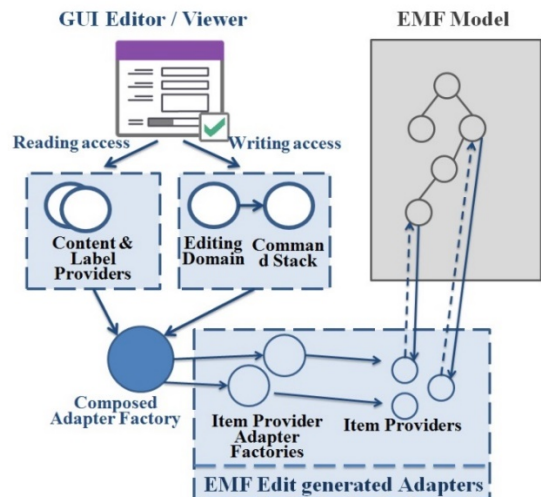


Figure 6. The architecture of a graphical editor built using EMF Edit.

Figure 7 shows a class diagram representing the EMF Edit generated code for the *Page* class from the *Multipage* metamodel.

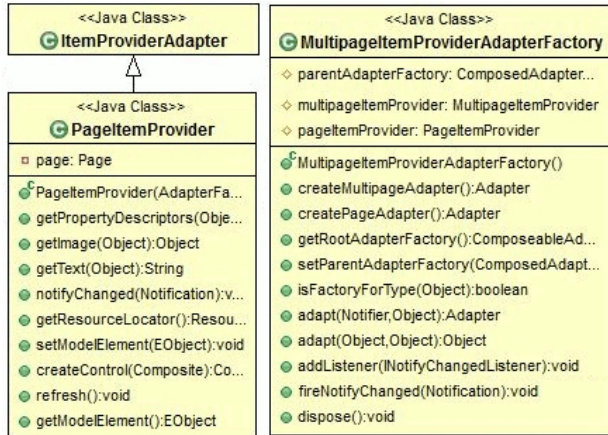


Figure 7. A class diagram representing EMF Edit generated classes for the metaclass *Page*.

8 shows an excerpt of the *Sheet* editor implementation. It is important to mention that the provision of the graphical content for each element of the model was centralized in a *createControl()* method attached to its adapter *ItemProvider*. The resulted editors match the screenshots shown in Figure 3.

The code generator facility

As mentioned earlier, the PoC aims to implement a code generator facility that allows generating a Java code overlay of the generated EMF Edit code. It must provide the necessary infrastructure to make it possible to quickly build (compose) new DSMLs based on the composition of their metamodels; using the composition rules defined in Subsection 4.2.

```

/**
 * This is an example of a Sheet model editor.
 */
public class SheetEditor extends EditorPart implements IResourceChangeListener, IResourceDeltaVisitor {
    ... ..
    public void createPartControl(Composite parent) {
        // This is the one adapter factory used for providing views of the model.
        ComposedAdapterFactory adapterFactory = new ....
        SheetItemProvider sItemItemprovider = new SheetItemProvider(adapterFactory);
        adapterFactory.addAdapterFactory(sItemItemprovider);
        adapterFactory.addAdapterFactory(...);
        ... ..
        // Load the resource through the editing domain.
        Sheet sheet = getEditingDomain().getResourceSet().getResource(resourceURI, true).getContents()....;
        // Attach the model element to its item provider
        sItemItemprovider.setModelElement(sheet);
        ... ..
        // create the Sheet GUI component
        adapterFactory.adapt(sheet, SheetItemProvider.class).createControl(parent);
    }
    ... ..
}

/**
 * This is the item provider adapter for a Sheet object.
 * @generated
 */
public class SheetItemProvider extends PageItemProvider {

    /** @generated not */
    public Composite createControl(Composite parent) {
        //Content & Label Provider initialisation
        ITreeContentProvider contentProvider = new SheetViewTreeContentProvider();
        DelegatingStyledCellLabelProvider labelProvider = new DelegatingStyledCellLabelProvider(...);
        Composite content = new Composite(parent, SWT.NONE);
        ... ..
        // create the model element GUI component
        TreeViewer treeViewer = new TreeViewer();
        treeViewer.setContentProvider(contentProvider);
        treeViewer.setInput((Sheet)getModelElement());

        // Iterate over the sheet element to create columns and rows using Content and Label Provides
        ... ..
        return content;
    }
    ... ..
}
    
```

Figure 8. An excerpt from the code of the Sheet Editor.

Therefore, the adapters mechanism used by EMF Edit is extended with the definition of a Java interface called *IExtendedGraphicalItemProvider*. It specifies a contract of five methods sufficient to create a graphical component connected to an element of the model, to refresh it and to dispose it:

- *createControl()*: It centralizes the provision of the graphical content of the model element.
- *setModelElement()*: It attaches the model element to its provider.
- *getModelElement()*: It accesses the model element from the provider.
- *refresh()*: It refreshes the graphical content of the model element.
- *dispose()*: It disposes the graphical content of the model element.

The implemented code generator facility modifies each edit *ItemAdapter* class, already generated by EMF Edit, in order to make it extend the *IExtendedGraphicalItemProvider* interface. These methods must be implemented and used for the construction of a new graphical editor. So far nothing new compared to the classic use of EMF Edit adapters. Now, if a composition rule is applied between two metamodels, these extended adapters come into play with the use of the aforementioned methods.

Let us consider the following example to better illustrate these statements. Let MM_A and MM_B be two metamodels related to two DSMLs (α) and (β). Let M_A be a model conforming to MM_A and M_B be a model conforming to MM_B . Let $\otimes Rule_C$ be a composition rule MM_A and MM_B in order to create a new DSML (ϑ). Considering that $\otimes Rule_C$ implies that an element A of the model M_A has to be linked to an element B of the model M_B . A graphical editor built using the architecture shown in Figure 6 makes that the *ItemProviderAdapterFactory* calls the *ItemProvider* IP_A related to the element A to display it in the editor of (α). Likewise, the *ItemProviderAdapterFactory* calls the *ItemProvider* IP_B related to the element B to display it in the editor of (β). In a new architecture built using the extended code generator facility, the generated adapters IP_A and IP_B will be linked with a generated link which reflects the $\otimes Rule_C$ rule. Thus, IP_A will directly call IP_B for displaying element B in the editor of (ϑ). Figure 9 schematizes this new architecture. For example, a specialization rule will imply, at the generated code, inheritance between IP_A and IP_B . Whereas containment reference rule will imply encapsulation of methods (defined by the interface *IExtendedGraphicalItemProvider*) of IP_B by those of IP_A .

Demonstration of the generator

Let us return back to our case study to illustrate the extended code generator facility through a second example. Let us consider the composition rule $\otimes Spe_{page}$ outlined in Subsection 4.3. The rule was applied on the implemented Ecore metamodels *Sheet* and *Multipage*. Indeed, Ecore makes it possible to describe a link of specialization between two metaclasses of two different metamodels. Then, the EMF generator facility was used to generate the EMF Impl code and the EMF Edit code. Finally, the implemented code generator prototype was used to generate the extension layer with the *ItemProvider*(s) that extend the *IExtendedGraphicalItemProvider* interface. Figure 10 shows a class diagram representing the generated *ItemProvider*(s).

The generated code was used to re-implement the graphical editor of the composed DSML resulting of $\otimes Spe_{page}$. An extended

Multipage editor has been obtained. It allows a multiple page to be composed of multiple sheet tabs. Figure 11 shows an excerpt from the code of the extended *Multipage* editor. It demonstrates how the *PageItemProvider* delegates the creation of the graphical component of an instance of *Sheet*, included under a *Multipage* instance, to a *SheetItemProvider*. It takes advantage of the polymorphism between the *PageItemProvider* and the *SheetItemProvider* to call the appropriate graphical interface creation method. In the same way, the code of the *MultipageItemProvider* demonstrates how the *refresh* and *dispose* methods can be called for each instance of *Page* contained in an instance of *Multipage*. It takes advantage of polymorphism to apply the appropriate method depending on whether the displayed instance is an instance of *Page* or an instance of *Sheet*.

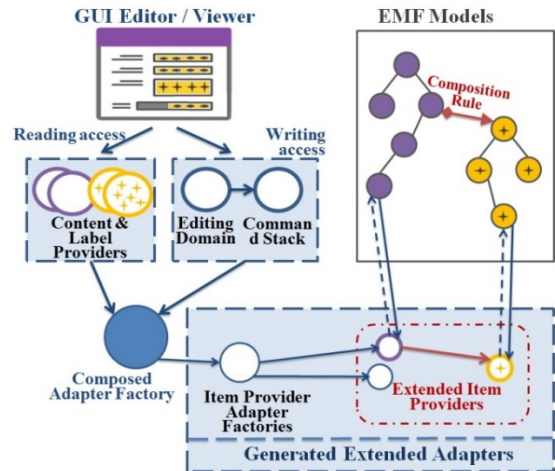


Figure 9. The architecture of a graphical editor built using the extended code generator facility.

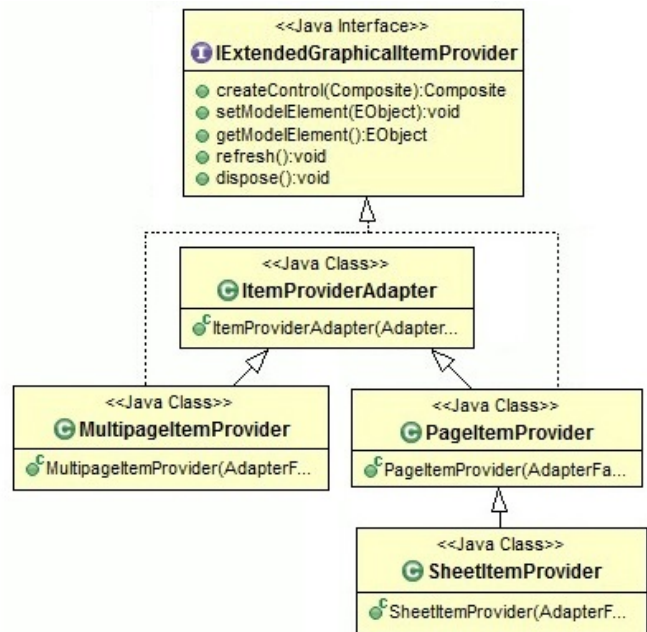


Figure 10. A class diagram representing the generated *Item Provider*(s) using the extended code generator facility.

In this paper we have conducted an exploratory study whose goal is to explore means of composing DSMLs by composing their

```

/**
 * This is an example of an Extended Multipage model editor.
 */
public class ExtendedMultipageEditor extends EditorPart
    implements IResourceChangeListener, IResourceDeltaVisitor {
    ... ..
public void createPartControl(Composite parent) {
    // This is the one adapter factory used for providing views of the model.
    ComposedAdapterFactory adapterFactory = new ...
    MultipageItemProvider mItemItemprovider = new MultipageItemProvider(adapterFactory);
    adapterFactory.addAdapterFactory(mItemItemprovider);
    adapterFactory.addAdapterFactory(...);
    ... ..
    // Load the resource through the editing domain.
    Multipage multipage = getEditingDomain().getResourceSet()...;
    // Attach the model element to its item provider
    mItemItemprovider.setModelElement(multipage);
    ... ..
    // create the Multipage GUI component
    tabFolder = (TabFolder) mItemItemprovider.createControl(parent);
    for (Page page : multipage.getChildren()) {
        String tabItemName = page.getName();
        TabItem tabItem = new TabItem(tabFolder, SWT.NONE);
        tabItem.setText(tabItemName);
        PageItemProvider pItemItemprovider = (PageItemProvider) new SheetItemProviderAdapterFactory().
            getAdaptedProvider(page, adapterFactory);

        pItemItemprovider.setModelElement(page);

        // The Page Item Provider delegates the creation of the GUI component to a Sheet Item Provider
        // in order to create a Sheet GUI component
        Composite tabItemComposite = pItemItemprovider.createControl(tabFolder);
        tabItem.setControl(tabItemComposite);
    }
}
... ..
}
/**
 * This is the item provider adapter for a Multipage object.
 * @generated
 */
public class MultipageItemProvider extends ItemProviderAdapter
    implements IItemPropertySource, IItemLabelProvider,
        IStructuredItemContentProvider, ITreeItemContentProvider,
        IExtendedGraphicalItemProvider, EditingDomainItemProvider{
    ... ..
/**
 * @generated
 */
@Override
public void refresh() {
    for (Page page : multipage.getChildren()) {
        // It delegates to the Sheet Item Provider if the page is an instance of Sheet
        adapterFactory.adapt(page, PageItemProvider.class).refresh(parent);
    }
}
/**
 * @generated
 */
@Override
public void dispose() {
    for (Page page : multipage.getChildren()) {
        // It delegates to the Sheet Item Provider if the page is an instance of Sheet
        adapterFactory.adapt(page, PageItemProvider.class).dispose(parent);
    }
}
... ..
}
}

```

Figure 11. An excerpt from the code of the Extended Multipage Editor.

metamodels and studies the projection on their associated graphical editors. Three rules of composition were defined: Reference, Specification and Fusion. Each has been applied to a case study to illustrate its use. The impact of this composition on graphic editors has been studied. We have shown how the actual syntaxes of the originals DSMLs have been reused and composed to give shape to the concrete syntax of the composed DSML.

5.1. Development Time

Saving time is saving money. Development time is a major factor in software development. Indeed, with the multitude of technology and the vertiginous speed with which languages of programming develop. It is essential to minimize the development time of new software. It is one of the MDE's battle horses as it introduces the necessary abstraction to safeguard knowledge and automate the projection to technological spaces. In our exploratory study we have implemented a code generation prototype that allows taking advantage of our composition rules. It allows generating a layer of code that facilitates the composition of the concrete syntaxes of composed DSMLs. We have implemented it on our case study and we have shown an example in the previous section. After the method usage, we can measure about some preliminary results about the gain obtained by our approach in term of development time by measuring the percentage of generated of code. Because the percentage of code generated is directly correlated to the development time earned. Indeed, the less time spent writing the automatically generated code represents a time gained directly on the development time. In addition, we measure this value on both EMF Edit and our prototype. In this way we show what we also gain compared to the EMF Edit Framework.

Table 1 presents a comparison results summary. The first column lists the global number of line of code of each DSML. The second column compares the number of line of generated code by EMF Edit and by our generator facility. It worth to note that our generator is only used after a composition. Therefore, it has been used only for DSMLs (d), (e), (f) and (g). The third column shows a percentage comparing between the two tools. This last result is exploited in Fig. 12 to show by interpolation the potential gain in terms of development time. It is important to remember that our generator is an overlay of EMF Edit.

5.2. Code Reuse

One of our stated objectives in this study was the reuse of software components. We explored the reuse of existing DSMLs to extend or compose new ones. This reuse is reflected on the code of the obtained DSML. Thus we measured the percentage of reused code each time we extended our DSML in the application of the exploratory study to the case study.

Table 2 shows the percentage of code reuse each time we extended our DSMLs of the case study. It represents what we reused after applying each composition rule.

5.3. Learnability

In our case study, 100% of the graphic components of the original DSMLs were reused. Very few new graphical features have been introduced in composed DSMLs. This is a very important factor for the ease of learning of users. The learnability of software is often overlooked. However, it is the most influential aspect leading to the success of a software application.

In [78], authors noted that experience with similar software is a major dimension of learnability.

Table 1. Comparison between EMF Edit and Our Generator.

	lines of code	generated Lines		% of generated code	
		EMF Edit	Our Generator	EMF Edit	Our Generator
DSML(a)	1108	803		72%	
DSML(b)	2637	2369		90%	
DSML(c)	2735	1112		41%	
DSML(d)	3581	1532	3481	43%	97%
DSML(e)	3631	1532	3481	42%	96%
DSML(f)	4589	1916	4309	42%	94%
DSML(g)	4535	1901	4264	42%	94%

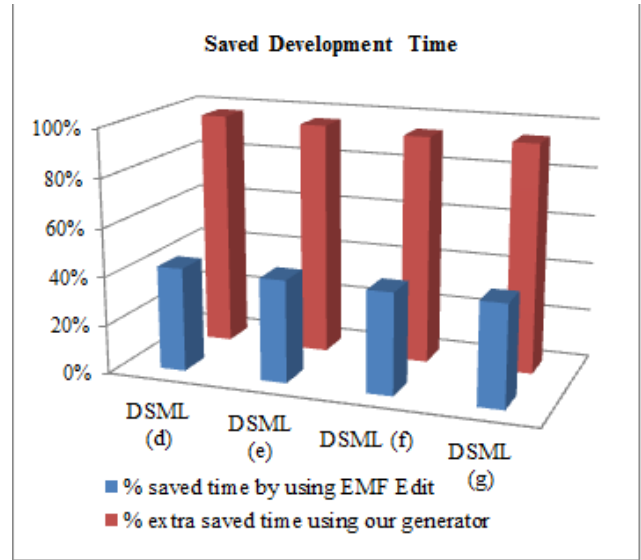


Figure 12. Gain in terms of development time.

Table 2. Percentage of reused after extending DSMLs.

	lines of code	% of reused code
DSML(d)	3581	97%
DSML(e)	3631	99%
DSML(f)	4589	97%
DSML(g)	4535	97%

6. Conclusions

This paper investigated the problem of extending DSMLs by composition their metamodels through an exploratory study. It exposes how DSMLs can be reused to rapidly create new ones with low cost. For this purpose three rules to compose DSMLs metamodels were specified: reference, specialization and fusion. A case study was used to illustrate the approach. In addition, the paper presented the implementation of a prototype of a code generator facility based on the aforesaid three composition rules. This prototype is then applied to the case study in order to validate our approach and measure its advantages. Compared to other works, our approach presents advantages, mainly by providing a higher level of reuse of DSMLs artefacts and by providing an automatic generation that facilitates the implementation of DSMLs tools and save development time. In addition, it keeps the graphics interfaces of the original DSMLs thus significantly improving the ease of learning of the new DSMLs.

The main contributions of the paper were: (i) the evaluation of the approach through an exploratory method; and (ii) the

implementation and the experimentation of the code generator facility prototype. Nevertheless, this work is only at its beginning. Indeed, it can be interesting to enlarge the set of composition rules, getting inspired by other principles and patterns coming from modeling languages and programming languages such as: encapsulation, substitution, adaptation and many others. Moreover, it can be interesting to take into account the composability properties of metamodels. Otherwise, the case study used in this study is very simple. It is a choice of writers to better illustrate the approach. However, it can exaggerate the results obtained from the fact of this simplicity.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

We express our respected gratitude goes to Ibn Zohr University, LabSIV laboratory, Ibn Zohr Doctoral Study Center and Faculty of Sciences of Ibn Zohr University for funding this research..

References

[1] A. Abouzahra, A. Sabraoui, K. Afdel, "A Metamodel Composition Driven Approach to Design New Domain Specific Modeling Languages" in 1st European Conference on Electrical Engineering and Computer Science, Bern, Switzerland, 2017. <https://doi.org/10.1109/EECS.2017.30>

[2] T. Mens, "A State-of-the-Art Survey on Software Merging" *IEEE Trans. Softw. Eng.* 28(5), 449-462, 2002. <https://doi.org/10.1109/TSE.2002.1000449>

[3] S. Kent, "Model Driven Engineering" *Lect. Notes. Comput. Sc.*, 2335, 286-298, 2002. https://doi.org/10.1007/3-540-47884-1_16

[4] D. C. Schmidt, "Guest Editor's Introduction: Model-Driven Engineering" *Computer.*, 39(2), 25-31, 2006. <https://doi.org/10.1109/MC.2006.58>

[5] R. Reddy, R. France, S. Ghosh, F. Fleurey, B. Baudry, "Providing Support for Model Composition in Metamodels" in 11th IEEE International Enterprise Distributed Object Computing Conference, Annapolis, MD, USA, 2007. <https://doi.org/10.1109/EDOC.2007.55>

[6] J. Estublier, G. Vega, A. D. Ionita, "Composing Domain-Specific Languages for Wide-Scope Software Engineering Applications" *Lect. Notes. Comput. Sc.*, 3713, 69-83, 2005. https://doi.org/10.1007/11557432_6

[7] F. Fleurey, B. Baudry, R. France, S. Ghosh, "A Generic Approach for Automatic Model Composition" *Lect. Notes. Comput. Sc.*, 5002, 7-15, 2008. https://doi.org/10.1007/978-3-540-69073-3_2

[8] J. Bézin, R. F. Paige, U. Assmann, B. Rumpe, D. C. Schmidt, "Manifesto - Model Engineering for Complex Systems" *Dagstuhl Seminar Proceedings*, 08331, 2008.

[9] D. S. Kolovos, L. M. Rose, N. Matragkas, R. F. Paige, E. Guerra, J. S. Cuadrado, J. De Lara, I. Ráth, D. Varró, M. Tisi, J. Cabot, "A research roadmap towards achieving scalability in model driven engineering" in *Workshop on Scalability in Model Driven Engineering (BigMDE '13)*, New York, USA, 2013. <https://doi.org/10.1145/2487766.2487768>

[10] B. Rumpe, "Towards model and language composition" in 1st Workshop on the Globalization of Domain Specific Languages (GlobalDSL '13), New York, USA, 2013. <https://doi.org/10.1145/2489812.2489814>

[11] A. Horst, B., Rumpe, "Towards Compositional Domain Specific Languages" *Ceur. Workshop. Procee.*, 1112, 7-16, 2013.

[12] U. Hohenstein and C. Elsner, "Model-driven development versus aspect-oriented programming a case study" in 9th International Conference on Software Paradigm Trends (ICSOFT-PT), Vienna, Austria, 2014. <https://doi.org/10.5220/0004999901330144>

[13] S. Dmitriev, Language oriented programming - the next programming paradigm, <http://www.onboard.jetbrains.com/articles/04/10/lop/>, accessed 21 November 2018.

[14] J.A. Pereira, K. Constantino, E. Figueiredo, "A Systematic Literature Review of Software Product Line Management Tools" *Lect. Notes. Comput. Sc.*, 8919, 73-89, 2014. https://doi.org/10.1007/978-3-319-14130-5_6

[15] J. White, J. H. Hill, J. Gray, S. Tambe, A. S. Gokhale, D. C. Schmidt, "Improving Domain-Specific Language Reuse with Software Product Line Techniques" *IEEE Software.*, 26(4), 47-53, 2009. <https://doi.org/10.1109/MS.2009.95>

[16] The Epsilon Homepage, <https://www.eclipse.org/epsilon/>, accessed 21 November 2018.

[17] The Epsilon Merging Language (EML) Homepage, <http://www.eclipse.org/epsilon/doc/eml/>, accessed 21 November 2018.

[18] D. S. Kolovos, R. F. Paige, F. A. C. Polack, "Merging models with the epsilon merging language (EML)" *Lect. Notes. Comput. Sc.*, 4199, 215-229, 2006. https://doi.org/10.1007/11880240_16

[19] D. Kolovos, "Merging Models with the Epsilon Merging Language - A Decade Later", in 19th ACM/IEEE International Conference on Model Driven Engineering languages and Systems, Saint-Malo, France, 2016.

[20] M. D. Del Fabro, P. Valduriez, "Towards the efficient development of model transformations using model weaving and matching transformations" *Softw. Syst. Model.*, 8(3), 305-324, 2009. <https://doi.org/10.1007/s10270-008-0094-z>

[21] The MOMENT web site, <http://moment.dsic.upv.es/>, accessed 21 November 2018.

[22] A. Boronat, "MOMENT: A Formal Framework for Model management" Ph.D Thesis, Universitat Politècnica de València, 2007.

[23] The Eclipse Modeling Framework (EMF) Homepage, <http://www.eclipse.org/modeling/emf/>, accessed 21 November 2018.

[24] A. Boronat, J. A. Carsi, I. Ramos, "Automatic Support for Traceability in a Generic Model Management Framework" *Lect. Notes. Comput. Sc.*, 3748, 316-330, 2005. https://doi.org/10.1007/11581741_23

[25] QVT, The MOF Query/View/Transformation specification page, <http://www.omg.org/spec/QVT/>, accessed 21 November 2018.

[26] A. Boronat, J. A. Carsi, I. Ramos, P. Letelier, "Formal model merging applied to class diagram integration" *Electron. Notes Theor. Comput. Sci.*, 166, 5-26, 2007. <https://doi.org/10.1016/j.entcs.2006.06.013>

[27] T. Degueule, B. Combemale, A. Blouin, O. Barais, J.M. Jézéquel, "Melange: a meta-language for modular and reusable development of DSLs" in 2015 ACM SIGPLAN International Conference on Software Language Engineering (SLE 2015), New York, USA, 2015. <https://doi.org/10.1145/2814251.2814252>

[28] T. Degueule, B. Combemale, A. Blouin, O. Barais, J. M. Jézéquel, "Safemodell polymorphism for flexible modeling" *Comput. Lang. Syst. Struct.*, 49(C), 176-195, 2017. <https://doi.org/10.1016/j.cl.2016.09.001>

[29] S. Kelly, K. Lyytinen, M. Rossi, "Metaedit + a fully configurable multi-user and multi-tool case and came environment" *Lect. Notes. Comput. Sc.*, 1080, 1-21, 1996. https://doi.org/10.1007/3-540-61292-0_1

[30] S. Kelly, J. P. Tolvanen, *Domain-specific modeling: enabling full code generation*, John Wiley & Sons, 2008.

[31] S. Erdweg, T. van der Storm, M. Völter, L. Tratt, R. Bosman, W. R. Cook, A. Gerritsen f, Angelo Hulshout g, Steven Kelly h, Alex Loh c, Gabriël Konat i, Pedro J. Molina j, Martin Palatnik, R. Pohjonen, E. Schindler, K. Schindler, R. Solmi, V. Vergu, E. Visser, K. van der Vlist, G. Wachsmuth, J. van der Woning, "Evaluating and comparing language workbenches: existing results and benchmarks for the future" *Comput. Lang. Syst. Struct.*, 44(PA), 24-47, 2015. <https://doi.org/10.1016/j.cl.2015.08.007>

[32] H. Berg, B. Møller-Pedersen, "Type-Safe Symmetric Composition of Metamodels Using Templates" *Lect. Notes. Comput. Sc.*, 7744, 160-178, 2012. https://doi.org/10.1007/978-3-642-36757-1_10

[33] H. Berg, B. Møller-Pedersen, "Metamodel and Model Composition by Integration of Operational Semantics" *Comm. Com. Inf. Sc.*, 580, 172-189, 2015. https://doi.org/10.1007/978-3-319-27869-8_10

[34] Schmidt, M., Wenzel, S., Kehrer, T., Kelter, U., "History-based merging of models" in 2009 ICSE Workshop on Comparison and Versioning of Software Models (CVSM '09), Washington, USA, 2009. <https://doi.org/10.1109/CVSM.2009.5071716>

[35] H. K. Dam, A. Eged, M. Winikoff, A. Reder, R. E. Lopez-Herrejon, "Consistent merging of model versions" *J. Syst. Softw.*, 112(C), 137-155, 2016. <https://doi.org/10.1016/j.jss.2015.06.044>

[36] D. Zhang, S. Li, X. Liu, "An Approach for Model Composition and Verification" in 1th IEEE Computer Society International Joint Conference on INC, IMS and IDC, Seoul, South Korea, 2009. <https://doi.org/10.1109/NCM.2009.271>

[37] The Alloy Homepage. <http://alloy.mit.edu>, accessed 21 November 2018.

[38] D. Jackson, "a lightweight object modelling notation" *ACM Trans. Softw. Eng. Methodol.*, 11(2), 256-290, 2002. <https://doi.org/10.1145/505145.505149>

[39] The Eclipse Graphical Modeling Framework (GMF) Homepage, <http://www.eclipse.org/modeling/gmf/>, accessed 21 November 2018.

[40] The Eclipse Modeling Project (EMP) Homepage, <http://www.eclipse.org/modeling/>, accessed 21 November 2018.

[41] A. Lédeczi, A. Bakay, M. Maroti, P. Volgyesi, G. Nordstrom, J. Sprinkle, G. Karsai, "Composing domain-specific design environments", *Computer.*, 34(11), 44-51, 2001. <https://doi.org/10.1109/2.963443>

[42] A. Lédeczi, G. Nordstrom, G. Karsai, P. Volgyesi, M. Maroti, "On metamodel composition" in 2001 IEEE International Conference on Control Applications (CCA'01), Mexico City, Mexico, 2001. <https://doi.org/10.1109/CCA.2001.973959>

- [43] J. M. Jézéquel, "Model driven design and aspect weaving" *Softw. Syst. Model.*, 7(2), 209-218, 2008. <https://doi.org/10.1007/s10270-008-0080-5>
- [44] O. Barais, J. Klein, B. Baudry, A. Jackson, S. Clarke, "Composing multi-view aspect models" in 7th International Conference on Composition-Based Software Systems (ICCBSS 2008), Washington, USA, 2008. <https://doi.org/10.1109/ICCBSS.2008.12>
- [45] P. Sánchez, L. Fuentes, D. Stein, S. Hanenberg, R. Unland, "Aspect-oriented model weaving beyond model composition and model transformation" *Lect. Notes. Comput. Sc.*, 5301, 766-781, 2008. https://doi.org/10.1007/978-3-540-87875-9_53
- [46] A. Hovsepyan, S. Van Baelen, Y. Berbers, W. Joosen, "Specifying and Composing Concerns Expressed in Domain-Specific Modeling Languages", In M. Oriol, B. Meyer (Ed.), *Objects, Components, Models and Patterns*, *Lect. Notes. Bus. Inf.*, 33, 116-135, 2009. https://doi.org/10.1007/978-3-642-02571-6_8
- [47] M. P. Cardoso, T. Carvalho, J. G. F. Coutinho, W. Luk, R. Nobre, P. Diniz, Z. Petrov, "LARA: an aspect-oriented programming language for embedded systems" in 11th annual international conference on Aspect-oriented Software Development, New York, USA, 2012. <https://doi.org/10.1145/2162049.2162071>
- [48] P. Pinto, T. Carvalho, J. Bispo, M. A. Ramalho, J. M. P. Cardoso, "Aspect composition for multiple target languages using LARA" *Comput. Lang. Syst. Struct.*, 53, 1-26, 2018. <https://doi.org/10.1016/j.cl.2017.12.003>
- [49] J. Whittle, P. Jayaraman, A. Elkhodary, A. Moreira, J. Araújo, "MATA: A unified approach for composing UML aspect models based on graph transformation" *Lect. Notes. Comput. Sc.*, 5560, 191-237, 2009. https://doi.org/10.1007/978-3-642-03764-1_6
- [50] J. Whittle, P. Jayaraman, "MATA: A Tool for Aspect-Oriented Modeling Based on Graph Transformation" *Lect. Notes. Comput. Sc.*, 5002, 16-27, 2008. https://doi.org/10.1007/978-3-540-69073-3_3
- [51] M. Schöttle, O. Alam, F.P. Garcia, G. Mussbacher, J. Kienzle, "TouchRAM: a multitouch-enabled software design tool supporting concern-oriented reuse" in 13th International Conference on Modularity. New York, USA, 2014. <https://doi.org/10.1145/2584469.2584475>
- [52] M. Schöttle, N. Thimmegowda, O. Alam, J. Kienzle, G. Mussbacher, "Feature modelling and traceability for concern-driven software development with TouchCORE" in 14th International Conference on Modularity, New York, USA, 2015. <https://doi.org/10.1145/2735386.2735922>
- [53] M. Voelter, "Language and IDE modularization, extension and composition with MPS" *Lect. Notes. Comput. Sc.*, 7680, 383-430, 2013. https://doi.org/10.1007/978-3-642-35992-7_11
- [54] M. Voelter, J. Warmer, B. Kolb, "Projecting a modular future" *IEEE. Software.*, 32(5), 46-52, 2015. <https://doi.org/10.1109/MS.2014.103>
- [55] M. Voelter, B. Kolb, T. Szabó, D. Ratiu, A. van Deursen, "Lessons learned from developing mbeddr: a case study in language engineering with MPS" *Softw. Syst. Model.*, 17(66), 1-46, 2017. <https://doi.org/10.1007/s10270-016-0575-4>
- [56] A. M. Şutii, "MetaMod: a modeling formalism with modularity at its core" in 30th IEEE/ACM International Conference on Automated Software Engineering, Lincoln, USA, 2015. <https://doi.org/10.1109/ASE.2015.29>
- [57] A. M. Şutii, M. Van Den Brand, T. Verhoeff, "Exploration of modularity and reusability of domain-specific languages: an expression DSL in MetaMod" *Comput. Lang. Syst. Struct.*, 51(C), 48-70, 2018. <https://doi.org/10.1016/j.cl.2017.07.004>
- [58] D. H. Lorenz, B. Rosenan, "Cedalion: a language for language oriented programming" *SIGPLAN. Not.*, 46(10), 733-752, 2011. <https://doi.org/10.1145/2076021.2048123>
- [59] D. H. Lorenz, B. Rosenan, "Code reuse with language oriented programming" *Lect. Notes. Comput. Sc.*, 6727, 167-182, 2011. https://doi.org/10.1007/978-3-642-21347-2_13
- [60] D. H., Lorenz, B. Rosenan, "CEDALIONS Response to the 2016 Language Workbench Challenge", in LWC@SLE 2016 Language Workbench Challenge at the 2016 ACM SIGPLAN International Conference on Object-Oriented Programming, Systems, Languages, and Applications, Amsterdam, Netherlands, 2016.
- [61] L. C. L. Kats, E. Visser, "The Spoofox language workbench: rules for declarative specification of languages and IDEs" *SIGPLAN. Not.*, 45(10), 444-463, 2010. <https://doi.org/10.1145/1932682.1869497>
- [62] M. Voelter, S. Benz, C. Dietrich, B. Engelmann, M. Helander, L. C. L. Kats, E. Visser, G. Wachsmuth, *DSL engineering: Designing, implementing and using domain-specific languages*, dslbook.org, 2013.
- [63] The Xtext Homepage, <http://www.eclipse.org/Xtext/>, accessed 21 November 2018.
- [64] L. Bettini, *Implementing domain-specific languages with Xtext and Xtend*, Packt Publishing Ltd, 2016.
- [65] H. Krahn, B. Rumpe, S. Völkel, "Monticore: Modular development of textual domain specific languages" *Lect. Notes. Bus. Inf.*, 11, 297-315, 2008. https://doi.org/10.1007/978-3-540-69824-1_17
- [66] H. Krahn, B. Rumpe, S. Völkel, "MontiCore: a framework for compositional development of domain specific languages" *Int. J. Softw. Tools Technol. Transf.*, 12(5), 353-372, 2010. <https://doi.org/10.1007/s10009-010-0142-1>
- [67] L. Pedro, V. Amaral, D. Buchs, "Foundations for a domain specific modeling language prototyping environment: A compositional approach" in 8th OOPSLA workshop on domain-specific modeling. In Companion to the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications (OOPSLA Companion '08). ACM, New York, USA, 2008. <https://doi.org/10.1145/1449814.1449886>
- [68] L. Pedro, M. Risoldi, D. Buchs, B. Barroca, V. Amaral, "Composing Visual Syntax for Domain Specific Languages" *Lect. Notes. Comput. Sc.*, 5611, 889-898, 2009. https://doi.org/10.1007/978-3-642-02577-8_97
- [69] B. Meyers, "A Multi-Paradigm Modelling Approach for the Engineering of Modelling Languages" *Ceur. Workshop. Proce.*, 1321, 2-9, 2015.
- [70] B. Meyers, A. Cicchetti, E. Guerra, J. de Lara, "Composing textual modelling languages in practice" in 6th International Workshop on Multi-Paradigm Modeling, New York, USA, 2012. <https://doi.org/10.1145/2508443.2508449>
- [71] J. De Lara, E. Guerra, "Deep metamodelling with metaDepth" *Lect. Notes. Comput. Sc.*, 6141, 1-20, 2009. https://doi.org/10.1007/978-3-642-13953-6_1
- [72] C. Herrmann, H. Krahn, B. Rumpe, M. Schindler, S. Völkel, "An Algebraic View on the Semantics of Model Composition" *Lect. Notes. Comput. Sc.*, 4530, 99-113, 2007. https://doi.org/10.1007/978-3-540-72901-3_8
- [73] S. Kelly, J. P. Tolvanen, *Domain-Specific Modeling: Enabling Full Code Generation*, Wiley, 2008.
- [74] S. Völkel, "Kompositionale entwicklung domänenspezifischer sprachen," Ph.D Thesis, Technical University Carolo-Wilhelmina, 2011.
- [75] The Xml Metadata Interchange (XMI) Specification page, <http://www.omg.org/mof/> <http://www.omg.org/spec/XMI/>, accessed 21 November 2018.
- [76] The MetaObject Facility (MOF) Specification page, <http://www.omg.org/mof/>, accessed 21 November 2018.
- [77] The Eclipse Standard Widget Toolkit (SWT) Homepage, <https://www.eclipse.org/swt/>, accessed 21 November 2018.
- [78] T. Grossman, G. W. Fitzmaurice, R. Attar, "A survey of software learnability: metrics, methodologies and guidelines" in 27th International Conference on Human Factors in Computing Systems, New York, USA, 2009. <https://doi.org/10.1145/1518701.1518803>

Modeling an Energy Consumption System with Partial-Value Data Associations

Nong Ye*, Ting Yan Fok, Oswald Chong

School of Computing, Informatics and Decision Systems Engineering, Ira A. Fulton School of Engineering, Arizona State University, 85287-8809, USA

ARTICLE INFO

Article history:

Received: 13 September, 2018

Accepted: 20 November, 2018

Online: 05 December, 2018

Keywords:

Partial-Value Association,

Data Mining,

Energy Consumption,

Structural System Model

ABSTRACT

Many existing system modeling techniques based on statistical modeling, data mining and machine learning have a shortcoming of building variable relations for the full ranges of variable values using one model, although certain variable relations may hold for only some but not all variable values. This shortcoming is overcome by the Partial-Value Association Discovery (PVAD) algorithm that is a new multivariate analysis algorithm to learn both full-value and partial-value relations of system variables from system data. Our research used the PVAD algorithm to model variable relations of energy consumption from data by learning full-and partial-value variable relations of energy consumption. The PVAD algorithm was applied to data of energy consumption obtained from a building at Arizona State University (ASU). Full- and partial-value variable associations of building energy consumption from the PVAD algorithm are compared with variable relations from a decision tree algorithm applied to the same data to show advantages of the PVAD algorithm in modeling the energy consumption system.

1. Introduction

Our research is an extension of work originally presented in the 2018 IEEE ICCAR Conference [1]. Many complex systems, such as energy consumption systems and transportation systems, involve both engineered and non-engineered system factors. For example, the energy consumption system of a building involves both engineered system factors, (e.g., AC equipment, pump for water use, lighting system, computers, and network equipment) and non-engineered system factors, (e.g., social/behavioral factors, such as occupants' activities, and environmental/natural factors such as outside climate), which are intertwined to drive the energy consumption and demand of the building [2-4]. For another example, the transportation system involves both engineered system factors, (e.g., the transportation infrastructure including highways, streets and roads, and traffic control mechanisms, such as traffic lights) and non-engineered system factors, (e.g., social/behavioral factors such as traffic flows, drivers, pedestrians, and car accidents, as well as natural/environmental factors, such as weather conditions).

Although models of engineered systems may be available, models of mixed-factor systems are usually not available due to unknown interconnectivities and interdependencies of many

engineered and non-engineered system factors. A complete, accurate system model, which clearly defines relations of system variables including interconnectivities and interdependencies of engineering and non-engineered system factors, is highly desirable for many applications. For example, variable relations of energy consumption are required to enable the accurate estimation of energy consumption/demand and the close alignment of energy production with energy demand to achieve energy production and use.

Utility/energy companies currently rely heavily on the past data of electricity loads in base, average and peak to project energy production/supply. This statistical investigative activity is done without adequate and accurate models of energy consumption systems [3]. Power plants often generate enough power to satisfy base loads and meet the difference between peak and base loads, sudden demand surge or any gap of energy supply and demand through their excess production capacities or by procuring from other energy sources [5, 6]. Historical data lack critical real-time features (e.g., the lag effect of historical data, and lack of finer levels and finer divisions in time and space) for the accurate projection and estimation of energy demand and consumption. Without adequate and accurate models of energy consumption systems, it is extremely difficult to obtain an accurate projection and estimation of energy demand and consumption. As a result,

* Nong Ye, Arizona State University, Email: nongye@asu.edu

energy has to be produced in excess in order to meet potential rise in demand. Energy production in excess is a significant cause of waste and inefficiency. Even with current technologies to obtain dynamic data of energy consumption systems in real time, the lack of adequate and accurate energy system models renders real-time dynamic system data useless for closely aligning energy production with energy demand to achieve energy production efficiency and energy use reduction. The ultimate energy efficiency through smart energy production and use will enable a shift from the existing code-, standard- and experience-based forecasting approach to a more dynamic, real-time and smart technology environment based on real-time data, models and analytics for the real-time, accurate estimation of energy consumption and smart technologies to align energy production with energy demand closely for energy use reduction and energy production efficiency.

Many statistical modeling, data mining and machine learning techniques for system modeling, including decision trees, regression analysis, artificial neural network, and Bayesian networks, have been used to analyze and model energy consumption and efficiency of equipment, homes and buildings [7-16]. System modeling techniques based on many existing statistical analysis, machine learning and data mining have a shortcoming of building variable relations for the full ranges of variable values using one model, although certain variable relations may hold for only some but not all variable values. This shortcoming is overcome by the PVAD algorithm that is a new multivariate analysis algorithm to learn both full-value and partial-value relations of system variables from system data. Our research used the PVAD algorithm to model variable relations of energy consumption from data by learning full-and partial-value variable relations of energy consumption. The PVAD algorithm was applied to building energy consumption data at ASU.

2. Shortcomings of existing techniques of system modeling from data

Existing methods of learning system models from data include statistical analysis [17-24] and data mining techniques [23-32]. With system modeling from data, classification and prediction can be performed to explain or find relations among system variables. Depending on the nature of data, there are several methods to analyze data using statistical techniques such as parametric, non-parametric and logistic regression. For example, when modeling categorical dependent variables, logistic regression can be applied [17, 21, 22]. In addition to decision and regression trees [23, 24], random forest and support vector machine are also considered [25-28, 29, 31]. However, the above methods assume that the role of a variable in a variable relation is known (i.e., which variable is an independent or dependent variable) and a variable plays only one role of being either an independent variable or a dependent variable in one layer of variable relations. Once a variable is considered as an independent variable, it can no longer be utilized as a dependent variable which is a main disadvantage especially when the role of a variable is not known or when multiple layers of variable relations are required where a variable can play different roles of being an independent or dependent variable in different variable relations at different layers.

Bayesian networks [23, 24, 35-37], structural equation models [33, 34] and reverse engineering methods [38-47] are examples of a few options left that can provide system modeling without prior knowledge of variables. However, those techniques discover only variable relations for full ranges of all variable values instead of relations for specific values only. This can be seen from the Fisher’s Iris data set [48] in which the classification of the target variable (Plant Type) using independent variables works for only the values of Iris Versicolor and Iris Virginica) for the target variable but not for another target value of Iris Sentosa. For such data where variable relations hold for partial ranges of variable values only or different variable relations hold for different ranges of variable values, the model of the same variable relations for all variable values do not fit all data values well, that is, the model explains or represents the whole data set poorly.

The PVAD algorithm was developed as a new system modeling technique [49-51] to overcome the above shortcomings. Variable value associations can be used to construct associative networks as multi-layer structural system models. The application of the PVAD based system modeling technique is part and parcel of our research of energy consumption in systems.

3. The energy consumption data and the PVAD application

The PVAD algorithm is presented in detail in [49-51]. This section shows the PVAD application to data of energy consumption collected from an ASU building in January 2013 for modeling energy consumption. There was a data sample every 15 minutes. The data set has 2976 data records or instances. Each data record contains four numeric values for the consumption of electricity (E), cooling (C), heating (H), and air temperature (A), respectively, as well as TimeStamp (T). T is important because changes of T are associated with changes in presence and activities of occupants and changes of E, C and H.

To apply the PVAD algorithm, in Step 1 the numeric variables of A, H, C, and E, were transformed into categorical variables as shown in Fi To apply the PVAD algorithm, in Step 1 the numeric variables of A, H, C, and E, were transformed into categorical variables as shown in Figures (1)-(4). More details of Step 1 are in [1].

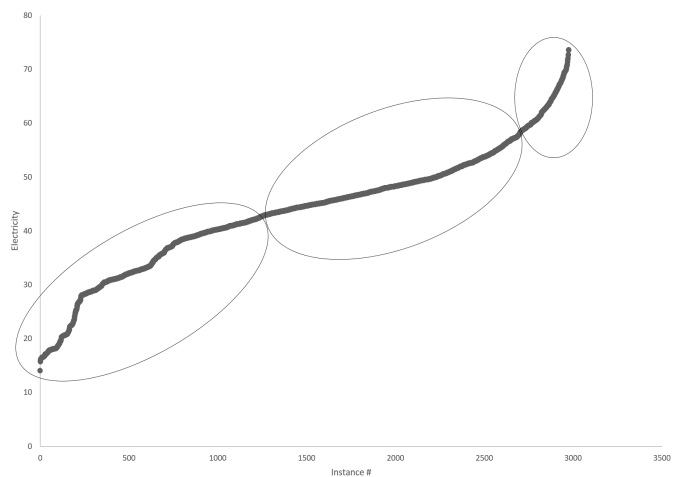


Figure 1. An example of plotting E values to determine data clusters and categorical values.

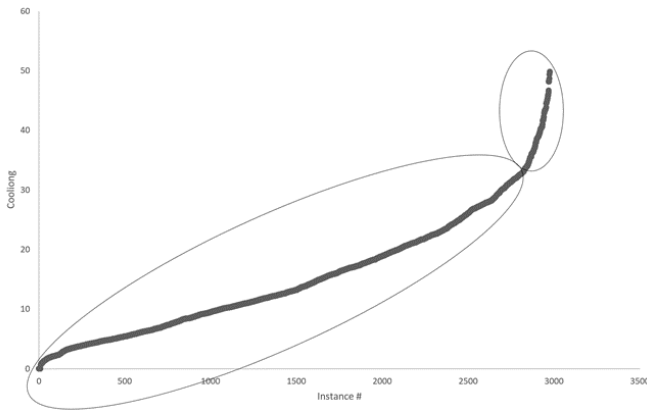


Figure 2. An example of plotting C values to determine data clusters and categorical values.

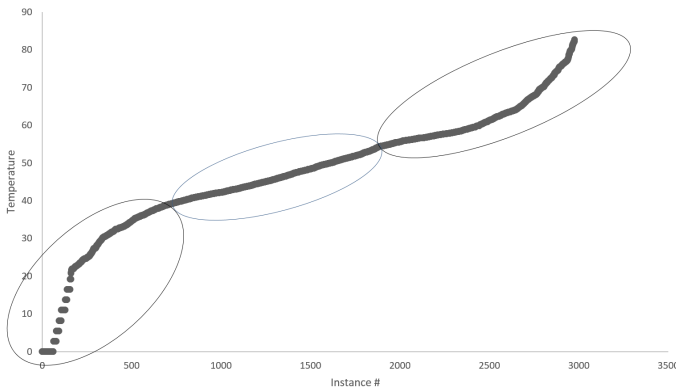
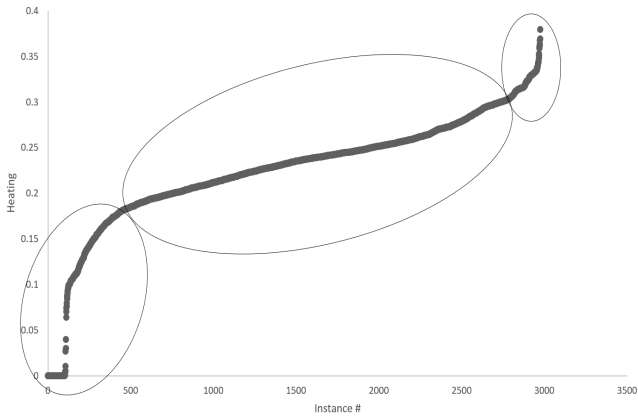


Figure 4. An example of plotting A values to determine data clusters and categorical values.

Step 2.1 generated candidate 1-to-1 associations of partial variable values, $x = a \rightarrow y = b$, where $x = a$ is the conditional variable value (CV) and $y = b$ is the associative variable value (AV), and computed the co-occurrence ratio (cr) of each candidate association as follows:

$$cr(x = a \rightarrow y = b) = \frac{N_{x=a,y=b}}{N_{x=a}} \quad (1)$$

If cr is greater than or equal to the parameter α , we had an established association. For example, Table (1) shows 1-to-1 associations having CV: C = High together with their respective cr values and $\alpha = 0.8$.

Table 1: 1-to-1 associations of ASU energy consumption data with CV: C=High

#	Association	cr	Co-Occurrence Frequency ($N_{x=a,y=b}$)	Type of Association
1	C=High \rightarrow T=12:15 PM to 5:30 PM	0.48	62	Candidate
2	C=High \rightarrow T=5:45 PM to 11 PM	0.43	55	Candidate
3	C=High \rightarrow T=8:15 AM to 12 PM	0.09	11	Candidate
4	C=High \rightarrow E=High	0.20	25	Candidate
5	C=High \rightarrow E=Medium	0.80	103	Established
6	C=High \rightarrow H=Low	0.73	94	Established
7	C=High \rightarrow H=Medium	0.27	34	Candidate
8	C=High \rightarrow A=High	0.96	123	Established
9	C=High \rightarrow A=Medium	0.04	5	Candidate

In addition to parameter α , two other parameters, β and γ , are also needed. β is used to remove associations whose number of supporting instances (the instances containing variable values in the numerator of equation 1) is smaller than β . γ is used to remove an association with a common CV or AV that appears in more than γ of the data set. In this example, β is set to be 50 while α and γ are set to 0.8 and 0.95, respectively.

Step 2.2 uses two methods, YFM1 and YFM2, to examine and establish p -to- q associations, $X = A \rightarrow Y = B$, where X and Y represent multiple variables. For example, using #5, 6 and 8 in Table (1), we applied YFM1 which considers all combinations of AVs covered in those associations so as to find 1-to- q associations, where $q > 1$. To find 1-to-2 established associations, we first computed $N_{CV} = 103 \div 0.8046875 = 128$. Then we considered all possible combinations of two-variable AVs from the established 1-to-1 associations:

1. C=High- \rightarrow E=Medium, H=Low (from #5 and #6)
2. C=High- \rightarrow E=Medium, A=High (from #5 and #8)
3. C=High- \rightarrow H=Low, A=High (from #6 and #8).

For each 1-to-2 candidate associations above, $N_{CommonSubset}$, the number of instances in the common subset of supporting instance, was computed to calculate cr for the 1-to-2 association. The results are given in Table (2). In this case, C=High- \rightarrow H=Low, A=High is the only established association.

Table 2: Calculation for 1-to-2 associations with CV: C=High

#	Association	$N_{CommonSubset}$	cr ($= N_{CommonSubset} / 128$)
1	C=High \rightarrow E=Medium, H=Low	87	0.6796875
2	C=High \rightarrow E=Medium, A=High	103	0.125
3	C=High \rightarrow H=Low, A=High	94	0.8046875

YFM2 is used to find 2-to-1 associations. YFM2 considers all candidate associations (cr value in (0, 1]) not just established associations ($cr \geq \alpha$). Table (3) is used to illustrate YFM2 in the following.

- 1) Determinem_i = $\lceil n_i \times \alpha \rceil$. If we pick C=High- \rightarrow T=12:15 PM to 5:30 PM to start with, $n_1 = 62$. Then $m_1 = \lceil n_1 \times \alpha \rceil = 50$ and the 2-to-1 association that we would like to generate will have C=High, T=12:15 PM to 5:30 PM as CV. Note that if $m_i < \beta$, the whole group is dropped as the number of instances covered by the new CV is just the occurrence frequency which should be $\geq \beta$.
- 2) Iterate through all other associations from Table (2). Skip

immediately to the next line if the AV of that association is the same as one picked in the previous step. For example, #2 has AV: T=5:45 PM to 11 PM that represents Timestamp. While the AV of the association also represents timestamp (T=12:15 PM to 5:30 PM), we skip to #3 without looking at the intersection of the instances.

2ii) Generate 2-to-1 association if $n_{\text{intersection}} \geq 50$. Table (3) lists the $n_{\text{intersection}}$ and the corresponding cr value.

Table 3: Calculation for 2-to1 associations of ASU energy consumption data with CV "C=High, T=12:15 PM to 5:30 PM"

Combination (Type of Association)	Association	$n_{\text{intersection}}$	$cr (= n_{\text{intersection}} / 62)$
1 & 4 (Candidate)	C=High, T=12:15 PM to 5:30 PM → E=High	3	0.04
1 & 5 (Established)	C=High, T=12:15 PM to 5:30 PM → E=Medium	59	0.9516
1 & 6 (Candidate)	C=High, T=12:15 PM to 5:30 PM → H=Low	59	0.9516
1 & 7 (Candidate)	C=High, T=12:15 PM to 5:30 PM → H=Medium	3	0.04
1 & 8 (Established)	C=High, T=12:15 PM to 5:30 PM → A=High	62	1
1 & 9 (Candidate)	C=High, T=12:15 PM to 5:30 PM → A=Medium	0	0

Following the same procedure, other p -to- q associations were generated by YFM1 and YFM2. Step 3 generalized and consolidated variable associations of partial values into associations of full value ranges if there are partial-value associations covering the full value range of the same variable.

The PVAD algorithm was used to analyze the energy consumption data using various values of $\alpha = 1, 0.9, \text{ and } 0.8$, $\beta = 50, 30, \text{ and } 10$, and $\gamma = 95\%$. The results for $\gamma = 95\%$, $\beta = 50$, and $\alpha = 0.8$ are most meaningful and presented in the next section.

4. Results of the PVAD Algorithm

Tables (4)-(5) list the most specific association(s) in each group of the associations with the same AV. Table (6) lists the most generic association(s) in each group of the associations with the same AV. Variable relations for energy consumption revealed by each association in Tables (4)-(6). In Tables (4)-(6), there are groups that give similar associations. For example, the associations in Group 1 and Group 2 in Table (6) are similar. For the groups with similar associations, we marked only one group using the symbol ^ in the column of group #. Most of the associations in Tables (4)-(6) involve C=Low for cooling being low in CV or AV, because most of instances in the data set (2848 out of totally 2976 instances) contain C=Low due to the month of January when the data was collected. Since C=Low is so common in the data set, C=Low can be dropped from the associations when interpreting associations.

Table 4: Specific associations in each group of associations with the same AV: Set 1

Group #	The most specific association(s) in group
1	A=Medium, [T=12:15 PM to 11 PM, E=High]/ [T=6:15 AM to 11 PM, E=Medium]/[T=11:15 PM to 6 AM, E=Low] → H=Medium, C=Low
2^	A=Medium, C=Low, [T=11:15 PM to 6 AM, E=Low]/[T=6:15 AM to 12 PM, E=Medium]/[T=12:15 PM to 11 PM, E=High/Medium] → H=Medium A=High, C=Low, E=Medium, T=8:15 AM to 12 PM → H=Medium
3^	A=High, E=Medium, C=High, T=12:15 PM to 5:30 PM → H=Low
4	C=High, E=Medium, T=12:15 PM to 5:30 PM → A=High H=Low
5^	H=High, E=Medium, T=6:15 AM to 8 AM → A=Low, C=Low
6	H=Medium, E=Low, T=12:15 PM to 5:30 PM → A=High, C=Low
7^	[E=Medium, C=*, H=Low]/[E=Low, C=Low, H=Medium], T=12:15 PM to 5:30 PM → A=High C=High, T=5:45 PM - 11 PM → A=High
8	H=Medium, E=High, T=12:15 PM to 5:30 PM → A=Medium, C=Low
9^	H=Medium, C=Low, E=High, T=12:15 PM to 5:30 PM, → A=Medium
10	H=High, C=Low, E=Medium, T=6:15 AM to 8 AM, → A=Low
11^	[A=Low, H=High]/[A=High, H=Low]/[A=Low/Medium, H=Medium], C=Low, T=11:15 PM to 6 AM → E=Low
12	[A=Low, H=High]/[A=High, H=Low]/[A=Medium/Low, H=Medium], T=11:15 PM to 6 AM → C=Low, E=Low

Table 5: Specific associations in each group of associations with the same AV: Set 2

13	A=Medium, H=High, T=5:45 PM to 11 PM → C=Low, E=Medium T=8:15 AM to 12 PM, H=Medium, A=High/Medium → E=Medium, C=Low
14^	A=High, H=Low, C=High, T=12:15 PM to 5:30 PM → E=Medium A=Medium, H=High, C=Low, T=5:45 PM to 11 PM → E=Medium A=Medium/High, H=Medium, C=Low, T=8:15 AM to 12 PM → E=Medium
15	H=Low, C=High, T=12:15 PM to 5:30 PM → A=High E=Medium
16^	A=High, C=High, T=12:15 PM to 5:30 PM, → H=Low E=Medium A=Medium/High, C=Low, T=8:15 AM to 12 PM → H=Medium E=Medium

The associative network of the energy consumption system model shown in Figure (5) was constructed using the associations in the groups marked with ^ in Table (6). Figure (5) shows the factors associated with the high, medium and low air temperatures (from the associations with A as the AV), the factors associated with the Medium and Low heating consumption (from the associations with H as the AV), and the factors associated with the medium and low electricity consumption (from the associations with E as the AV).

Figure (5) shows that E, C, H and A are related differently in different time periods. For example, in the afternoon, T = 12:15 PM to 5:30 PM, the medium heating consumption (H = Medium) along with the high electricity consumption (E = High) is associated with the medium air temperature (A = Medium), whereas in the early morning, T = 6:15 AM to 8 AM, the high heating consumption (H = High) is associated with the low air temperature (A = Low). Similarly, the most specific associations

in Tables (4)-(5), even the most generic associations in Table (6) and in Figure (5) show that associations of T, E, C, H and A differ in different value ranges of these variables. This illustrates that the PVAD algorithm can discover full/partial-value variable relations that exist in many real-world systems.

Table 6: Generic associations in each group of associations with the same AV.

Group #	The most generic association(s) in each group
1^	A=Medium/E=High → H=Medium
2	E=High/A=Medium, C=Low → H=Medium
3	E=Medium, C=High, A=High → H=Low
4	E=Medium, C=High → H=Low A=High
5^	H=High, T=6:15 AM to 8 AM → A=Low.
6^	E=Low, T=12:15 PM to 5:30 PM → A=High
7^	H=Low → A=High T=5:45 PM - 11 PM, C=High → A=High C=Low, E=Low, T=12:15 PM to 5:30 PM → A=High
8^	H=Medium, E=High, T=12:15 PM to 5:30 PM → A=Medium
9	H=Medium, C=Low, E=High, T=12:15 PM to 5:30 PM → A=Medium
10	H=High, C=Low, T=6:15 AM to 8 AM → A=Low
11	C=Low, T=11:15 PM to 6 AM → E=Low
12^	T=11:15 PM to 6 AM → E=Low
13^	T=8:15 AM to 12 PM → E=Medium H=High, T=5:45 PM to 11 PM → E=Medium
14	H=High, C=Low, T=5:45 PM to 11 PM → E=Medium C=High → E=Medium
15	C=High → A=High E=Medium
16^	A=High, C=High, T=12:15 PM to 5:30 PM → H=Low E=Medium A=Medium/High, C=Low, T=8:15 AM to 12 PM → H=Medium E=Medium

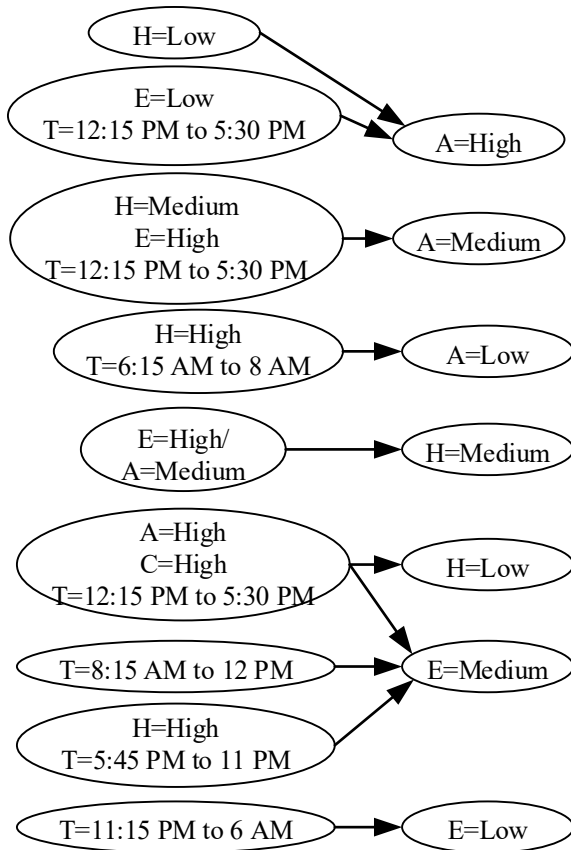


Figure 5. The most generic associations in the groups marked by ^ in Table (6) represented in an associative network.

5. Comparison of the PVAD algorithm with some data mining techniques

We considered two of the existing data mining techniques to compare with the PVAD algorithm: association rule and decision tree.

5.1. Comparison with the association rule technique

The association rule technique first uses the Apriori algorithm to determine frequent item sets that satisfy the minimum support [23-24]. Then each frequent item set is broken up into all possible combinations of association rules which are evaluated to see if any of them satisfy the minimum support and confidence. For a large dataset, frequent item sets and candidate association rules from frequent item sets can be enormous, requiring huge amounts of computer memory space and computation time. When the association rule technique was applied to the energy consumption data, there were too many frequent item sets and consequently association rules to be listed in this paper. While the performance of the association rule technique was hindered by the data size, the search space of associations in the PVAD algorithm is narrowed down by YMF1 and YFM2, along with parameters α , β and γ .

5.2 Comparison with the decision tree technique

Decision tree is a data mining technique to learn decision rules that express relations of the dependent variable y with independent variables x in a directed and acyclic graph [23-24]. The software, Weka, was used to construct decision trees of the energy consumption system data, To construct a decision tree in Weka, there are different algorithms such as ID3 [52] and J48 [53]. The later one is an extended version of ID3 with additional features like dealing with missing values and continuous attribute value ranges. It also addresses the over-fitting problem that decision trees are prone to by pruning. The pruning process requires the computation of the expected error rate. If the error rate of a subtree is greater than that of a leaf node, a subtree is pruned and replaced by the leaf node.

In our research, ID3 was used for the comparison with the PVAD algorithm because ID3 produces comparable results with associations produced by the PVAD algorithm. Leaf nodes produced by ID3 are pure in that the class labels of instances are the same in each leaf node. The purity of leaf node corresponds to AV in associations from the PVAD algorithm having the same variable value. The PVAD algorithm produces all associations up to N -to-1 associations, where $N+1$ is the number of variables. In other words, the PVAD algorithm can generate the longest CVs and find the AV that they are associated with. The combination of CVs corresponds to the path from the root of a decision tree down to a leaf node.

Because the decision tree technique requires the identification of one dependent variable (the target variable) and independent variables (attribute variables) for each decision tree, five decision trees need to be constructed for each of the five variables as the dependent variable. Tables (7)-(10) list decision rules produced by one of the five ID3 trees.

Although the decision rules from the decision trees appear to have the same form as associations from the PVAD algorithm, a

decision rule has a different meaning from an association from the PVAD algorithm. A decision rule derived from the root of a decision tree to a leaf node of the decision tree represents a frequent item set with instances in the leaf node having the values of the target variable and the attribute variables in the decision rule. This is why we see a path in a decision tree is also present in another tree even though different decision trees have different target variables. For example, the variable values in E=Medium, A=High, H=Medium, C=Low, T=12:15 PM to 5:30 PM, are found in all four decision trees. Note that the energy consumption data set has only five variables. Redundant paths of different decision trees can be found more often for larger data sets with more variables. This means the waste of computation time and space and the difficulty of sorting out results from a number of

Table 7: Decision rules from the ID3 Tree with Air Temperature as the target variable same as PVAD association rules

#	Decision rules that appear same as PVAD association rules
1	H=Medium, E=High, T=12:15 PM to 5:30 PM → A=Medium
2	H=Medium, E=Low, T=12:15 PM to 5:30 PM → A=High
3	H=Low, T=12:15 PM to 5:30 PM → A=High
4	H=High, E=Medium, T=6:15 AM to 8 AM → A=Low

Table 8: Decision rules from the ID3 Tree with Air Temperature = Low as the target variable

5	H=Medium, E=High, T=11:15 PM to 6 AM → A=Low
6	H=High, E=Low, T=11:15 PM to 6 AM → A=Low
7	H=High, E=Medium, T=11:15 PM to 6 AM → A=Low
8	H=High, E=Medium, T=12:15 PM to 5:30 PM → A=Low
9	H=High, E=Low, T=6:15 AM to 8 AM → A=Low
10	H=Medium, E=Low, T=6:15 AM to 8 AM → A=Low
11	H=High, E=Low, T=8:15 AM to 12 PM → A=Low
12	H=High, E=Medium, T=8:15 AM to 12 PM → A=Low
13	H=Low, C=Low, E=Medium, T=5:45 PM to 11 PM → A=Low

Table 9: Decision rules from the ID3 Tree with Air Temperature = Medium as the target variable

14	H=Low, T=6:15 AM to 8 AM → A=Medium
15	H=Medium, E=Low, T=11:15 PM to 6 AM → A=Medium
16	H=Low, E=Medium, T=11:15 PM to 6 AM → A=Medium
17	H=Medium, E=Medium, T=11:15 PM to 6 AM → A=Medium
18	H=High, E=Low, T=12:15 PM to 5:30 PM → A=Medium
19	H=High, E=Low, T=5:45 PM to 11 PM → A=Medium
20	H=High, E=Medium, T=5:45 PM to 11 PM → A=Medium
21	H=Medium, E=Medium, T=6:15 AM to 8 AM → A=Medium
22	H=Medium, E=High, T=8:15 AM to 12 PM → =Medium
23	H=Medium, E=Low, T=8:15 AM to 12 PM → A=Medium
24	H=Medium, C=Low, E=High, T=5:45 PM to 11 PM → A=Medium
25	H=Medium, C=Low, E=Medium, T=5:45 PM to 11 PM → A=Medium
26	H=Medium, C=Low, E=Medium, T=8:15 AM to 12 PM → A=Medium

decision trees. Hence, a decision rule corresponds to a frequent item set in the association rule technique, whereas an association from the PVAD algorithm corresponds to an association rule in the association rule technique. This is why there are decision rules in Tables (7) – (10) that are not found in associations of the PVAD algorithm because frequent item sets for those decision rules were eliminated in the process of forming associations. Hence, the PVAD algorithm has the advantage to the decision tree technique because the PVAD algorithm discovers associations rather than frequent item sets.

Table 10: Decision rules from the ID3 Tree with Air Temperature = High as the target variable

27	H=Low, E=Low, T=11:15 PM to 6 AM → A=High
28	H=Low, E=High, T=5:45 PM to 11 PM → A=High
29	H=Low, E=Low, T=5:45 PM to 11 PM → A=High
30	H=Low, E=Low, T=8:15 AM to 12 PM → A=High
31	H=Medium, C=High, E=High, T=5:45 PM to 11 PM → A=High
32	H=Medium, C=Low, E=Low, T=5:45 PM to 11 PM → A=High
33	H=Low, C=High, E=Medium, T=5:45 PM to 11 PM → A=High
34	H=Medium, C=High, E=Medium, T=5:45 PM to 11 PM → A=High
35	C=High, H=Low, E=Medium, T=8:15 AM to 12 PM → A=High
36	H=Medium, C=High, E=Medium, T=8:15 AM to 12 PM → A=High
37	H=Low, C=Low, E=Medium, T=8:15 AM to 12 PM → A=High
38	H=Medium, C=High, E=Medium, T=12:15 PM to 5:30 PM → A=High
39	H=Medium, C=Low, E=Medium, T=12:15 PM to 5:30 PM → A=High

There is another difference between the decision tree technique and the PVAD algorithm. Each step of constructing a decision tree performs the splitting of a data subset for data homogeneity based on the comparison of splits using only one variable and its values rather than combinations of multiple variables due to the large number of combinations and the enormous computation costs. Hence, the resulting decision tree contains decision rules with the consideration of only one variable at a time and may miss decision rules that can be generated if multiple variables and their values are considered and compared at a time. However, the PVAD algorithm examines one to multiple variables at a time and does not miss any associations that exist. The PVAD algorithm thus has the advantage to the decision tree technique by not missing any established associations and using YFM1 and YFM2 to cut down the computation costs.

Moreover, the decision tree algorithm requires the identification of the dependent variable (the target variable) and the independent variables (the attribute variables) although there may no priori knowledge for the identification of which variable is a dependent or independent variable. This is why five decision trees, with one decision tree taking each of the five variables as the target variable, had to be constructed for the energy consumption data. The PVAD algorithm does not require the distinction of dependent and independent variables but discovers variable value relations and the role of each variable in each variable value relation.

Furthermore, the PVAD algorithm can generate p -to- q associations with $q > 1$ that the decision tree technique cannot generate because a decision tree is constructed for only one target variable and produces only p -to-1 decision rules. Given the differences of the PVAD algorithm and the decision tree technique, the results of the PVAD algorithm are not comparable to the results of the decision tree technique. As discussed in Section 2, the PVAD algorithm overcomes shortcomings of existing statistical analysis and data mining techniques and produce partial/full-value associations that cannot be produced from other existing techniques.

5. Conclusion

Our research used the PVAD algorithm to learn and build the system model of energy consumption from data, especially learn relations of variables for both full and partial value ranges. The resulting partial-value associations of variables in the energy consumption system model reveal variable relations for partial value ranges that require not one but different models of variable relations over full value ranges of the variables. This finding shows that the PVAD algorithm has the advantage and capability of discovering variable relations for building a multi-layer, structural system model. Hence, the PVAD based system modeling technique can be useful in many fields to learn system models from data. The advantages of the PVAD algorithm to existing data mining, machine learning and statistical analysis techniques were also demonstrated by comparing the PVAD algorithm and its results from the application to the energy consumption data with the association rule technique and the decision tree technique.

References

- [1] N. Ye, T. Y. Fok, X. Wang, J. Collofello, N. Dickson, "Learning partial-value variable relations for system modeling", In Proceedings of the 2018 IEEE ICCAR Conference, Auckland, New Zealand, April 20-23, 2018, <https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=8384628>.
- [2] Kwok, K. green buildings: A comprehensive study of. Y. G., Statz, C., Wade, B., and Chong, W. K., "Carbon emissions modeling for calculation methods", ASCE Proceeding of the ICSDEC (ICSDEC), 118-126, 2012, <https://doi.org/10.1061/9780784412688.014>.
- [3] Swan, L. G., V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques" *Renewable and Sustainable Energy Reviews (Sust. Energy. Rev.)*, 13(8), 1819-1835, 2009.
- [4] Zhao, H.-X., Magoulès, F., "A review on the prediction of building energy consumption" *Renewable and Sustainable Energy Reviews (Renew. Sust. Energy. Rev.)*, 16(6), 3586-3592, 2012.
- [5] PSC Wisconsin, Electricity Use and Production Patterns. Public Service Commission of Wisconsin. Madison, WI: Public Service Commission of Wisconsin. Retrieved from <https://psc.wi.gov/thelibrary/publications/electric/electric04.pdf>, 2011.
- [6] Flex Alert., Peak Demand. Retrieved from Flex Alert Home: <http://www.flexalert.org/energy-ca/peak>, 2013.
- [7] Ahmed, A., et al. Ahmed, A., et al., "Mining building performance data for energy-efficient operation" *Advanced Engineering Informatics (Adv. Eng. Inform.)*, 25(2), 341-354, 2011.
- [8] Berges, M. E., et al. "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring", *Journal of Industrial Ecology (J. Ind. Ecol.)*, 14(5), 844-858, 2010.
- [9] Fan, C., et al., "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques" *Applied Energy (Appl. Energy.)*, 127(0), 1-10, 2014.
- [10] Hawarah, L., et al., User Behavior Prediction in Energy Consumption in Housing Using Bayesian Networks, *Artificial Intelligence and Soft Computing*, Springer Berlin Heidelberg, 6113: 372-379, 2010.
- [11] Khan, I., et al., "Fault Detection Analysis of Building Energy Consumption Using Data Mining Techniques" *Energy Procedia (Enrgy. Proced.)*, 42, 557-566, 2013.
- [12] Khansari, N., et al., "Conceptual Modeling of the Impact of Smart Cities on Household Energy Consumption" *Procedia Computer Science (Procedia. Comput. Sci.)*, 28, 81-86, 2014.
- [13] Kim, H., et al., "Analysis of an energy efficient building design through data mining approach", *Automation in Construction (Automat. Constr.)*, 20(1), 37-43, 2011.
- [14] Palizban, O., et al., "Microgrids in active network management – part II: System operation, power quality and protection", *Renewable and Sustainable Energy Reviews (Renew. Sust. Energy. Rev.)*, 2014.
- [15] Tso, G. K. F. and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks", *Energy*, 32(9), 1761-1768, 2007.
- [16] Vollaro, R. D. L., et al., "An Integrated Approach for an Historical Buildings Energy Analysis in a Smart Cities Perspective" *Energy Procedia (Enrgy. Proced.)*, 45(0), 372-378, 2014.
- [17] Friedman, J. H., "Multivariate Adaptive Regression Splines", *The Annals of Statistics (Ann. Stat.)*, 1-67, 1991.
- [18] Hastie, T., Tibshirani, R., and Friedman, J. H. *The Elements of Statistical Learning*, 2nd edition. New York, New York: Springer, 2009.
- [19] Zhang, H., and Singer, B. H., *Recursive Partitioning and Applications*, 2nd edition. New York, New York: Springer, 2010.
- [20] Freedman, D. A., *Statistical Models: Theory and Practice*. Cambridge University Press, 2009.
- [21] Bishop, C. M., *Pattern Recognition and Machine Learning*. Springer, 2006.
- [22] James, G., Witten, D., Hastie, T., and Tibshirani, R., *An Introduction to Statistical Learning*. Springer, 2013.
- [23] Ye, N., *Data Mining: Theories, Algorithms, and Examples*, Boca Raton, Florida: CRC Press, 2013.
- [24] Ye, N., *The Handbook of Data Mining*. Mahwah, New Jersey: Lawrence Erlbaum Associates, 2003.
- [25] Gasse, M., Aussem, A., and Elghazel, H., An experimental comparison of hybrid algorithms for Bayesian network structure learning. *Lecture Notes in Computer Science: Machine Learning and Knowledge Discovery in Databases*, 7523: 58-73., 2012.
- [26] Breiman, L., "Random forests" *Machine Learning (Mach. Learn.)*, 45(1), 5-32, 2001
- [27] Breiman, L., "Arcing classifier", *Ann. Statist (Ann. Stat.)*, 26(3), 801-849, 1998.
- [28] Breiman, L., "Bagging predictors", *Machine Learning (Mach. Learn.)*, 24(2), 123-140, 1996.
- [29] Freund, Y., and Schapire, R. E., "Decision-theoretic generalization of on-Line learning and an application to boosting". *Journal of Computer and System Sciences (J. Comput. Syst. Sci.)*, 55(1), 119-139, 1997.
- [30] Ho, T. K. "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE T. Pattern. Anal.)*, 20(8), 832-844, 1998.
- [31] Kam, H. Tim., "Random decision forest.", *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 1, 278-282, 1995.
- [32] Mason, L., Baxter, J., Bartlett, P. L., & Frean, M. R., "Boosting algorithms as gradient descent" *Advances in Neural Information Processing Systems (Adv. Neur. In.)*, 12, 512-518, 2000.
- [33] Schapire, R. E., "The strength of weak learnability" *Machine Learning (Mach. Learn.)*, 5(2), 197-227, 1990.
- [34] Jones B. D., Osborne J. W., Paretti M. C., Matusovich H. M., "Relationships among students' perceptions of a first-year engineering design course and their engineering identification, motivational beliefs, course effort, and academic outcomes", *International Journal of Engineering Education (Int. J. Eng. Educ.)*, 30(6), 1340-1356, 2014.
- [35] Kline, R. B., *Principles and Practice of Structural Equation Modeling*, New York, NY: The Guilford Press, 2011
- [36] Tsamardinos, I., Brown, L. E., and Aliferis, C. F., "The max-min hill-climbing Bayesian network structure learning algorithm", *Machine Learning (Mach. Learn.)*, 65, 31-78, 2006.
- [37] Ellis, B., and Wong, W. H., "Learning causal Bayesian network structures from experimental data" *Journal of the American Statistical Association (J. Am. Stat. Assoc.)*, 103(482), 778-789., 2008.
- [38] Akutsu, T., Miyano, S., and Kuhara, S., "Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function" *Journal of Computational Biology (J. Comput. Biol.)*, 9(3/4), 331-343, 2000.
- [39] Akutsu, T., Miyano, S., and Kuhara, S., "Inferring qualitative relations in genetic networks and metabolic pathways", *Bioinformatics*, 16(8), 727-734, 2000.

- [40] Akutsu, T., Miyano, S., and Kuhara, S., "Identification of genetic networks from a small number of gene expression patterns under the Boolean network model", In Proceedings of the Pacific Symposium on Biocomputing, 4, 17-28, 1999.
- [41] Bazil, J. N., Qiw, F., Beard, D. A., "A parallel algorithm for reverse engineering of biological networks", Integrative Biology (Integr. Biol.), 3, 1215-1223, 2011.
- [42] D'haeseleer, P., Liang, S., and Somogyi, R., "Genetic network inference: From co-expression clustering to reverse engineering" Bioinformatics, 16(8), 707-726, 2000.
- [43] Liang, S. "REVEAL, a general reverse engineering algorithm for inference of genetic network architectures", In Proceedings of the Pacific Symposium on Biocomputing, 3, 18-29, 1998.
- [44] Marback, D., Mattiussi, C., Floreano, D., "Combining multiple results of a reverse-engineering algorithm: Application to the DREAM five-gene network challenge", Annals of the New York Academy of Sciences (Ann. NY. Acad. Sci.), 1158(1), 102-113, 2009.
- [45] Margolin, A. A., Nemenman, I., Basso, K., Wiggin, C., Stolovitzky, G., Favera R. D., and Califano, A., "ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context", BMC Bioinformatics, 7, Suppl 1, 1-15, 2006.
- [46] Soranzo, N., Bianconi, G., and Altafini, C., "Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks: Synthetic versus real data", Bioinformatics, 23(13), 1640-1647, 2007.
- [47] Stolovitzky, G., and Califano, A. (eds.), Reverse Engineering Biological Networks: Opportunities and Challenges in Computational Methods for Pathway Inference. New York, New York: Blackwell Publishing on Behalf of the New York Academy of Sciences., 2007.
- [48] Frank, A., and Asuncion, A., UCI machine learning repository. <http://archive.ics.uci.edu/ml>, Irvine, CA: University of California, School of Information and Computer Science, 2010.
- [49] N. Ye, "Analytical techniques for anomaly detection through features, signal-noise separation and partial-value associations", Proceedings of Machine Learning Research, 77, 20-32, 2017. <http://proceedings.mlr.press/v71/ye18a/ye18a.pdf>.
- [50] N. Ye, "The partial-value association discovery algorithm to learn multi-layer structural system models from system data", IEEE Transactions on Systems, Man, and Cybernetics: Systems (IEEE T. SYST. MAN. CYB.: SYST.), 47(12), pp. 3377-3385, 2017.
- [51] N. Ye, "A reverse engineering algorithm for mining a causal system model from system data" International Journal of Production Research (Int. J. Prod. Res.), Vol. 55, No. 3, pp. 828-844, 2017. Published online in July 27, 2016. Eprint link: <http://dx.doi.org/10.1080/00207543.2016.1213913>.
- [52] Quinlan, J. R., "Induction of decision trees". Machine learning (Mach. Learn.), 1(1), 81-106, 1986.
- [53] Quinlan, J. R., C4. 5: programs for machine learning. Elsevier, 2014.

Using Fuzzy PD Controllers for Soft Motions in a Car-like Robot

Paolo Mercorelli *

Institute of Product and Process Innovation, Leuphana University of Lueneburg, Volgershall 1, D-21339 Lueneburg, Germany

ARTICLE INFO

Article history:

Received: 26 September, 2018

Accepted: 24 November, 2018

Online: 05 December, 2018

Keywords:

Fuzzy control

Nonholonomic car-like robot

Sensorless cascade control system

ABSTRACT

This paper deals with the control problem for nonholonomic wheeled mobile robots moving on the plane, and in particular, the use of a Fuzzy controller technique for achieving a given motion task which consists of following a rectilinear trajectory until an obstacle occurs on the path. After a background part, in which the fundamental knowledge of Fuzzy control is considered, the problem of the avoidance of an obstacle is taken into consideration. When an obstacle occurs on the path, the drive assistant provides for its avoidance calculating the minimal distance from which the avoidance maneuver starts. Conditions on the parameters of a PD controller are calculated using a Fuzzy based approach. An observer is designed to obtain unmeasurable states to be used in the control loop. In the Appendix of this paper a formal demonstration of a Proposition is proven in which the convergence of the system state estimation of the observer is shown. Simulations considering a real transporter vehicle for a storage service are shown.

1. Introduction

In the field of logistics, the demand for shorter lead times, the highest possible flexibility and high efficiency increases every year. In order to implement these requirements, it is of a great importance that the resources which are responsible for the flow of materials should be optimized and designed so that they can be intuitive and easy to be used. For transportation today as it was in the past, trucks and pallet trucks are used and it is essential to employ personnel for picking within the warehouse. Since a high potential for improving the picking process is still present, a high level of attention will be paid to the optimization of these tasks.

This paper is an extension of work originally presented in *ICCC conference 2018*, [1]. The difference between this work and that already published in [1] is a wide extension of the simulated results together with an extension of the background aspects related to the Fuzzy control and the Luenberger observer. Moreover, the demonstration of the convergence of the estimation is reported in the Appendix of this paper. In this sense, the problem of the observer is considered in depth because represents one of the most challenging problems. The idea is to control the vehicle using the measurement of the steering angular position and the position of the vehicle. The aim of this work is to create a simulation by which the evasive behavior of a newly developed Picking Vehicle can be tested. The results of this work can be applied at a later stage to a

second stage of the vehicle. The simulation allows to test the vehicle's behavior and it gives ideas for the design of the controller parameters. More in depth, this paper deals with the control problem for nonholonomic wheeled mobile robots moving on the plane, and in particular, the problem of the avoidance of obstacles is taken into consideration. When an obstacle occurs on the path, then the drive assistant provides for its avoidance calculating the minimal distance from which the avoidance maneuver starts. Using a Fuzzy approach, conditions on the parameters of a PD controller are calculated. A Luenberger observer is used in the control loop to minimize the number of the sensors. It is known that Luenberger observer is one of the most used observers in motion control and it is a high gain observer, see [2, 3] and [4] and Kalman Filters, see [5] and [6]. Even though the model of a mobile robot is well known in terms of its structure, for controlling mechanisms Fuzzy logic is often used, see [7]. The use of Fuzzy controller is not limited to the system without physical insights but it is shown in [8] that any system can take advantage from such a kind of control strategy. In particular, in [9] the authors describe the design and the implementation of a trajectory tracking controller using Fuzzy logic for mobile robots to navigate in the indoor environments. Most of the previous works used two independent controllers for navigation and avoiding obstacles. Also in [10] the salient feature of the proposed approach is that it combines the Fuzzy gain scheduling method and a Fuzzy proportional-integral-derivative (PID) controller to solve the nonlinear control problem. The paper is organized in the

*Corresponding Author Paolo Mercorelli, Institute of Product and Process Innovation, Leuphana University of Lueneburg, Volgershall 1, D-21339 Lueneburg, Germany, Email: mercorelli@uni.leuphana.de

following way. Session 2 presents the model of a general nonholonomic system. Session 3 is devoted to the presentation of the Luenberger observer. Session 4 is dedicated to the Fuzzy and PD control strategy. Session 5 shows the simulated results and the conclusions close the paper. At the end of the paper in an Appendix, a Proposition is proven in which the convergence of the estimation of the system state variables is shown.

2. The Nonholonomic System

The first step in developing the kinematic model, which will be used below, is the consideration of nonholonomic conditions. The model used is that of robots with car-like properties. Therefore, the properties of the robot will be described by means of the position and the orientation of the vehicle along the route and the angle of the wheels to be controlled. The main feature of the mobile vehicle similar to a robot is the presence of nonholonomic constraints, since they are based on the condition of getting around without slipping on the ground. This means that the movements of the vehicle are restricted and the vehicle cannot move freely in any direction, see [11]. One example is the parallel parking a car. To get into the parking space it is not possible to drive the vehicle easily sideways. It must be moved forward or backward and be maneuvered by means of changing the steering angle of the front wheels into the parking space. To understand the conditions of the nonholonomic system, at the beginning the case of each wheel will be considered. The speed of the wheel is orientated in the direction in which the wheel moves.

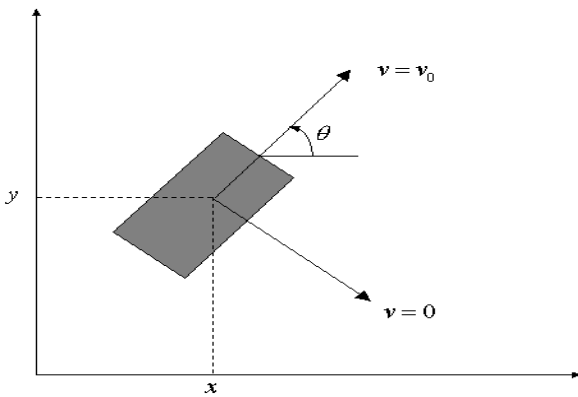


Figure 1. Nonholonomic system: (Mellodge, Feedback Control for a Path Following Robotic Car, 2002)

Due to the nonholonomic properties speed in other directions is not possible. The properties of each single wheel can be described using a vector consisting of three generalized coordinates. The position coordinates x , y within a fixed coordinate system in which the wheel touches the ground and the angle θ , which is the alignment of the wheel to the x -axis. The generalized velocities q cannot accept independent values, they are subject to the following constraint, see Figure 1 from [12]:

$$[\sin \theta \ -\cos \theta \ 0] \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = 0. \quad (1)$$

Equation (1) constitutes a particular expression of a kinematic constraint, a Pfaffian constraint. In this case, the Pfaffian

$C(q)\dot{q} = 0$, for example, is linearity in the generalized speeds. As a result, all the generalized velocities in the core of the matrix $C(q)$ are included, see [13].

For the individual wheel, this results in the following model:

$$\dot{q} = \begin{bmatrix} \cos \theta \\ \sin \theta \\ 0 \end{bmatrix} v_1 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} v_2. \quad (2)$$

In this case, v_1 represents the linear speed of the wheel and v_2 represents the angular velocity around the vertical axis.

2.1. Kinematic Model of the Vehicle with Global Coordinates

The kinematic model of a vehicle is often used because of its simplicity and its accuracy when the vehicle behavior must be shown under normal driving conditions, see [14]. The model of a car-like robot vehicle is an option for the kinematic model. This can be described (see Figure 2) on the basis of four variables, see [14].

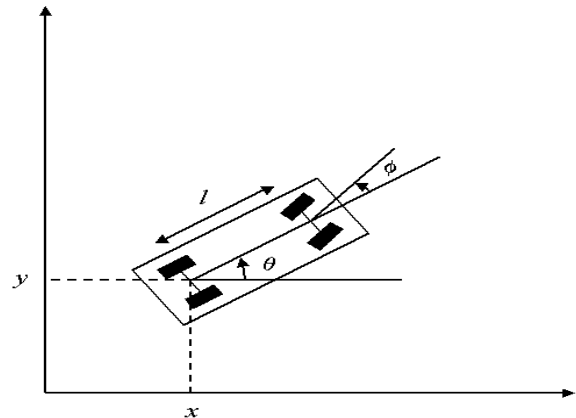


Figure 2. Car model with global coordinates (Source: (Mellodge & Kachroo, 2008), page 31)

x and y represent the Cartesian coordinates of the rear axle of the vehicle. The angle θ is the orientation of the vehicle with respect to the x -axis and the angle ϕ which shows the steering angle of the front wheels of the vehicle, θ with respect to the alignment. Because of front and rear axles, the system is subject to two different nonholonomic conditions:

$$\begin{aligned} \dot{x}_f \sin(\theta + \phi) - \dot{y}_f \cos(\theta + \phi) &= 0 \\ \dot{x} \sin \theta - \dot{y} \cos \theta &= 0, \end{aligned} \quad (3)$$

where x_f and y_f represent the coordinates of the front axle, x and y represent the coordinates of the rear one. From the conditions shown in Figure 2, the current velocities in x and y directions can be calculated, in which v_1 represents the speed at the rear wheels:

$$\begin{aligned} \dot{x} &= v_1 \cos \theta \\ \dot{y} &= v_1 \sin \theta. \end{aligned} \quad (4)$$

Taking the rear axle as a reference point and taking into account the distance between the two axes, the result for the front axle is:

$$\begin{aligned} x_f &= x + l \cos \theta \\ y_f &= y + l \sin \theta. \end{aligned} \quad (5)$$

This results in the position of the derivative to the speeds:

$$\begin{aligned} \dot{x}_f &= \dot{x} + l \dot{\theta} \cos \theta \\ \dot{y}_f &= \dot{y} + l \dot{\theta} \sin \theta. \end{aligned} \quad (6)$$

If the new conditions of formula (2.6) in the nonholonomic conditions of the front axle (2.3) are used and solved for θ , then the constraint of the front axle results to be:

$$\dot{\theta} = \frac{\tan \phi}{l} v_1. \quad (7)$$

From this, the entire model results as follows, see also [15]:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \\ \frac{\tan \phi}{l} \\ 0 \end{bmatrix} v_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} v_2, \quad (8)$$

v_1 thereby represents the speed of the rear wheels and v_2 represents the speed of the steering angle or the change of the steering wheels. In order to investigate the nonlinear system from the previous equation, the principle of the Lie algebra provides the ability to perform a controllability analysis. The existing system is known to be a nonlinear and a drift-free one due to the nonholonomic conditions of the system. A general description of the kinematic nonlinear system is shown below in equation (9):

$$\dot{q} = G(q)v. \quad (9)$$

Here, q represents the n vector of the general coordinates.

v is the m -vector of the input speeds, and it must be smaller than the amount of the general coordinates ($m < n$). The matrix G is divided into a plurality of uniform g_i columns, where ($i = 1, \dots, m$) is valid. For the model of the vehicle like robot which is used, $n = 4$ different coordinates and $m = 2$ controllable inputs are valid. This results in the following model which is used for the consideration of controllability, see [13].

$$\dot{q} = g_1(q)v_1 + g_2(q)v_2, g_1 = \begin{bmatrix} \cos \theta \\ \sin \theta \\ \frac{\tan \phi}{l} \\ 0 \end{bmatrix}, g_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \quad (10)$$

The two vector fields and g_1 and g_2 allow driving and steering of the model. Apparently only the directions given by g_1 and g_2 are possible, but in reality, it is known that practically any position can be approached with a car. To describe the process mathematics provides the method of Lie derivative, which creates a new vector field out of two vector fields. In [12] the Lie brackets are designed as follows:

$$[g_1, g_2, [g_1, g_2], [g_1, [g_1, g_2]]]. \quad (11)$$

The brackets g_1 and g_2 are already known, the other two arise by means of the Jacobian matrix:

$$J_f(x) = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} & \dots & \frac{df_1}{dx_n} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} & \ddots & \vdots \\ \frac{df_3}{dx_1} & \frac{df_3}{dx_2} & \dots & \frac{df_m}{dx_n} \end{pmatrix} \quad (12)$$

to:

$$[g_1, g_2](x) = \frac{dg_2}{dx} g_1(x) - \frac{dg_1}{dx} g_2(x) = \begin{bmatrix} 0 \\ 0 \\ -1 \\ \frac{1}{l \cos^2 \phi} \\ 0 \end{bmatrix}, \quad (13)$$

and

$$[g_1, [g_1, g_2]] = \frac{d[g_1, g_2]}{dx} g_1(x) - \frac{dg_1}{dx} [g_1, g_2](x) = \begin{bmatrix} -\frac{\sin \theta}{l \cos^2 \phi} \\ \frac{\cos \theta}{l \cos^2 \phi} \\ 0 \\ 0 \end{bmatrix}. \quad (14)$$

Further matrices are not present and therefore, the individual matrices can be merged to the overall matrix G . Finally, the rank of the matrix will be determined. Matrix G is the following:

$$G = \begin{bmatrix} \cos \theta & 0 & 0 & -\frac{\sin \theta}{l \cos^2 \theta} \\ \sin \theta & 0 & 0 & \frac{\cos \theta}{l \cos^2 \theta} \\ \frac{\tan \theta}{l} & 0 & \frac{1}{l \cos^2 \theta} & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (15)$$

From the consideration of matrix G , the rank of this matrix results to be equal to four, because all four columns are independent for all θ . From the middle two columns it is directly visible that they cannot be zero. The two outer columns should be specifically considered, because of the danger that for $\phi = 0$ a dependency can occur. For this, the determinant of the matrix G was calculated by the symbolic toolbox of MATLAB. This investigation showed that except in the case of a higher or lower steering angle of $\phi = \pm \pi/2$ there is no dependency. It follows that the system is controllable, as long as it is within the valid steering angle, see [5].

2.2. Kinematic Model of the Vehicle with Path Coordinates

The Global Coordinates model of the vehicle presented in subsection (3.1) is useful for the purpose of simulation and of good use to gain first insights into the behavior. However, for an application in practice, it is not particularly suitable in most cases because the sensors have difficulty to determine the position of the vehicle in relation to global coordinates. Therefore, a new model must be introduced for the rest of the work. This describes the position of the vehicle as dependent of the path (Figure 3). The distance, which the vehicle must keep to the wall and to obstacles

is indicated by the variable d and measured by the laser scanner, which is located centrally on the vehicle front side. The distance between the front and the rear axle is defined by l . The angle θ reproduces the orientation of the vehicle with respect to the x-axis again as also in the global coordinate system. θ_t indicates the angle between a tangent to the path and the x-axis. In order the vehicle to be able to follow the path, an angle θ_p which indicates the difference between the orientation of the vehicle and the angle of the tangent θ_t is defined:

$$\theta_p = \theta - \theta_t. \tag{16}$$

For the desired path, tracking this difference should tend as much as possible to zero. The distance traveled is defined by s and is an arbitrarily selectable starting position. The course of the track, or for instance, the curvature is given by $c(s)$ and defined as follows:

$$c(s) = \frac{d\theta_t}{ds} \text{ or } \dot{\theta}_t = c(s)\dot{s}. \tag{17}$$

Furthermore, the speeds along the track \dot{s} and the speed required perpendicular to the distance \dot{d} for the model will be required:

$$\begin{aligned} \dot{s} &= v_1 \cos \theta_p + \dot{\theta}_t d \\ \dot{d} &= v_1 \sin \theta_p. \end{aligned} \tag{18}$$

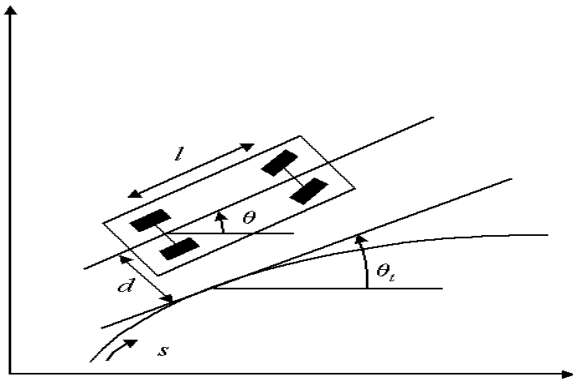


Figure 3. Kinematic Model with Path coordinates (Source: (Mellodge & Kachroo, 2008), page 33)

The overall model based on the redefined variables with Path Coordinates, as it can be seen in Equation (19), can be set up.

$$\begin{bmatrix} \dot{s} \\ \dot{d} \\ \dot{\theta}_p \\ \dot{\varphi} \end{bmatrix} = \begin{bmatrix} \frac{\cos \theta_p}{1-dc(s)} \\ \sin \theta_p \\ \frac{\tan \theta}{l} - \frac{c(s) \cos \theta_p}{1-dc(s)} \\ 0 \end{bmatrix} v_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} v_2. \tag{19}$$

3. Observer

The state variables must be supplied to the controller for the control of dynamic systems or control of the physical parameters of the track. These quantities are usually measured. However, if this is not possible, or only possible with a great effort, they have to be determined in another way. For this, an observer can be used, which reconstructs a state (condition) based on the course of input

and output variables. Here, the model of the controlled system is connected in parallel to the actual process and supplied with the same input dimension. If the model is correct, it performs the same action as the control path and differences can thus only occur through the different initial conditions. If the control system is stable and it is possible to wait long enough, the observer and the controlled system commute into the same sizes and it is: $\hat{y}(t) = y(t)$, and $\hat{x}(t) = x(t)$ and $\hat{x}(t)$ can be used as a state feedback. This simple variant has major lacks:

- The control path $x(t)$ will only consider a disturbance size, while the model $\hat{x}(t)$ remains unchanged, when a disturbance occurs. Therefore, this method is only suitable for undisturbed, uninteresting cases for control.
- Secondly, a difference $x(t) - \hat{x}(t)$ generated by different initial states \hat{x}_0 and x_0 is only compensated when the controlled system is stable. Out of the equation of the state of the controlled system:

$$\dot{x}(t) = Ax(t) + Bu(t), x(0) = x_0 \tag{20}$$

with $\hat{x}(t)$ the expression follows:

$$x(t) - \hat{x}(t) = e^{At}(x_0 - \hat{x}_0). \tag{21}$$

This however is only possible for a stable matrix A against zero, regardless of whether the unstable controlled system is stabilized within the control loop by a state feedback. Since only the system matrix of the controlled system and not of the closed circuit enters into the relationship, the simple observer is useful only for stable systems. Based on this principle, David G. Luenberger has developed an extension of this model in 1964. However, this observer is expanded to the difference between the output of the model y and the output of the control path \hat{y} . This difference is used in order to equalize the state of the model to the route. This method has already proven itself in the control loop, there, however, the deviation of the controlled variable is used to perform a control intervention, which influences the controlled variable. Based on the following non-jump-capable control system the subsequent design of the observer follows:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ x(0) &= x_0 \\ y(t) &= Cx(t). \end{aligned} \tag{22}$$

In the design of the observer, the system model is supplemented by the input u_B :

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + u_B(t), \hat{x}(0) = \hat{x}_0 \\ \hat{y}(t) &= C\hat{x}(t). \end{aligned} \tag{23}$$

In the new input u_B the difference between the measured output of the control path y and the output of the model \hat{y} is returned to be:

$$u_B(t) = L(y(t) - \hat{y}(t)). \tag{24}$$

If equations (3.4) and (3.5) are put together, we obtain the following relationship:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + LC(x(t) - \hat{x}(t)). \quad (25)$$

After merging, the effect of the newly introduced input is recognized. As long as $\hat{x}(t) = x(t)$ is true, the right part falls away, because the addend becomes zero and the model thus runs in parallel to the control path. If a deviation between the input variables, which influence the outputs, occurs, so the behavior of the model due to the feedback will be influenced. In order to reduce the deviation, a matrix L must be found which decreases the difference of $x(t) - \hat{x}(t)$. The input vector $x(t)$ is not known in practice as a rule, however, the output vector $y(t)$, serving as an input of the observer. The result for the observer is a controlled dynamic system with the two inputs $u(t)$ and $y(t)$ as in equation (3.7).

$$\dot{\hat{x}}(t) = (A - LC)\hat{x}(t) + Bu(t) + Ly(t). \quad (26)$$

In the simulation, the observer has the following characteristics:

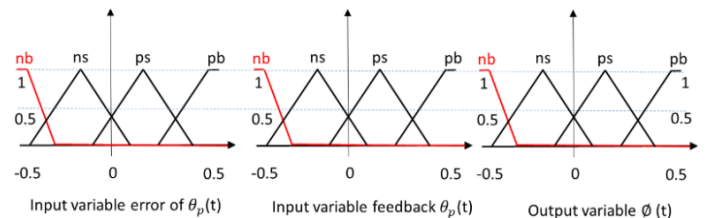
$$\begin{bmatrix} \dot{\hat{s}} \\ \dot{\hat{d}} \\ \dot{\hat{\theta}}_p \\ \dot{\hat{\phi}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{s} \\ \hat{d} \\ \hat{\theta}_p \\ \hat{\phi} \end{bmatrix} + \begin{bmatrix} \frac{\cos \theta_p}{1-dc(s)} \\ \sin \theta_p \\ \frac{\tan \phi}{l} - \frac{c(s) \cos \theta_p}{1-dc(s)} \\ 0 \end{bmatrix} v_1 + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} v_2 + \begin{bmatrix} k1 \\ k2 \\ k3 \\ k4 \end{bmatrix} [d(t) - \hat{d}(t)]. \quad (27)$$

The principle of the Luenberger observer is based on a linear system, in which matrices with fixed values are integrated. However, the system in the work is a nonlinear one, in which the matrix B represents a field with variables. The principle of the observer remains the same as the one in the linear version. The stability of the observer seems to be given, however, it was only heuristically determined. To investigate the stability more precisely, the stability theory of Lyapunov is used and in the Appendix at the end of the chapter a formal analysis is conducted.

4. Fuzzy Control Algorithm integrated with a PD controller

In addition to methods for mathematically analytical problem solving of control and monitoring tasks, now human experience and knowledge that cannot be expressed mathematically for problem solving are being increasingly used. For this purpose, the Fuzzy logic is often used. In contrast to the classical control engineering, the behavior of a system is tried to be described with traditional means and then a controller will be tried to be designed with analytical methods. Fuzzy systems are particularly well suitable for the design of systems under a vague knowledge. For example, they are well suitable for not exactly known process or a not exactly known controller behavior. Due to the completely different approaches for both controller designs, completely different solution approaches for Fuzzy and classical approaches for regulation are used. While in the classical, for instance PD scheme, in a first step a model of the controlled-track is formed,

in the second step the associated controller design follows only using the formed path. This procedure can therefore be described as a model-driven interpretation. The Fuzzy controller represents a contrast for it, since, this is controller-oriented. Subsequently, the basic concepts of Fuzzy control, and the steps of the design of a Fuzzy controller are described. For execution of the control mechanism in Fuzzy logic three actions must be performed. In the first part, the sharp values, which are supplied by the input variable, will be transformed into the fuzzification. The physical values of input values become the converted Fuzzy sets. These are a specific form of Fuzzy sets, in which the values are not described as usually in mathematics by numerical variables in the form of sharp numbers, but by colloquial expressions. Here formulations arise such as: when the distance to the wall is too small, the distance must be increased somewhat by increasing the steering angle. Here, the descriptions “distance slightly too small” and “steering angle slightly increase” represent the blurry relationship between the distance to the wall and the steering angle. The variables, which result from this, are called linguistic variables and they will be assigned to a membership function. The membership function makes an indication, in what proportion between 0 and 1, the value of a blurry statement is true. It also provides a statement to what extent an element belongs to a certain set. In literature fuzzy controller is used also for nonholonomic systems. In [16] a fuzzy PD controller for a dynamic model of nonholonomic mobile manipulator in order to treat the trajectory tracking control and to eliminate the effect of external force on the end-effector is proposed. Concerning the robustness of the Fuzzy approach, the paper in [17] addresses the output feedback trajectory tracking problem for a nonholonomic wheeled mobile robot in the presence of parameter uncertainties, external disturbances, and a lack of velocity measurements. A combination between a heuristic Fuzzy and PID controller is also designed in [18] to move the robot upward or downward the inclined plane and approach the target point. In [19] fuzzy adaptive observers together with parameter adaptation laws are designed to estimate the state-dependent disturbances in both kinematics and dynamics in order to adapt a Fuzzy adaptive controller.



In [20] a holistic, for holonomic and nonholonomic systems, intelligent control strategy is proposed in which only position measurements are used.

In Figure 4, the two inputs and the output of one of the two Fuzzy controllers are presented as an example. Here the different

Fuzzy sets of the two inputs and the output can be seen. The linguistic variable “ns” means “negative small” and is associated with an interval 0.1 to -0.4. Within the interval there are different degrees of association, the value 0.15 with a factor of 1 is the most strongly associated value and the association decreases in negative and positive direction until at 0.1 and -0.4 the association of 0 is achieved. Such linguistic variables are defined over the entire area to be controlled, if possible, they must have an overlap at their adjacent variable, so that no areas remain undefined. In order to provide a high degree of flexibility in different executions such as trapezoidal, Gaussian, or triangular shaped, the associated functions can be defined and combined with each other. MATLAB provides for different execution forms. In the present work, the trapezoidal and triangle-shaped execution have been chosen.

Table 1. Fuzzy rules

Output $\theta(t)$		Angle $\theta(t)$			
		nb	ns	ps	pb
Error $\theta(t)$	nb			nb	nb
	ns			ns	
	ps	ps	ps		
	pb	pb			

Output $d(t)$		Distance $d(t)$			
		nb	ns	ps	pb
Error $d(t)$	nb			nb	nb
	ns			ns	ns
	ps	ps	ps	ps	ps
	pb	pb	pb	pb	pb

In Table 1 all the used fuzzy roles in this application are indicated. For instance, the linguistic variable “nb” stands for “negative big,” “ps” stands for “positive small” and “pb” stands for positive big.

After the fuzzification was performed, the inference follows. The rules for evaluation will be developed with help of an expert or by a private experience. These consist of two parts, the if-condition (premise): IF the distance is too small and the Fuzzy inference (conclusion): THEN the steering angle must be increased slightly. The input variables will be linked by means of an AND-operation in the IF part. Then the resulting rules are connected by means of an OR-operation. So, the following rules resulted for the conditions used in this work. The possible Fuzzy sets for the deviation from the desired value are referred here as an error. The Fuzzy sets that are delivered as a response of the system, are referred to as a feedback. A 4x4 matrix follows from it, in which not every field is filled. The actual inference takes place within three steps after the rules having been defined:

- 1) Aggregation (evaluation of the rule premises): The rule premises are evaluated here, that is the IF-parts of rules corresponding to their affiliations and selected operators. The input variables were associated with an AND-operator, in this case a min-operator results to be used, which always selects the smallest affiliation value of a Fuzzy set. From a human point of view, it acts pessimistically and not compensatory.
- 2) Implication (evaluation of the conclusions (THEN-part)): Evaluation of the conditions occurs during the implication. The rules can also only be partially true in the Fuzzy control,

in contrast to the classical control. Therefore, in this step, a definition follows in which the dimension of the condition is true or not. The minimum procedure will be used for it (see Figure 5), which assumes as a conclusion the minimum of the premise (IF-part) and conclusion (THEN- part). In simple words, this means that the output function as a result of this operation, is cut off at the level of fulfillment of the premise and an area in the output Fuzzy set arises.

- 3) Accumulation (summary of all rules); The conclusions of all the rules will be summarized in the accumulation. This is necessary because, as a rule not only one rule is true, but several rules can be at least partially valid. Due to the fact that the rules are connected with each other through the Or-links, here the maximum operator is used in the selection of the conclusions. Here the maximum value of the Fuzzy sets is used as the output function, see Figure 6. Since the Min-method was used for the implication and the Max-method was used for the accumulation, the process which is applied in the work is called a max / min inference.

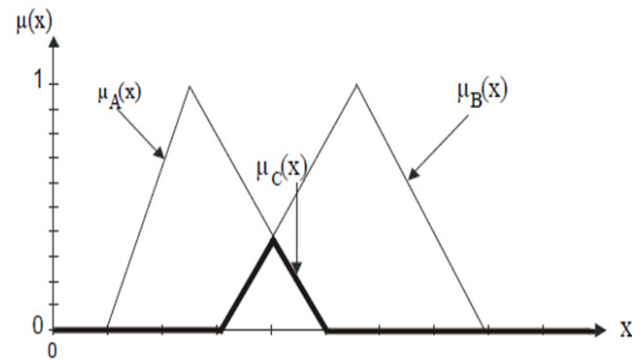


Figure 5. Minimum Operator

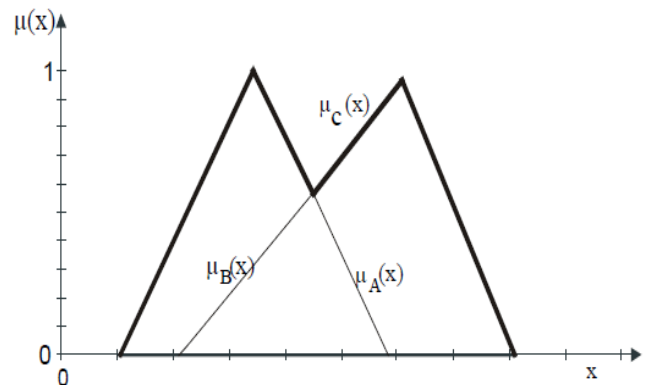


Figure 6. Maximum Operator

Defuzzification is carried out during the last step, after the inference was completed. A new Fuzzy set will turn out to be from the inference, and thus a Fuzzy information will be delivered. So that other participants can process the signals, sharp output values must be generated again. This is done using the gravity method. Here, a Fuzzy set resulting from the inference is considered as a totality.

As shown in Figure 7, the Fuzzy set became a set B* which determines the output variable. The centroid is now determined by the gravity method and the gravity coordinate u_s gives it as a

defuzzification result. The coordinate of the centroid is calculated by the following formula:

u_s = sharp output value

u_i = abscissa bases

$\mu_B * (u_i)$ = membership degree for u_i

q = number of abscissa bases

$$u_s = \frac{\int_{u_{min}}^{u_{max}} u * \mu_B * u (d_u)}{\int_{u_{min}}^{u_{max}} \mu_B * u (d_u)} \approx \frac{\sum_{i=1}^q u_i * \mu_B(u_i)}{\sum_{i=1}^q \mu_B(u_i)} \quad (28)$$

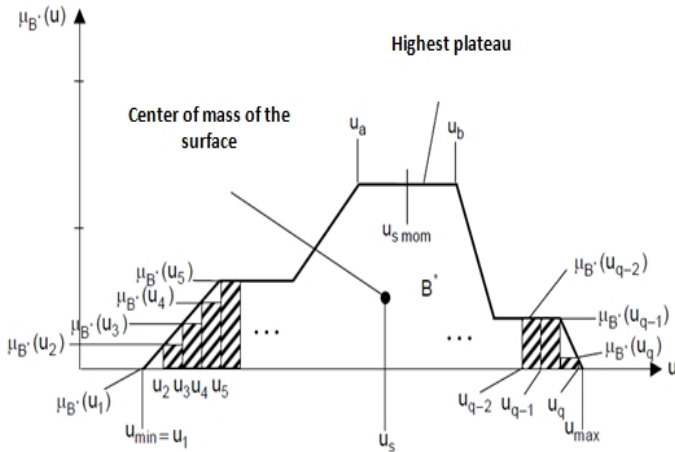


Figure 7. Defuzzification by the centroid method (centroid)

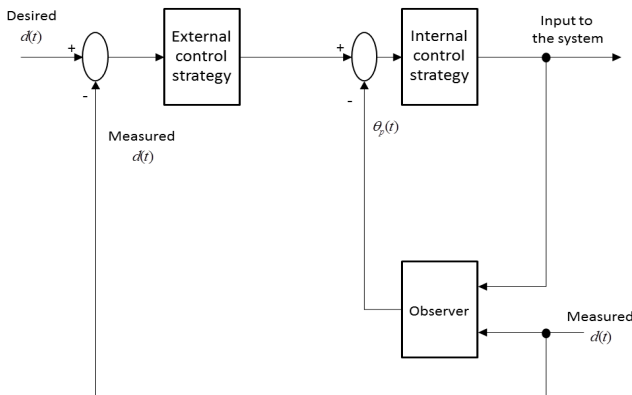


Figure 8. Simplified diagram of the cascaded control loop

During the calculation according to the gravity method, all active rules determine the sharp output. The calculation methods and criteria for the selection of operators are selected in Simulink within the Fuzzy Editor and run subsequently automatically.

Figure 8 shows the general structure of the cascade control scheme including the Luenberger observer. In Figure 8 it is visible how the observer is issued. Through the measurement of the distance between the car and the wall all other state variables can be observed.

Here, the outer and the inner circle of the PD and the Fuzzy control with help of an adder were interconnected. This is shown in Figure 9.

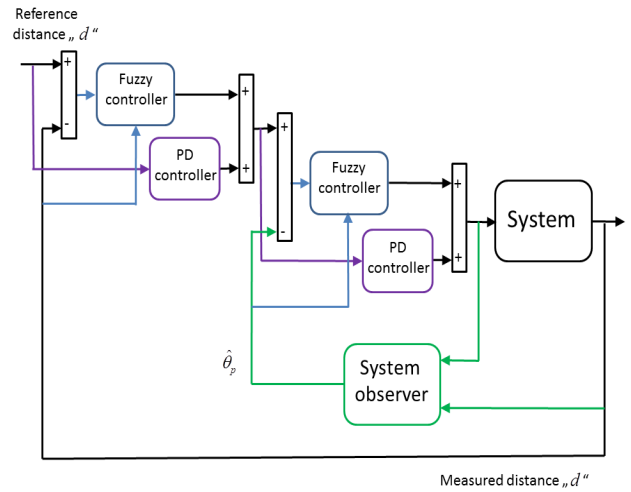


Figure 9. Simulink diagram combining PD and Fuzzy controllers in a cascade control structure together with an observer

Remark 1

The choice of this particular Fuzzy logic is due to the fact that, considering more variables than proposed, not relevant improvements are obtained in terms of reduction of error.

4.1. A Gaussian Curve to Avoid Obstacle

A Gaussian function offers the possibility to obtain the highest possible, incremental and without jumps resulting function to avoid obstacles. This guarantees due to its properties, in addition to a slow and continuous adjustment of the steering angle, a high degree of stability in its derivation. This property proves to be of a great advantage, because the derivative is required for further calculations within the simulation. To achieve a high degree of flexibility and influence a formed function, which is similar to a Gaussian function, but offers more free parameters, and thus becomes a Gaussian-like function. The calculation is based on the freely selectable amplitude H, which determines the maximum level of the function. In order to influence the abdomen of a function, upsetting or a stretching factor was introduced. This will be automatically calculated based on the obstacle width in a sub-function. The following formula is a result for the curve:

H = amplitude of the curve

o = factor for upsetting or stretching of the curve

t = velocity factor

$$G(t) = H * e^{-\left(\frac{t}{o}\right)^2} \quad (29)$$

The derivative of the curve is thus obtained as follows:

$$\dot{G}(t) = -H * e^{-\left(\frac{t}{o}\right)^2} * 2 \left(\frac{t}{o}\right) * \frac{1}{o} \quad (30)$$

Factor “t” states, in this case, the speed of the vehicle and the phase shift, which serves as a correction factor. In the following part the overall configurations of the various control concepts and their implementation in Simulink were presented.

5. Results

Figure 10 shows that the cascade PD control structure increases significantly faster the distance to the wall than the other two control configurations. This requires only 8 seconds, in the meantime the cascade Fuzzy control structure, represented in Figure 11, requires about 12 seconds to bring the vehicle at the same distance. The combination of both controllers manages to do it in just over 10 seconds, see Figure 12. During the fallback procedure, which starts at about 25 seconds, the quality of the results of the controller changes, the PD controller shows here his weakness. In the case of the example shown, this operation supplies during the evacuation process the worst results. It swings at the desired distance from $\approx \mp 6$ cm. The Fuzzy Control System in comparison provides results of $\approx \mp 5$ cm, which accounts for a difference of $\approx 17\%$. The combination of the two controllers presents the best results. This results in a maximum deviation of $\approx \mp 2$ cm, which means a difference of $\approx 67\%$ compared to the PD controller. Thus, the combination of the two controllers in the simulation provides the best results, although it increases the distance to the wall something slower than the PD it performs, but it has a significantly higher stability during the fallback procedure. It is important to be able to configure the smallest possible protective field because the aisles in which the vehicle should move anyway already offer little space and thus the smallest possible protective field is necessary. A comparison between the model of the kinematic system and the estimated values of the observer based on the distance d is also visible.

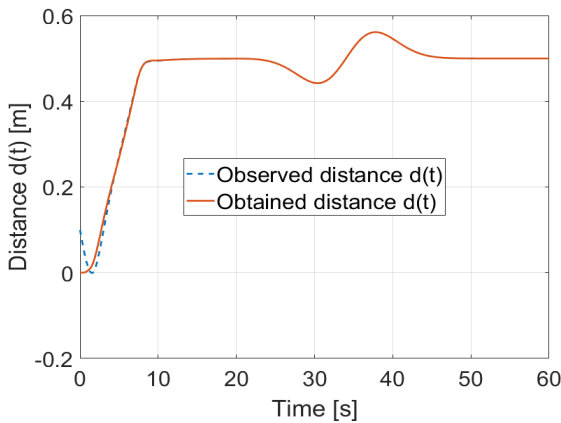


Figure 10. Distance of the car from the wall using the cascade PD control structure

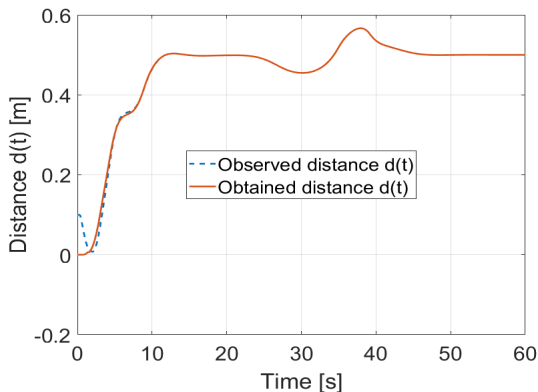


Figure 11. Distance of the car from the wall using the cascade Fuzzy control structure

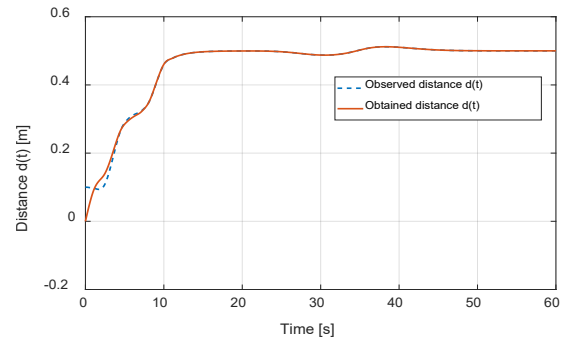


Figure 12. Distance of the car from the wall using the combination between PD and Fuzzy cascade control structure

To ensure that the vehicle does not slip during the evacuation process even with a heavy load, or does not even lose some of the load and also to be able to configure a minimum possible protective field, the steering angle may not exceed a maximum impact of $\mp 10^\circ$. Figures 13, 14 and 15 show the time progress of the steering angle in degrees using PD, Fuzzy and a combination of PD and Fuzzy controllers respectively. As it can be seen there, the requirement of the maximum steering angle has been met for all three schemes.

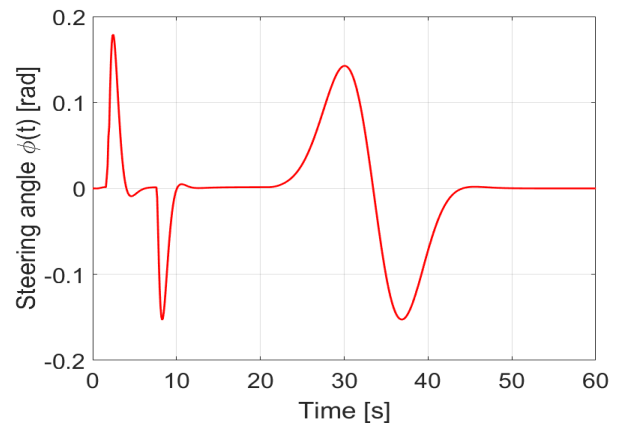


Figure 13. Steering angle ϕ using the PD cascade controller structure

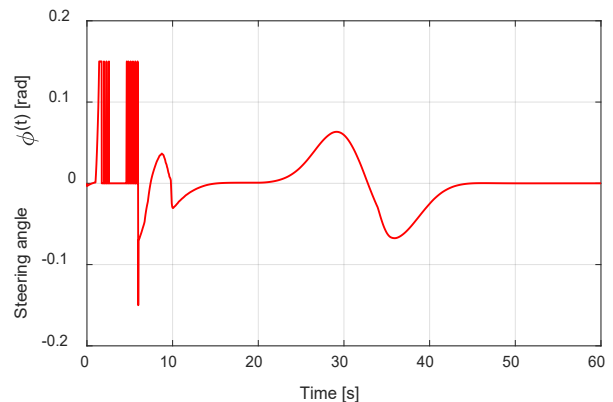


Figure 14. Steering angle ϕ using the Fuzzy cascade controller structure

However, it was shown that the PD control with $\mp 10^\circ$, during the evasive maneuver, is still within the required range, but the other two controls are significantly lower. Thus, during the evasive maneuver, the Fuzzy control with a range of $\approx \mp 8^\circ$ and

the combination of both controls, PD and Fuzzy controllers, lead the comparison with a range of $\approx \pm 6^\circ$. As it has already been described, also the combination of the two controls leads if we compare the simulation results, because it requires the least amount of the steering angle. Advantages of this are, as it has already been explained above, that no abrupt changes of direction as a result of a strong steering maneuver occur due to the small steering angle and the avoidance maneuver has fewer risks like slippage of the vehicle, or a shifting or falling down of the load. Figures 13, 14 and 15 show the simulation results concerning these aspects.

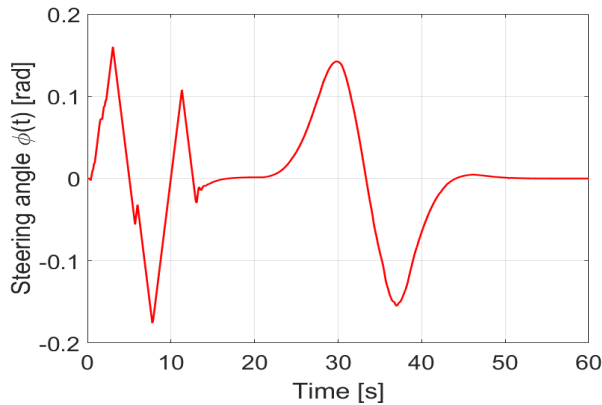


Figure 15. Steering angle ϕ using the combination between PD and Fuzzy cascade control structure

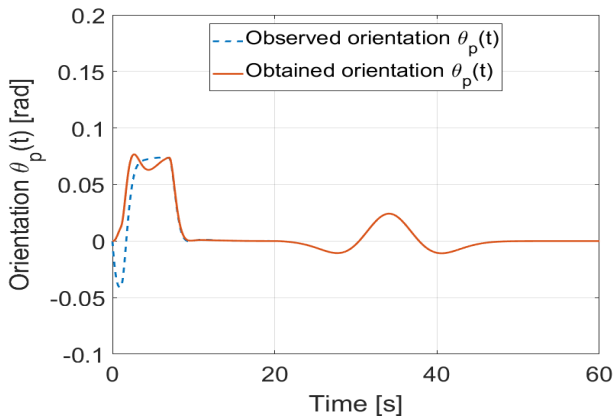


Figure 16. Angle θ_p using the PD cascade controller structure

5.1. Results of the Vehicle Orientation

The vehicle as it is described must follow the predetermined path. For this purpose, the differential angle θ_p was introduced which describes the difference between the angle to the path θ_t and the actual vehicle heading θ . Figures 16, 17 and 18 show the time course of the difference angle θ_p expressed in degrees. For the route to follow this best, the difference angle must be kept as small as possible. The initially large deflection is due to the correction, in order to achieve the already mentioned predetermined distance and in this case, less relevant, since this is not to be taken into account for the actual viewing of the path. Interestingly, the course is between 20 and 50 seconds since the avoidance of the obstacle occurs, which goes along with the path. Here, PD and Fuzzy controllers deliver almost equivalent results, but the

maximum excursions hardly differ so they both achieve maximum angle of $\approx \pm 1.5^\circ$ during the evasive maneuver. It is different with the result of the combination of the two schemes, this is as well as the distance and the steering angle significantly better. Here is a maximum angle of $\approx \pm 0.5^\circ$ during the evasive maneuver and thus the best result. A comparison between the model of the kinematic system and the estimated values of the observer based on the angle θ_p is also visible.

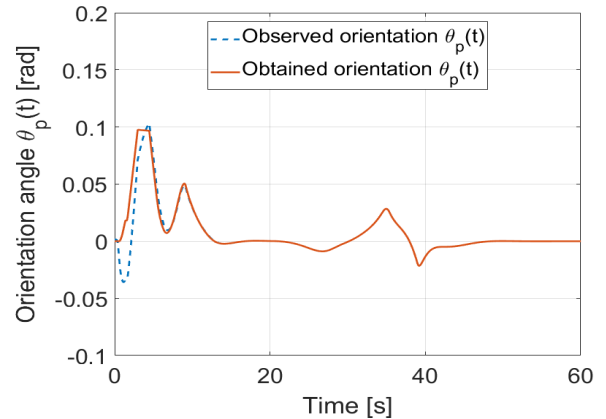


Figure 17. Angle θ_p using the Fuzzy cascade controller structure

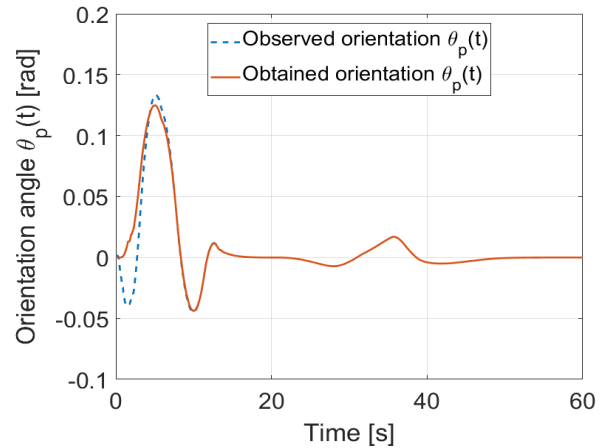


Figure 18. Angle θ_p using the combination between PD and Fuzzy cascade control structure

Figure 19 shows the error of the distance and Figure 20 shows the error of the orientation of the car. In both cases it is possible to see how after 45 seconds the error results to be almost equal to zero.

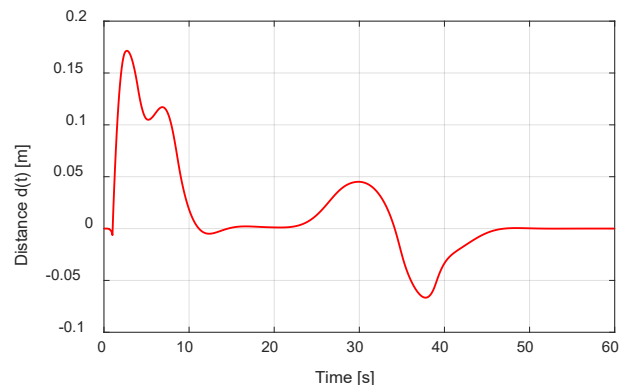


Figure 19. Error of the distance of the car from the wall

Remark 2

To discuss the obtained results with the already existing contributions it is possible to say that in this contribution an original combination of Fuzzy control strategy combined in a PD controller in a cascade structure is presented. This idea is an original one, which is not present in the already existing literature as already discussed through the cited literature.

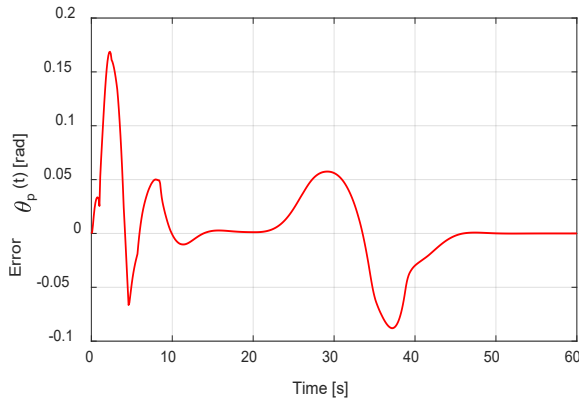


Figure 20. Error of angle θ_p

6. Conclusions

This paper deals with the control problem for nonholonomic wheeled mobile robots moving on the plane to avoid obstacles. Parameters of a PD controller are calculated using a fuzzy based approach. To estimate the orientation of the vehicle a Luenberger observer is involved in the control scheme. In the context of the Luenberger observer, the demonstration of the convergence of the estimation of the proposed observer is shown in the Appendix of the paper. Simulations considering a real transporter vehicle for a storage service are shown.

References

[1] P. Mercorelli, "Fuzzy based control of a nonholonomic car-like robot for driveassistant systems", *2018 19th International Carpathian Control Conference (ICCC)*, Szilvasvarad, pp. 434-439, 2018.

[2] P. Mercorelli, "An adaptive and optimized switching observer for sensorless control of an electromagnetic valve actuator in camless internal combustion engines". *Asian Journal of Control*, 16:959-973, 2014.

[3] P. Mercorelli, "A two-stage sliding-mode high-gain observer to reduce uncertainties and disturbances effects for sensorless control in automotive applications". *IEEE Transactions on Industrial Electronics*, 62(9):5929-5940, 2015.

[4] P. Mercorelli, "A motion-sensorless control for intake valves in combustion engines". *IEEE Transactions on Industrial Electronics*, 64(4):3402-3412, 2017.

[5] P. Mercorelli, "A two-stage augmented extended Kalman filter as an observer for sensorless valve control in camless internal combustion engines". *IEEE Transactions on Industrial Electronics*, 59(11):4236-4247, 2012.

[6] P. Mercorelli, "A hysteresis hybrid extended Kalman filter as an observer for sensorless valve control in camless internal combustion engines". *IEEE Transactions on Industry Applications*, 48(6):1940-1949, 2012.

[7] C. Zheng, Y. Su, P. Mercorelli, "A simple fuzzy controller for robot manipulators with bounded inputs". *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, 737-1742, 2017.

[8] G. Feng, "A Survey on Analysis and Design of Model-Based Fuzzy Control Systems". *IEEE Transactions on Fuzzy Systems*, 14: 676-697, 2006. doi: 10.1109/TFUZZ. 2006.883415.

[9] H. Omrane, M. S. Masmoudi, M. Masmoudi, "Fuzzy Logic Based Control for Autonomous Mobile Robot Navigation", *Computational Intelligence and Neuroscience*, 2016, Article ID 9548482.

[10] Y. L. Sun, M. J. Er, "Hybrid fuzzy control of robotics systems", *IEEE Transactions on Fuzzy Systems*, 12: 755-765, 2004. doi: 10.1109/TFUZZ.2004.836097.

[11] F. Fahimi, *Autonomous Robots Modeling, Path Planning, and Control*. New York: Springer Science & Business Media, 2009.

[12] P. Mellodge, "Feedback Control for a Path Following Robotic Car". *M.S. Thesis*, Blacksburg: VirginiaTech, 2002.

[13] J.-P. Laumond, *Robot Motion Planning and Control*. London: Springer, 1998.

[14] P. Mellodge, K. Pushkin, *Model Abstraction in Dynamical Systems: Application to Mobile Robot Control*. Berlin Heidelberg: Springer, 2008.

[15] A. De Luca, G. Oriolo, C. Samson, "Feedback control of a nonholonomic car-like". In Jean-Paul Laumond, editor, *Robot Motion Planning and Control*, 4: 171-253. Berlin Heidelberg: Springer, 1998.

[16] A. Karray and M. Feki, "Tracking control of a mobile manipulator with fuzzy PD controller", *2015 World Congress on Information Technology and Computer Applications (WCITCA)*, Hammamet, pp. 1-5, 2015.

[17] S. Peng, W. Shi, "Adaptive Fuzzy Output Feedback Control of a Nonholonomic Wheeled Mobile Robot", in *IEEE Access*, vol. 6, pp. 43414-43424, 2018.

[18] M. Roozegar, M. J. Mahjoob, "Modelling and control of a non-holonomic pendulum-driven spherical robot moving on an inclined plane: simulation and experimental results", in *IET Control Theory & Applications*, vol. 11, no. 4, pp. 541-549, 24 2, 2017.

[19] D. Chwa, "Fuzzy Adaptive Tracking Control of Wheeled Mobile Robots With State-Dependent Kinematic and Dynamic Disturbances", in *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 3, pp. 587-593, 2012.

[20] Y.-C. Chang, B.-S. Chen, "Intelligent robust tracking controls for holonomic and nonholonomic mechanical systems using only position measurements", in *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 491-507, 2005.

[21] G. Zhu, A. Kaddouri, L. A. Dessaint, O. Akhrif. "A nonlinear state observer for the sensorless control of a permanent-magnet ac machine". *IEEE Transactions on Industrial Electronics*, 48(6):1098-1108, 2001.

Appendix

Stability analysis of the Luenberger observer

To analyze the convergence of the estimation of the proposed Luenberger observer, the well-known Lyapunov approach is adopted, [4], [21]. In the following part the stability analysis is proposed.

Proposition 1

Let us consider the continuous observer defined by (3.8), if the nonlinear functions present in this system are Lipschitz ones, then there exists a matrix

$$K = \begin{bmatrix} k1 \\ k2 \\ k3 \\ k4 \end{bmatrix}$$

such that

where the corresponding estimation error dynamics are given by:

$$\dot{e}(t) = (A - KC)e(t) + \Delta f(t), \tag{A1}$$

with

$$\hat{e}(t) = \begin{bmatrix} \hat{s} \\ \hat{d} \\ \hat{\theta}_p \\ \emptyset \end{bmatrix} - \begin{bmatrix} \hat{s} \\ \hat{d} \\ \hat{\theta}_p \\ \hat{\emptyset} \end{bmatrix}$$

and with $A_0 = A - KC$

where

$$\Delta f(t) = \begin{bmatrix} \frac{\cos \theta_p}{1-dc(s)} \\ \sin \theta_p \\ \frac{\tan \emptyset}{l} - \frac{c(s) \cos \theta_p}{1-dc(s)} \\ 0 \end{bmatrix} v_1 - \begin{bmatrix} \frac{\cos \hat{\theta}_p}{1-dc(s)} \\ \sin \hat{\theta}_p \\ \frac{\tan \emptyset}{l} - \frac{c(s) \cos \hat{\theta}_p}{1-dc(s)} \\ 0 \end{bmatrix} v_1,$$

in which term Δf states the uncertainty due to the not perfect cancellation between the dynamics of the system and the dynamics of the observer.

Proof 1: With (A,C) being an observable pair, matrix A_0 for a suitable choice of the observer gain K is a Hurwitz matrix. This means that there exist symmetric and positive matrices P_0 and Q_0 which satisfy the so called Lyapunov equation

$$A_0^T P_0 + P_0 A_0 = -Q_0. \tag{A2}$$

To show the asymptotic stability of (A1), this Lyapunov function is introduced:

$$V(e(t)) = \frac{e^T(t)P_0e(t)}{2}.$$

The time derivative is given by:

$$\dot{V}(e(t)) = \dot{e}^T(t)P_0e(t) + e^T(t)P_0\dot{e}(t).$$

From (A1) it follows

$$\dot{V}(e(t)) = (A_0e(t) + \Delta f)^T P_0e(t) + e^T(t)P_0(A_0e(t) + \Delta f)P_0.$$

This yields:

$$\dot{V}(e(t)) = (e^T(t)A_0^T + \Delta f^T)P_0e(t) + e^T(t)P_0(A_0e(t) + \Delta f)P_0$$

At the end considering (A2), it follows:

$$\dot{V}(e(t)) = -e^T(t)Q_0e(t) + \Delta f^T P_0e(t) + e^T(t)P_0\Delta f.$$

Thus Δf is a Lipschitz function, then there is a positive constant L such as

$$\|\Delta f(x_1, x_2)\| \leq L\|x_1, x_2\|.$$

If ρ_1 is the small eigenvalue of matrix Q_0 and ρ_2 largest eigenvalue of matrix P_0 , if the following conditions result satisfied:

$$\rho_1 \geq 2\rho_2,$$

then

$$\dot{V}(e(t)) \leq -(\rho_1 - 2\rho_2)e^2(t).$$

Once suitable matrices Q_0 and P_0 are chosen, we can also choose matrix K such that A_0 has negative real eigenvalues to guarantee (A2).

Robot Self-Detection System

Ivaylo Penev^{*1}, Milena Karova², Mariana Todorova¹, Danislav Zhelyazkov¹

¹Technical University of Varna, Department of Computer Science and Engineering, 9010 Varna, Bulgaria

²Technical University of Varna, Department of Automation, 9010 Varna, Bulgaria

ARTICLE INFO

Article history:

Received: 15 August, 2018

Accepted: 18 November, 2018

Online: 07 December, 2018

Keywords:

Robotics

Image recognition

2D coordinates

3D coordinates

Arduino

ABSTRACT

The paper presents design and implementation of a mobile robot, located in an accommodation. As opposed to other known solutions, the presented one is entirely based on standard, cheap and accessible devices and tools. An algorithm for transformation of the 2D coordinates of the robot into 3D coordinates is described. The design and implementation of the system are presented. Finally, experimental results with different devices are shown.

1. Introduction

This paper is an extension of work, originally presented in [1].

Robot orientation is a field of scientific and practical interest for many years. Different methods and approaches for different instances of the robot orientation problem are proposed and studied, e.g. [2-9]. Most often communication technologies for robot orientation and path planning are used, for example RFID [10-12].

On the basis of the literature review several main challenges, concerning robot orientation, arise.

How to recognize the robot, using images of the robot and the environment;

This is a problem of image analysis and computer vision areas. Different methods for image cutting and segmentation are known, e.g. [13, 14]. The algorithms for robot localization use lasers, sonar sensors and stereo vision systems, e.g. [15, 16]. Although these solutions show good results, it is not always possible to provide the robot with lasers, sonars or vision system.

How to convert the 2D into 3D coordinates;

There known algorithms for transformation of coordinates from 2D to 3D plane rely on laser range finders or vision systems, e.g. [17-19].

How to design the applications, concerning their usage and scalability.

Image processing of robot and environment usually require many resources of the robot device (CPU and RAM). Most of the known applications rely on groups of robots or mobile agents, e.g. [20-23]. Of course, it is not always possible to provide groups of robot devices for experiments.

As a final conclusion of the problems and the known solutions, described above, we could summarize, that there are not complete and working solutions of the robot orientation problem, based on ordinary devices with limited resources.

The current work presents a robot orientation system, using a cheap Arduino based robot, supplied with LED sensors, and an ordinary mobile device. The paper layout is concentrated on three main points:

- Robot recognition from the image, shot by the device's camera;
- Converting the 2D coordinates to 3D;
- Implementation of the system by wide spread hardware and programming tools.

The purpose of the presented approach is to provide methodology for applying the robot orientation problem, using standard, cheap, wide spread devices.

*Corresponding Author: Ivaylo Penev, Email: ivailo.penev@tu-varna.bg

2. Methodology

2.1. System overview

The “Selfie robot” consists of a robot and a mobile device with a camera. The camera takes a picture of the robot and sends it to the robot. The robot’s control unit computes its self-coordinates as well as the camera’s coordinates. Afterwards it builds a path for moving to the center of the frame.

The robot is supplied with LED sensors. The system is implemented by standard Arduino architecture and OpenCV library.

2.2. Robot recognition

The Robot recognition algorithm consists of several separate steps together solving the common problem.

1) Robot marking

Simple yet efficient approach should be using light-emitting diodes (LEDs). LED is an electronic diode, which converts electricity to light – this effect is called electroluminescence. When an appropriate voltage is applied to it an energy is being disposed in the form of photons. The color of the light is defined by the semiconductor’s energy gap.

Simplified LEDs are colorful light sources which makes them easy for recognition in environments where light is reduced. Marking the robot with such diodes could make it look unique regardless of the daylight. The solution should work regardless of the environment’s brightness.

2) Camera properties

The previous pictures are taken by a camera, which properties are set to default – most commonly to auto-adjustment values, which ensures better quality of the picture. The current project does not depend on high end quality. All that is needed is a recognition of the robot.

The following settings deal with light perception:

a) ISO

This parameter measures the the camera sensitivity regarding the light.

b) Shutter speed

The shutter speed measures the time period necessary to fire the camera.

c) Aperture

Aperture (also known as *f-number*) presents a port into the lens, thorough which light moves into the body of the camera.

d) Exposure value

The value of this parameter is determined by the aperture and the shutter speed in such way ensuring that combinations giving the same exposure have the same exposure value (1).

$$exposure\ value = \log_2\left(\frac{aperture^2}{shutter_speed}\right) \quad (1)$$

Practically the exposure value indicates how much a picture is illuminated or occulted.

Given these definitions a simple approach comes up for reducing the unnecessary light from the taken pictures. They could be darkened by reducing the ISO and Exposure values.

Figure 1 and Figure 2 present photos, taken with ISO set to 50 and exposure value set to -2. An improvement can be easily noticed. By setting the ISO to a low value the camera sensor’s sensitivity was reduced. By decreasing the exposure value less light enters the camera. The effect of the reflected light from solid objects was highly filtered while the light from the LEDs has not changed at all.

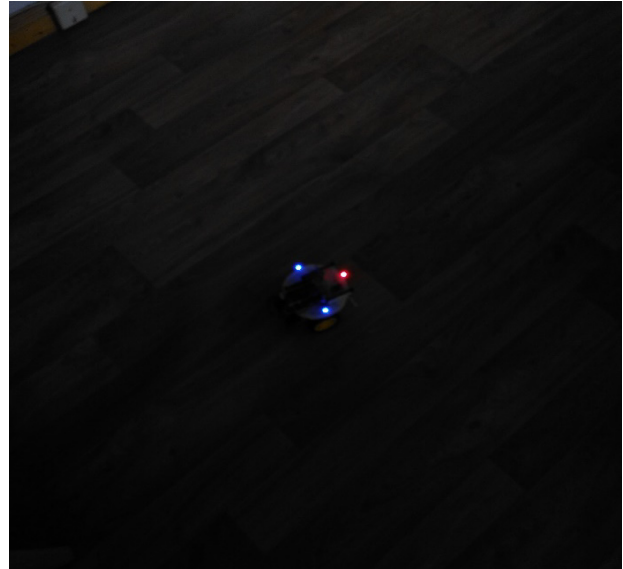


Figure 1. Dark room; reduced light sensitivity

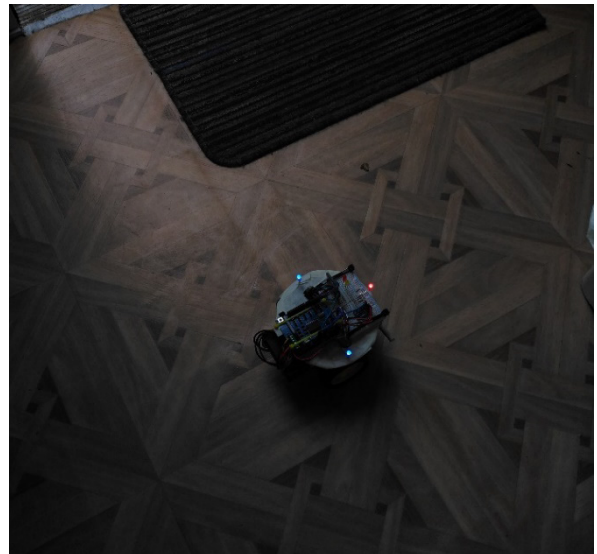


Figure 2. Light room; reduced light sensitivity

The LEDs’ light now can be easily extracted even if the robot is in a brighter environment. However they are seen as little blobs and the light reflected by highly specular objects could be still present as noise, thus presupposing algorithm confusion and failure.

3) Camera focus

Most cameras support different focus modes, which helps the user adjust the focus:

- Auto-focus: automatically finds the best focus range, where most of the objects are sharp;
- Infinity: the focus is set to the farthest range leaving close objects dull;
- Macro: the focus is set to the closest range making nearby objects look sharp;
- Manual: leaves focus range to be adjusted manually by the user.

Figure 3 is taken with ISO 50, EV -2 and Macro focus in a bright environment. It appears, that reducing the camera focus not only blurs / removes noise from strong reflections, but also improves the LEDs light enlarging and saturating their blobs.



Figure 3. Lighter room; defocused with reduced light sensitivity

4) Image analysis

So marking the robot with LEDs and adjusting the camera settings ensures more or less the same picture format with clearly expressed form of the robot regardless of the environment's illumination. The taken pictures are analyzed with a computer vision algorithm. The taken approach relies on the following two color ranges:

a) RGB

Red-green-blue color model is an additive color model consisting of red, green and blue lights which combined give a wide range of different colors. This model is the base of the color space and is considered as the most identical and the easiest to understand by the human eye. Identical values of the three lights give a shade of the gray color. Keeping up the lights (at least one) gives a bright color while if all three of them are low will result in a dark color, close to the black one. In electronic devices a color in that space is represented by 3 Bytes – a Byte for the value of each light. Could be graphically represented by a cube (Figure 4a).

b) HSV

Hue-saturation-value is another range based on the RGB model. It is considered to be more convenient for usage in computer graphics and image editing. Hue represents the color intensity, saturation – the color completeness and the value is often call the brightness of the color. This color space is one of the most common cylindrical representation of the RGB model. Like the RGB color space a color of this space can be also stored in 3 Bytes – one for each channel (Figure 4b).

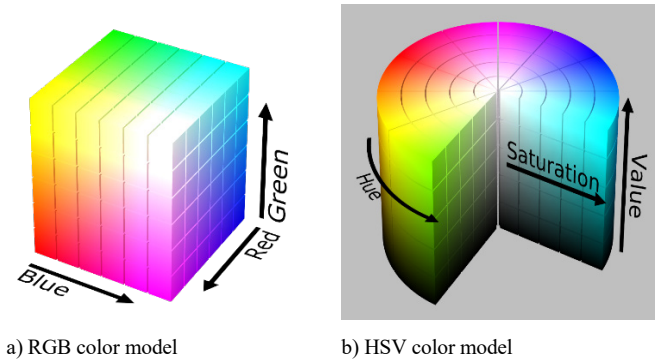


Figure 4. Color model

The goal is to extract the **blue** LEDs for example. First an image representing the blue channel from the RGB color space is obtained (Figure 5).



Figure 5. RGB Blue channel

The image is grayscale because it has only a single channel. The blue blobs can be easily noticed, because they are colored in high grey value. But so does the door's threshold. In fact every whiter object from the original image will have a high grey value in the blue channel image, because the white color in the RGB color space consists of high values in all of the three channels – red, green, blue.

The difference between the blue LEDs and the door's threshold is obvious – LEDs are more colorful in the input image. A good representation of the color intensities is the *saturation* channel from the HSV color variant of the input image (Figure 6).

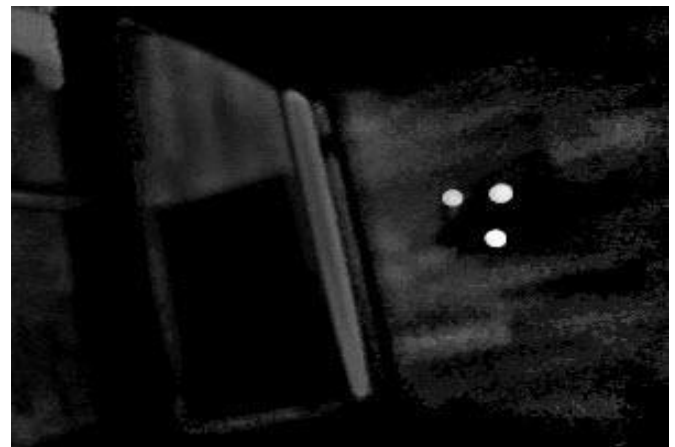


Figure 6. HSV Saturation channel

The difference is increased. The values of the LEDs are much higher than the one of the door's threshold. The saturation image could be a perfect mask for ignoring white objects in all of the three channels from the RGB color space.

Next a masking of the blue channel is performed with the saturation image – applying a logical AND between the images' binary data (Figure 7).

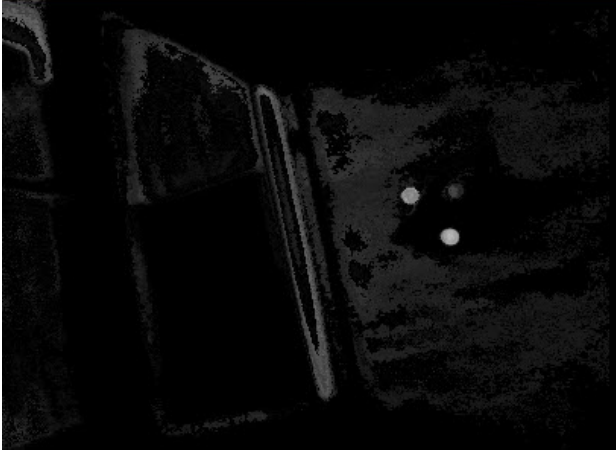


Figure 7. Merged blue and saturation (blue_sat) channels

c) Binary image

Binary images are simple images whose pixels can have only two possible states – 0/1 or black/white. In most cases white pixels are considered as part of an object or of the foreground of the image and black – as the background. In computer vision binary images are mostly used as a result of different processing operations such as segmentation, thresholding, and dithering.

d) Image segmentation

Image segmentation is a technique to divide a digital image into multiple zones. The purpose is to detect objects and boundaries within the image.

e) Thresholding

Thresholding is one of the simplest image segmentation methods. It basically separate a greyscale image into a background and foreground areas as a resulting binary image. The algorithm is simple – if a pixel value is higher than a given constant (called threshold), it is classified as a part of the foreground regions and its value is set to 1 in the binary image. Otherwise the pixel is considered as a background one and a 0 is assigned to its value. The most challenging task while using this method is determination of a correct threshold value. Often a technique called Otsu's method is used to solve this task. It calculates optimal threshold value by finding the minimal intra class variance between two class of points – one with lower values and one with higher values.

These image processing techniques are used by the LEDs recognition algorithm. After masking white regions in the blue image, a binarization is required. After the filtering a lot of dark regions have occurred. For certain LEDs' pixel values are highest. Histogram analysis must be done only in the upper half of the histogram.

This is the extracted upper half of the merged image. Almost every time the LEDs' pixel values are represented by the first

highest peak. Otsu's method is used over this part of the histogram. The calculated threshold is colored in green (Figure 8).

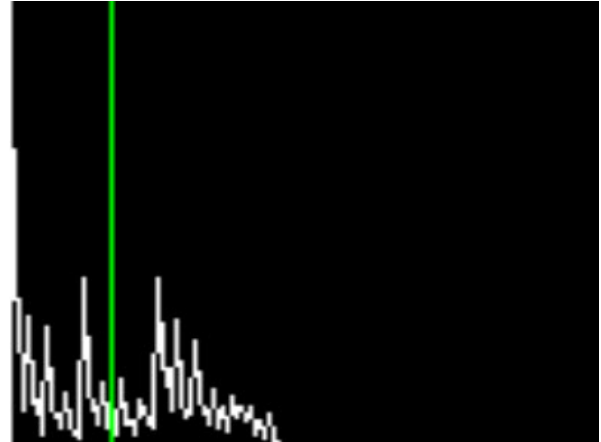


Figure 8. Blue_sat histogram

This binary image is a result of thresholding and consequential dilation. With the use of connected-component labeling technique regions could be easily extracted and their centroids could be computed by the average point formula. These centroids represent the {x, y} coordinates of the blue LEDs (Figure 9).

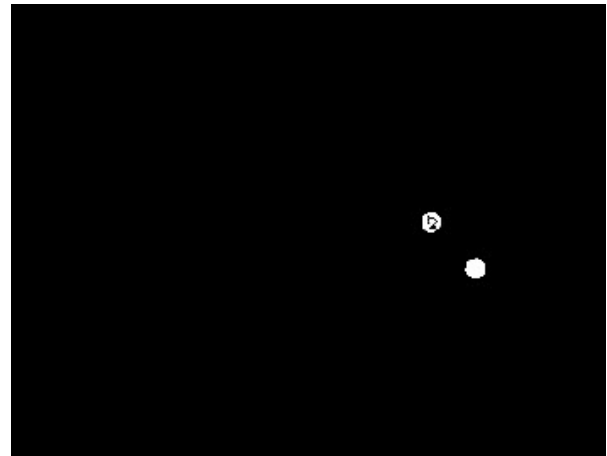


Figure 9. Blue_sat binary image

The same algorithm could be applied for the other two RGB channels. Since there are no green LEDs results will be shown only for the red channel (Figure 10, Figure 11).



Figure 10. RGB red channel



Figure 11. Merged red and saturation (red_sat) channels

At the end $\{x, y\}$ coordinates of the red and blue LEDs are present which makes the task of recognizing the robot completed.

2.3. Transformation of 2D coordinates to 3D

The camera of the device shoots the robot's sensors. The coordinates of the robot are calculated and sent to the robot using the Bluetooth interface. This process is presented at Figure 12.

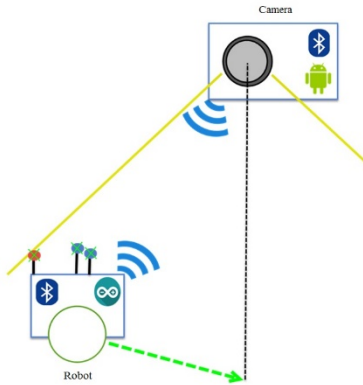


Figure 12. Sending coordinates to the robot

The robot moving to the goal consists of turning and moving forward. Turning uses the angle between the direction of the robot and the goal direction. The LEDs coordinates, extracted from the image by the recognition algorithm, put in a Cartesian coordinate system result in the following model (Figure 13):

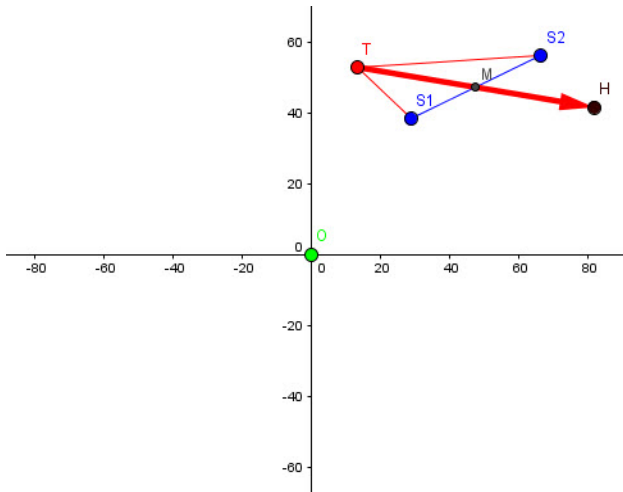


Figure 13. Extracted LEDs coordinates in Cartesian coordinate system

The model is an approximate example. For increasing the detail in the interest area future models will be limited only to the first quadrant.

$$\overline{TH}$$

The environment in which the robot and the camera exist is three dimensional. Two dimensional object representations in such 3D world exist in so called planes. The camera takes 2D images which belongs to a plane determined by the camera. The robot on the other hand is a 3D object. Its representation in the 2D image is a projection of itself in that plane. Robot's LEDs are attached in a way that keeps them on equal distance from the floor. It could be assumed that they are laying in a 2D plane parallel of the ground plane. What appears is that the extracted coordinates of the LEDs are their projections on the phone's plane, moreover the center of the image is a projection of the robot's target.

The following task arises: determine the robot's position and orientation against its target by the given projections' coordinates. After little mathematical analysis it appears that these coordinates are not enough for solving that given task. But that is not all. The robot knows the position of its LED sensors (Figure 14).

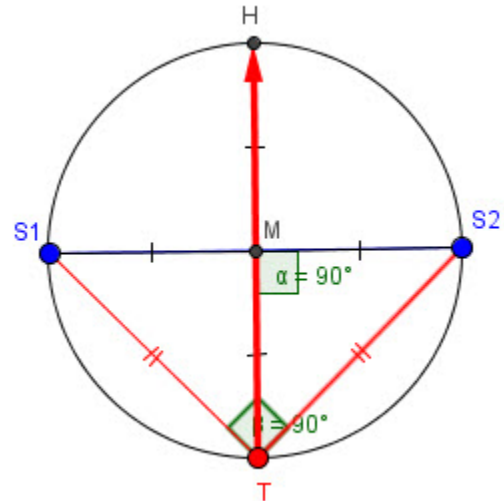


Figure 14. Robot representation in its plane

The symbols in the diagram have the following meaning:

- T – tail point, representing the red LED with the extracted coordinates,
- S1, S2 – side points, representing the blue LEDs with the extracted coordinates,
- – the center of the image/coordinate system, representing the goal's position,
- M – midpoint of the S1S2 segment, representing the center of the robot around which the robot should turn
- H – the T's mirror point against M, representing the head of the robot,
- \overline{TH} – a vector indicating robot's size and the direction it is looking at.

This model is representation of the robot in its plane. It knows where are its head, center and tail, what size is its diameter and that the LEDs form an isosceles right triangle. All that is left to be

found is the target’s coordinate against the robot in that plane – the O point. After that the value of the turn angle is defined by the points H, M, O. The distance to O is defined by (2):

$$distance = k \cdot robot_diameter \tag{2}$$

where $k = \frac{MO}{HT}$.

Since the robot knows its physical diameter and the distance depends on a proportional variable the robot’s plane doesn’t need to be the same size – just the same proportions.

2.4. Analytic Geometry for Transformation

The technique for accomplishing the task is simple and relies on basic laws from the Analytic geometry. Analytic geometry studies geometry using coordinate system – mostly the Cartesian one to deal with equations for planes, lines, points and shapes in both two and three dimensions. It defines, represents and operates with geometrical shapes using numerical information.

The current algorithm operates only with 2D coordinates {x,y} and the following shapes are being used:

Point

Represented by one pair of {x, y} coordinates which defines its position on the coordinate system.

Angle

Represented by single real value in radians or in degrees in the intervals respectively $[0, 2\pi)$ or $(-\pi, \pi]$ and $[0, 360)$ or $(-180, 180]$.

Line

Represents an unlimited straight ray. It is defined by (2):

$$y = K \cdot x + C \tag{2}$$

, where

K – the tangent of the angle between $Ox \rightarrow$ and the line; it has a constant value,

C – the offset between the intersection of $Ox \rightarrow$ and the line and the center point, it has a constant value.

Laws

The following laws are used by the algorithm:

Line from two points p1 and p2 (3)(4)

$$K = \frac{y_2 - y_1}{x_2 - x_1} \tag{3}$$

$$C = y_1 - K \cdot x_1 \tag{4}$$

Line parallel to line l1 passing through point p1 (5)(6)

$$K = K_1 \tag{5}$$

$$C = y_1 - K \cdot x_1 \tag{6}$$

Intersection point of two lines l1 and l2 (7)(8)

$$x = \frac{C_2 - C_1}{K_1 - K_2} \tag{7}$$

$$y = K_1 \cdot x + C_1 \tag{8}$$

Distance between two points p1 and p2 given by the Pythagorean theorem (9)

$$distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{9}$$

Angle α from three points p1, p2 and p3 (10)

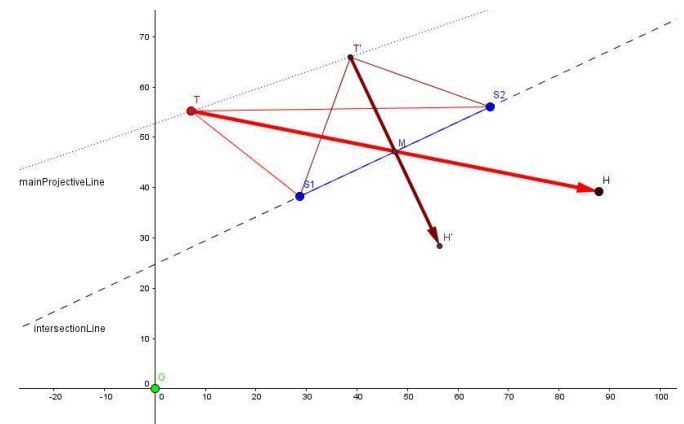
$$\alpha = atan2(y_1 - y_2, x_1 - x_2) - atan2(y_3 - y_2, x_3 - x_2) \tag{10}$$

where

$$atan2(y, x) = \begin{cases} \arctan\left(\frac{y}{x}\right), & \text{if } x > 0 \\ \frac{\pi}{2} - \arctan\left(\frac{x}{y}\right), & \text{if } y > 0 \\ -\frac{\pi}{2} - \arctan\left(\frac{x}{y}\right), & \text{if } y < 0 \\ \arctan\left(\frac{y}{x}\right) \pm \pi, & \text{if } x < 0 \\ \text{undefined}, & \text{if } x = 0 \text{ and } y = 0 \end{cases}$$

2.5. Transformation Algorithm Definition

There are two planes – the phone’s one and the robot’s one, several laws, extracted points and real proportions. The two planes have to intersect somewhere in the space. If not that means they are parallel – the camera is parallel to the ground and the searched target’s position is in fact the center of camera image. But in most cases it is not and the two planes intersect. If the actual intersection line is found, distance from the camera to the robot could be computed. But since the result could be proportional that operation is unnecessary and the intersection line could be any line from the two planes. Of course the shapes need to keep original robot proportions. It is assumed that the intersection line is the one that passes through the two blue points giving the following model as a result (Figure 15).



Intersected planes with an example projection line

The dark color has shapes in the robot’s plane while brighter ones – in the image plane; a shape’s projection has the same name with suffix ‘’.

T', H' – projections of T and H , showing how would T and H coordinates be if the image's plane was parallel to the robot's one.

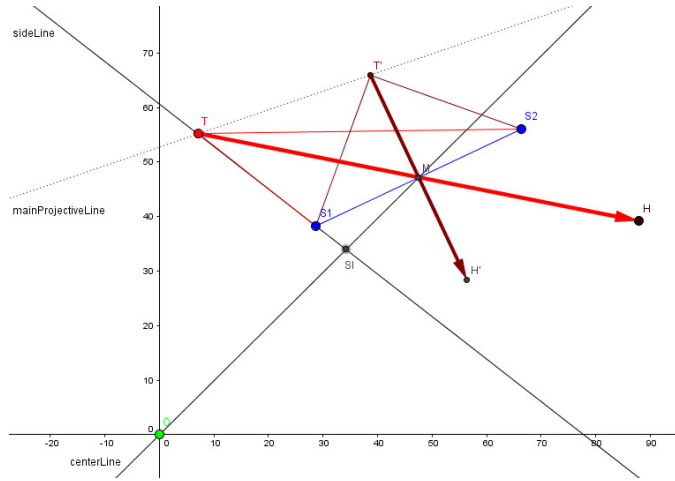
As it can be seen on Fig. 4 $SI, S2$ and T' also define an isosceles right triangle in 2D – the proportions are kept.

A line could be defined from either T and T' or H and H' and could be called “main projective line”. All newly defined projective lines have to be parallel to this main one.

Let's define the following shapes (Figure 16):

- Line from O and M called “center line”;
- Line from T and either called “side line” that can intersect the “center line”;

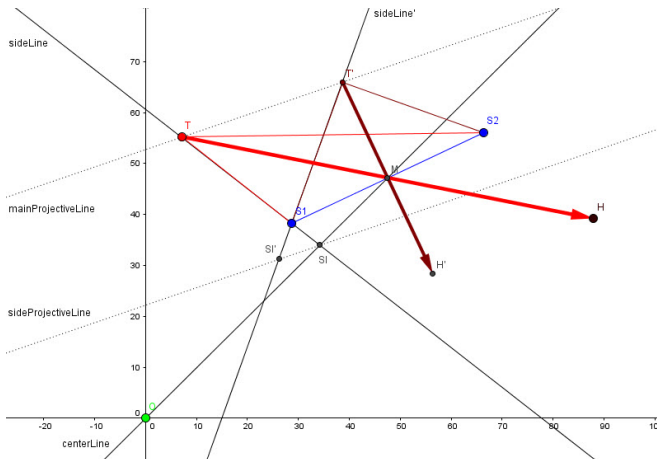
Intersection point SI from the center and side lines.



The “side line” is defined using SI which in this case is correct since the side line is not parallel to the center one. But there is a possibility that the line from T and SI may not intersect the center line. If such case occurs the side line should be defined using $S2$. For certain $T, SI, S2$ form a triangle so if SI is not appropriate for defining the side line, $S2$ will definitely be.

Next step consists of defining the following shapes (Figure 17):

- Side line's projection – line from T and SI (or $S2$ if SI was not relevant during the previous step);
- Projection point of SI on the side line's projection.

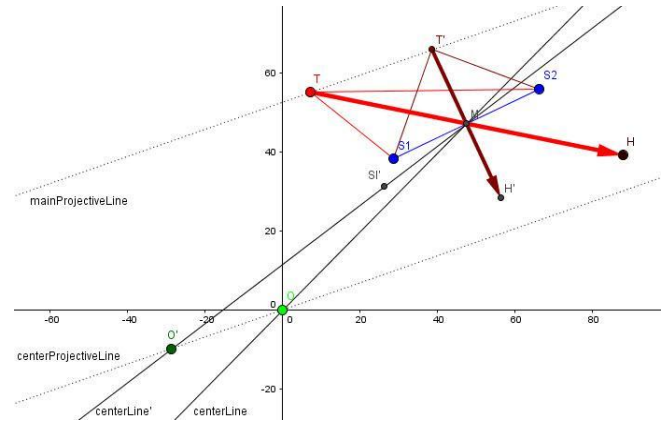


Projecting the side intersection line

Defining a projection of a point on a line consists of two steps. Firstly a projective line must be defined. That line passes through the projected point and is parallel to the main projective line. The projection point is considered the intersection point of the projective line and the line that it is supposed to lie on. In this case the projected point is SI , the projective line is *sideProjectiveLine* and the projection point is SI' that lies on *sideLine'*.

The final algorithm step consists of defining the following shapes (Figure 18):

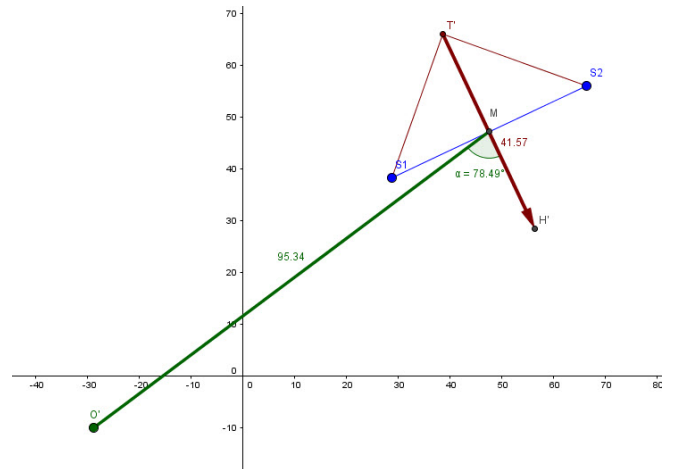
- Define a line from M and SI' ; since SI' is a projection of SI along with the fact that SI and M lay on the center line, it could be stated that this new line is the projection of the center line;
- Define the projection of O on the center line's projection.



O' is the projection of O which is defined the same way SI' was. But this time the projective line is *centerProjectiveLine* and the second line for intersection is *centerLine'*.

O' now represents the location of the target in the robot's plane.

The result projection is shown on Figure 19.



Given that point the angle $H'MO'$ could be easily computed using the declared law.

The lengths of the segments MO' and $H'T'$ are also known which enables the evaluation of (1):

$$k = \frac{MO'}{H'T'} = \frac{95.34}{41.57}$$

3. System requirements and specification

3.1. Mobile robot

The robot used in this project is based on the Arduino platform. It is an open-source microcontroller-based kit for building digital devices and interactive objects that could interact with the environment with sensors and actuators. Arduino board provides set of analog and digital I/O pins which enable connectivity with external physical devices. Programs for Arduino are written in either C, C++ or Processing.

3.2. Arduino board

Robot’s controller board is an Arduino UNO R3, which is a basic microcontroller board, suitable for simple startup projects. It supports Universal Serial Bus (USB) with the help of which a serial communication could be established with PC and other devices and could also supply power to the board. Usually programs are built and compiled on a PC and further transferred to the board using the USB. It has also a 2.1mm center-positive power jack for external power supplies.

The board, used for assembling the robot, has the following parameters:

- Microcontroller: ATmega328
- CPU: 8-bit AVR
- Clock: 16MHz
- Memory of 23KB flash, 2KB SRAM and 1KB EEPROM
- 14 I/O pins
- 6 Analog Input pins
- Physical parameters: 68.6x53.4mm , 25g weight

3.3. Robot body

For basic movements a simple 2 Wheel Drive (2WD) structure is used. This is a DIY 2WD double leveled plastic chassis. It consists of 2 wheels with separate gear motors, full-degree rotating caster wheel as a third strong point and two decks. The wheels are positioned almost in the center of the structure. Their rotation in opposite directions will result in the whole structure spinning over its center. The two decks are ideal for separating the mechanics from the electronics.

3.4. Motor shield

The motor shield of the robot provides simple motor control. It provides a control of a motor’s speed, direction and braking along with sensing of the absorbed current. This board is stacked over the Arduino board.

3.5. Bluetooth module

The robot is supplied with additional Bluetooth communication module.

3.6. Magnetometer

HMC5883L is a Triple-axis Magnetometer (Compass) board. With the help of this sensor the robot can determine the direction it is facing. Furthermore using it the turning accuracy could be increased.

3.7. Camera

The project definition assumes that the robot establishes a communication with a remote device with a camera. It has one communication module – Bluetooth, therefore the remote device is required to have not only a camera but also Bluetooth communication technology.arduino board.

4. System Design

4.1. General Design

Figure 20 gives an overview of the system’s design.

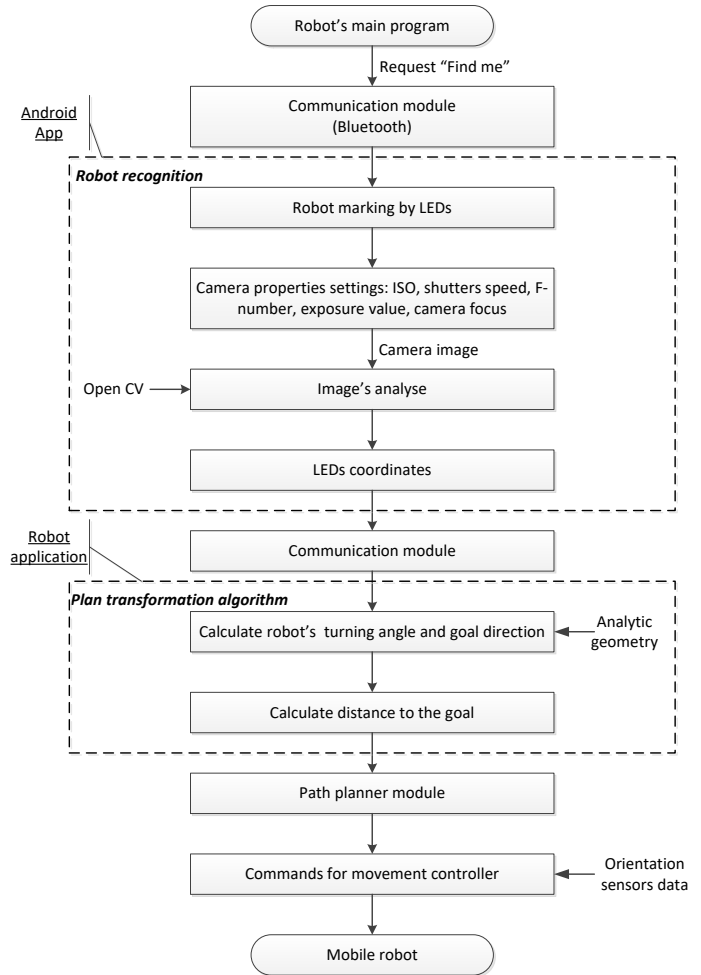


Figure 20. Flowchart of the system’s general design

4.2. Robot Software Design

The software for the robot is build based on the Arduino API keeping identical design to the general one.

The design is similar to the robot side of the general diagram. The difference is that the software is built over Arduino core API while the other is more abstract.

Abstract2WDController and Motor classes are meant to be an API to the Arduino Motor shield. They encapsulate pins reservation along with basic Arduino methods for setting different digital and analog levels to the pins.

- Motor – class, representing single motor channel of the Motor shield. It adjust its pins modes and levels using Arduino basic

subroutines. Represents an API to the Motor module from the general diagram.

- Abstract2WDController – abstract class, encapsulating the declaration of two motor channels with the specific pin mapping for the Arduino Motor Shield. It also defines a basic interruptible motor control leaving the motors' parameters to be adjusted by the class' successors.
- Compass – an interface declaring a minimum behavior that one true compass object should implement.
- AngularController2 – an extension of the 2WDController providing forward movement and turning operations. In order to do accurate turns it uses a Compass object to determine its orientation. It has a predecessor sharing the same name and interface but not so successful. More information about them in the Implementation chapter.

The two controller classes could be considered as a representation of the Movement Controller module from the general diagram.

- ExSoftwareSerial – extends the Arduino's SoftwareSerial library providing methods for reading not-string based data types from the upcoming stream. These methods support error handling and time-out periods. This class represents the API to the robot's Communication module from the general diagram.
- TargetFinder – class implementing the main Plane transformation algorithm. It takes the {x,y} coordinates of the tail and the two side points and determines what turn has to be made in order to make the robot face the target and how far it is. This class does relates to the Path Planner module from the general diagram.
- Main – class representing the Main program. It binds the three features of the robot (communication, brain and motion control) together to complete the main goal of the project – staying in the center of the camera frame. This class delegates communication with the remote camera and controls the data flow from it through the TargetFinder to the Movement Controller.

4.3. Camera Software Design

The software written for the remote camera is built based on the Android API but keeps its core identical to the general idea. Keeping the fact that the software is more or less a mobile application it has to have highly intractable user interface in order to monitor the system. The design extends the general one providing additional monitoring support using Android UI tools.

4.4. Application Overall Design

This design shows how application's modules are bound together.

- Bluetooth package – this package contains all defined Bluetooth classes. These classes are façades of the Android's Bluetooth API and provides additional exception handling techniques and device discovery tools.
- Camera package – contains Camera classes extending the functionality of the Android's Camera API. It implements different tools for serving the application's specification.

Camera preview class is present for handling each frame captured by the camera sensor. Additional Image processor is defined for analyzing images and a Custom camera class providing programmable interface to the project specific camera parameters for adjustment.

- Robot package – classes for handling robot requests. They implement features as robot communication managing, robot data translation and robot recognition.
- Graphical user interface (GUI) package – classes observing the core objects and updating the user interface when events occur
- MainActivity – manager class; creates all class hierarchy binding different modules together. It is responsible for resource acquiring and releasing such as Camera and Bluetooth hardware modules reserving, communicator and image processing threads handling on application starting and closing. It also appears as a controller to the main UI window granting classes from the GUI package an access to the UI objects.

4.5. Core Design

The design, presenting the core of the application, does not differ much from the general one.

- Bluetooth classes – façades to the Android's Bluetooth API; represent the interface to the device's Bluetooth (the Communication module of the mobile device).
- BluetoothModule – manager class; creates BluetoothConnection objects, deals with the Bluetooth hardware's settings and takes the responsibility of acquiring and releasing it along with closing all opened connections.
- BluetoothConnection – establishes bidirectional RF connection to a single remote device and is responsible for all data that flows in and out the device.
- RobotCommunicator – manager class; as the name suggests it completely implements the functions of the RobotCommunicator module from the general design. Additionally it uses the BluetoothModule to connect to the robot and a separate decoder called RobotTranslator to get a RobotRequestEvaluator object for handling the current request.
- RobotTranslator – factory for RobotRequestEvaluator objects; it takes the received request from the communicator and decides what evaluator to return.
- RobotRequestEvaluator – interface that unifies the type of all evaluators' responses – a byte array as specified by the BluetoothConnection's method "sendData".
- CustomCamera – façade based on the Android's Camera API providing interface to the Camera module and access to the following adjustable camera parameters (focus mode, ISO, Auto-exposure lock and exposure compensation)
- CustomCameraView – camera preview; implements frame handling that submits each new camera frame to the CameraImageProcessor and to a UI preview surface.
- CameraImageProcessor – singleton; fully implements the Image Analysis algorithm using external image processing

native library – OpenCV; the result represents all extracted red and blue regions; represents the ImageProcessor module from the general design.

- RobotLEDImageRecognizer – singleton; analyzes the extracted regions from the CameraImageProcessor trying to recognize the robot. If successful, it converts the appropriate coordinates to a binary data. If the robot is not successfully detected from the current regions, modifications of the camera’s exposure are made and next result is awaited. Combinations of all available low exposure parameters are being tried.

4.6. User Interface Design

21). It consists of objects called *Observers* that are listening for activity or changes in other objects called *Subjects* or *Observables*. Observer register to a Subject which hold a list with subscribed Observers. Each time a Subject do some activity or change state it notifies all its registered listeners providing them with the required information. Using this pattern constant looping (listening) in the listener classes is avoided.

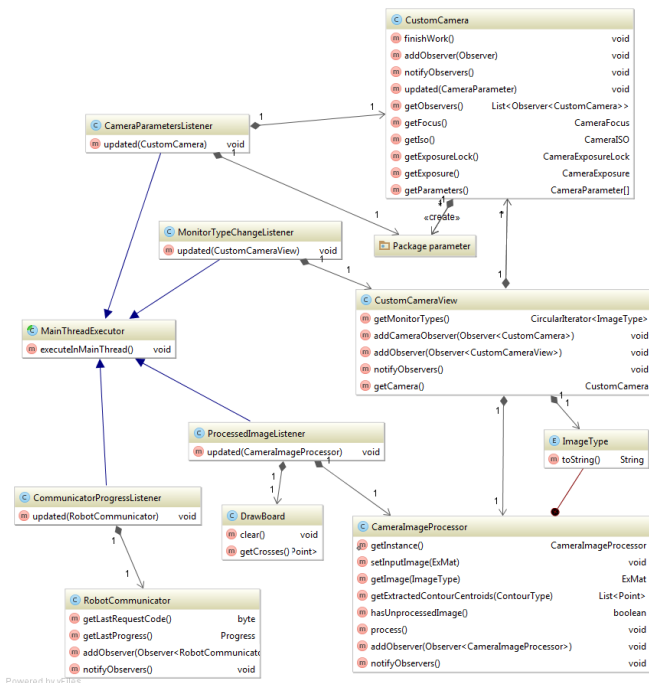


Figure 21. UML class diagram of the Android application's UI layer along with the observed core classes

- *MainThreadExecutor* – it is an abstract class providing support for executing subroutines in the main thread. Android’s specification states that all interactions with user interface objects have to be done by the main thread also called *UI thread*. Also it highly recommends long-running tasks to be not executed by the main thread because it may result in UI freezing or lagging. *MainThreadExecutor* is ideal for classes that mainly interact with the UI but are being accessed by objects running in separate threads – like the *RobotCommunicator* or the *CameraImageProcessor*.
- *CommunicationProgressListener* – this class observes the *RobotCommunicator*. Each time a communication activity is being done the communicator notifies its listeners. Interaction

between the communicator and its observers is being done using *Progress* typed objects. The *CommunicationProgressListener* updates an UI object each time new *Progress* object becomes present.

- *ProcessedImageListener* – an observer of the *CameraImageProcessor*. When an image has been processed the processor updates its observers. The listener retrieves the extracted regions from the picture and displays them to the screen using a *DrawBoard*. *DrawBoard* is an which represents an Android UI Surface that is positioned above any else UI object in the main UI window. It implements functionality for drawing crosses.
- *MonitorTypeChangeListener* - observes the *CustomCameraView*. By definition this custom view class handles all preview images from the Camera and sends them to both the Image processor and an UI preview class. Some additional monitoring was implemented giving the opportunity of displaying not only the original image but also intermediate processed image from the *Image analysis algorithm*. The *Image processor* provides access to such images via *ImageType* object. Clicking the view changes the type of the send image to the UI preview. The *MonitorTypeChangeListener* is notified in case of such events and reflects the displayed image type’s string representation to a UI textbox.

Figure 22 demonstrates the common work of the application.

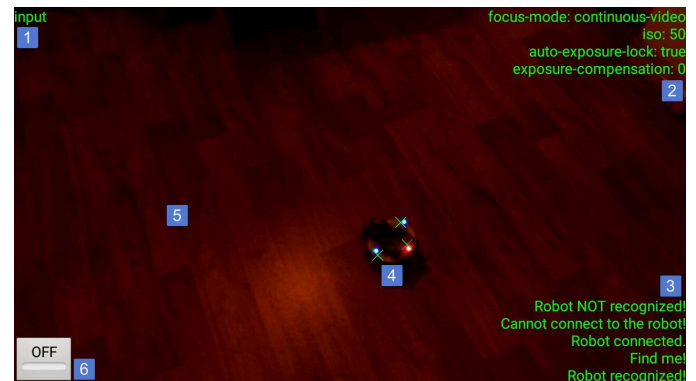


Figure 22. Screenshot of the Android application with marked UI objects

The number markers has the following meaning:

- This is a text label that is being handled by the *MonitorTypeChangeListener*. It displays the preview image’s type.
- A multiline not-editable text area. It is modified by the *CameraParametersListener* each time a camera parameter is changed. Presents the current camera’s settings.
- Another multiline area that keeps track of the *RobotCommunicator*’s progress. It is being handled by the *CommunicationProgressListener*.

- Crosses drawn by invoking DrawBoard methods from the ProcessedImageListener.
- Touching the screen will change the preview image. The new image's type will be presented in the Label 1.
- It is a switch button that forces the ImageProcessor work even when robot communication is not present. Used for high-end monitoring but consumes more battery power.

5. Testing and Evaluation

5.1. Android Application

1) Robot recognition algorithm

The application was tested on three devices with quite different hardware specification. The results are summarized in Table 1.

Table 1. Camera application test results

Device model	Android	CPU	Camera parameters	Remarks
Asus ZenFone 2 Laser	API 21 v5.0 Lollipop	Qualcomm Octa-core	Focus mode ISO Exposure	Smooth performance; Fast image analysis; Robot cannot be recognized while on bright surface
LG Optimus G	API 19 v4.4 KitKat	Qualcomm Quad-core	ISO Exposure	Smooth performance; Fast image analysis; Robot cannot be recognized while on bright surface or if some noise is present.
Sony Live	API 14 v4.0 Ice Cream	Qualcomm Single-core	Exposure	Smooth performance; Fast enough image analysis; Robot cannot be recognized at all.

The following conclusions are summarized:

- The presence of several background threads doesn't badly affect overall performance on a Qualcomm CPU with less cores;
- Automatic or not-present ISO setting is not recommended;
- Support of a close focus range could ignore noise;
- The recognition depends on the supported editable camera settings. An advanced intelligent robot recognition algorithm is needed to remove this dependency.

2) Application Stability

After a lot of testing and debugging it could be stated that *OpenCV's* matrix object *Mat* is not being automatically released

on garbage collection as defined in the library's documentation. Therefore it should be programmatically released before dereferenced. Otherwise a memory leak occurs which could cause system's instability.

5.2. Robot behaviour

1) Plane transformation algorithm

The plane transformation algorithm is tested with different coordinates of the LED sensors, i.e. with three input data into a Cartesian coordinate system with range [-5; -5] ÷ [5; 5].

For resolution 2 of the value of coordinates 6 values per a coordinate are possible: -5, -3, -1, 1, 3, 5. Three input points have two coordinates with $6^6 = 46656$ test cases (Table 2).

Table 2. Results from test 1

Experiments	Number of tests	Percent	Conclusion
Total	46656	100%	
Passed	36524	78.28%	
Exceptions	10132	21.72%	
• Handled	6048	12.96%	The three points lay on one line or two of them overlap.
• Not handled	4048	8.75%	To be analyzed in future.

The tests from resolution 1 are presented in Table 3:

- each coordinate has 11 possible values in range [-5; 5];
- test cases: $11^6 = 1\ 771\ 561$ possible combinations.

Table 3. Results From Test 2

Experiments	Number of tests	Percent	Conclusion
Total	1 771 561	100%	Conclusion
Passed	1 613 506	91.08%	
Exceptions	158 055	8.92%	
• Handled	84 473	4.77%	The three points lay on one line or two of them overlap.
• Not handled	73 582	4.15%	Further analysis

More combinations lead to less chance for an exception. Even a camera with low quality can take images with resolution of IMP. This means, that the number of combinations is great and the occurrence of exceptions is practically not possible.

2) Turning accuracy

After continuous manual testing an anomaly has been discovered. The magnetometer (compass) sensor does not always gives adequate data and thus resulting in Movement controller's confusion effecting the turning accuracy. The sensor has been tested in separate project with all magnetic components (motors, battery) detached from the robot but the problem still occurs. The issue could be caused by hardware defect but another module is not available at this state.

6. Conclusions and Future Work

The results from the experimental tests prove that the system is stable and effective with different devices and resolutions of the

camera. This means, that a mobile robot, equipped with a usual camera, could be practically exploited for solving tasks as moving through dangerous spaces, searching for victims of disaster events, etc.

The future work will be directed to implementation of intelligent software controller, using neural network for detection and recognition of the robot image. More advanced orientation sensors (accelerometer, gyroscope) will be integrated and robot's physical parameters (wheel diameter, motor RPM, robot size) will be used by the algorithm.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] M. Karova, I. Penev, M. Todorova, and D. Zhelyazkov, "Plane Transformation Algorithm for a Robot Self-Detection", Proceedings of Computing Conference 2017, 18-20 July 2017, London, UK, ISBN (IEEE XPLORE): 978-1-5090-5443-5, ISBN (USB): 978-1-5090-5442-8, IEEE, 2017.
- [2] C. Connolly, J. Burns, and R. Weiss, "Path planning using Laplace's Equation", IEEE Int. Conf. on Robotics and Automation, pp. 2101-2106, 1990.
- [3] D. Lima, C. Tinoco, J. Viedman, and G. Oliveira, "Coordination, Synchronization and Localization Investigations in a Parallel Intelligent Robot Cellular Automata Model that Performs Foraging Task", Proceedings of the 9th International Conference on Agents and Artificial Intelligence – Vol. 2: ICAART, pp. 355-363, 2017.
- [4] J. Barraquand, and J. C. Latombe, "Robot motion planning: A distributed representation approach", Int. J. of Robotics Research, Vol. 10, pp. 628-649, 1991.
- [5] O. Cliff, R. Fitch, S. Sukkarieh, D. Saunders, and R. Heinsohn, "Online Localization of Radio-Tagged Wildlife with an Autonomous Aerial Robot System", Proceedings of Robotics: Science and Systems, 2015.
- [6] S. Saeedi, M. Trentini, M. Seto, and H. Li, "Multiple-Robot Simultaneous Localization and Mapping: A Review", Journal of Field Robotics 33.1, pp. 3-46, 2016.
- [7] T. Kuno, H. Sugiura, and N. Matoba, "A new automatic exposure system for digital still cameras", Consumer Electronics, IEEE Transactions on 44.1, pp.192-199, 1998.
- [8] T. Sebastian, D. Fox, W. Burgard, and F. Delaert, "Robust Monte Carlo localization for mobile robots", Artificial Intelligence, Vol. 128, Iss. 1-2, pp. 99-141, 2001.
- [9] W. Yunfeng, and G. S. Chirikjian, "A new potential field method for robot path planning", IEEE, DOI: 10.1109/ROBOT.2000.844727, San Francisco, CA, USA, 2000.
- [10] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, and M. Philipose, "Mapping and localization with RFID technology", *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 1, pp. 1015-1020, 2004.
- [11] J. Flores, S. Srikant, B. Sareen, and A. Vagga, "Performance of RFID tags in near and far field", *Proc. IEEE Int. Conf. Pers. Wireless Commun.*, pp. 353-357, 2005.
- [12] D. Lima, and G. de Oliveira, "A cellular automata ant memory model of foraging in a swarm of robots.", *Applied Mathematical Modelling* 47, pp. 551-572, 2017.
- [13] B. Siciliano, and O. Khatib, *Handbook of Robotics*, Ed. 2, ISBN: 978-3-319-32550-7, Springer-Verlag, Berlin, Heidelberg, 2016.
- [14] J. Shi, and J. Malik, "Normalized cuts and image segmentation", *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 22.8, pp. 888-905, 2000.
- [15] R. Stengel, "Robot Arm Transformations, Path Planning, and Trajectories", *Robotics and Intelligent Systems*, MAE 345, Princeton University, 2015.
- [16] S. Se, D. Lowe, and Jim Little, "Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks", *The International Journal of Robotics Research*, Vol 21, Iss. 8, pp. 735 – 758, 2002.
- [17] T. Kuno, H. Sugiura, and N. Matoba, "A new automatic exposure system for digital still cameras", *Consumer Electronics*, IEEE Transactions on 44.1, pp.192-199, 1998.
- [18] D. Hahnel, W. Burgard, and S. Thrun, "Learning compact 3D models of indoor and outdoor environments with a mobile robot", *Robotics and Autonomous Systems*, Vol. 44, Iss. 1, 2003.
- [19] L. Lee, R Romano, and G Stein, "Introduction to the special section on video surveillance", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8, pp. 740-745, 2000.
- [20] M. Beetz, "Plan-Based Control of Robotic Agents: Improving the Capabilities of Autonomous Robots", ISSN-0302-9743, Springer-Verlag, Berlin, Heidelberg, New York, 2002.
- [21] C. Richter, S. Jentzsch, R. Hostettler, J. Garrido, E. Ros, A. Knoll, F. Rohrbein, P. van der Smagt, and J. Conradt, "Musculoskeletal robots: scalability in neural control", *IEEE Robotics & Automation Magazine*, Vol. 23, Iss. 4, pp. 128-137, 2016.
- [22] D. Lima, and G. de Oliveira, "A cellular automata ant memory model of foraging in a swarm of robots.", *Applied Mathematical Modelling* 47, pp. 551-572, 2017.
- [23] M. Alkilabi, A. Narayan, and E. Tuci, "Cooperative object transport with a swarm of e-puck robots: robustness and scalability of evolved collective strategies", *Swarm Intelligence*, Vol. 11, Iss. 3-4, pp. 185-209, 2017.

Analysis of Garri Frying Machine Manufacturing in Nigeria: Design Innovation

Rufus Ogbuka Chime^{*1}, Odo Fidelis O²

¹*Mechanical Engineering, Institute of Management and Technology, 400001 Enugu Nigerian*

²*Food Tech, Institute of Management and Technology, 400001. Enugu, Nigeria*

ARTICLE INFO

Article history:

Received: 14 August, 2018

Accepted: 20 November, 2018

Online: 07 December, 2018

Keywords:

Design

Garri frying

Sustainability analysis

Environmental

Agriculture

Innovation and Manufacturing

ABSTRACT

Production of garri (edible processed and granulated form of cassava) and sustaining the production process has become so laborious, time consuming, and predisposes one to some form of danger, especially as it concerns the hot fire that one is disposed to during the process. In west African tradition, where this garri serves as one of the major staple food in the sense that it can be taken soaked in water, or served with soup in the form of “eba”, is produced mainly by women. They do this by frying it over a wood fire in a shallow-ware cast iron pans (Agbada). For even circulation of heat that comes from that wood fire via the iron pans, they use spatula – like paddles of wood or calabash on the hot surface, where placed the granules for processing, to vigorously turn to avoid caking. This process brings about great discomfort as the processor (the woman) will have to sit sideways, just close to the frying fire from the wood. It is against this backdrop that this work seeks to establish a mechanized method of garri processing which is design innovation and sustainability analysis of garri frying machine. This mechanized method appears in form of machine with stainless steel drum, rotary conveyor and paddles fixed along the conveyor to a slower rotation in the same axis of the drum. This project also presents how important this design innovation is especially in the areas of economic growth, welfare and job creation.

1. Introduction

Garri which is creamy white grainy flour made from fermentation and gelatinization of which has faintly flavor, sour taste as a result of fermentation of new of cassava tubers. It is consumed in Brazil and in most African countries especially in Nigeria where most of its preparations are done using local processing techniques. Researches into the mechanization method involved into the unit operations as peeling and washing of tubers, grating, dewatering, fermentation, sieving, frying, and cooling in garri production were done in the past years [1] Some machines were designed to help in the large scale [2-5]. Frying operation which is the unit operation that approves the value of the last product, in garri production. It was difficult to mechanize this operation perfectly and the processes was not well understood by many designers and manufacturers [6]. It was thought of by some people that garri frying means dehydration/roasting. The removal of water molecules or other liquids from solid to lessen the content or residual liquid to an acceptable value is otherwise known as

drying. The removal of some amount of water or other liquid from the solid material to reduce the substances or residual liquid to an acceptable low value is the means of solid drying [7]. The final step in a series of processing and handling operations and the goods from a dryer is often ready for final packaging. Drying is a relative term and methods of reduction moisture content from an initial value to some acceptable final value. Cassava has approximately 60% water content while final garri is between 10-15% [8] Garri frying (called gratification), though a dehydrating process, is not a straightforward drying process. The moisture value of the cassava mash that was removed during sieving should be from 51 -66% and should be 13% when the frying operation finishes. The product is allowed to dry from the bearest minimum of about 13% during the final operation known as frying operation. Recently, frying and drying machines are usually manufactured using stainless steel drum with rotary conveyor and paddles fixed along the conveyor to reduce rotation in the same axis of the drum. As mentioned earlier, traditionally, garri is fried by women in shallow earthenware of cast-iron pans (agbada) over a wood fire.

^{*}Corresponding Author: Rufus Ogbuka Chime, Email: rumechservices@yahoo.com

The operator sits sideways by the fireplace while frying, and this brings discomfort due to heat and the sitting posture [9]

Many years ago, garri is fried by women in shallow earthen-ware cast-iron pans (agbada, Nigerian Ibo) over a wood fire. Female use spatula-like paddles of wood or calabash half sections to press the sieved mash against the hot surface of the frying pan, while continuously turning it to avoid cake production. The operator who sits beside the fire during frying process faces challenges base on the heat produced from the wood fire, inspired researchers.

In Nigeria for instance, quality mechanized garri processing plants are few, and as a result engineers and manufactures in Nigeria are seeking for improvement in already existing models. Model by definition unattainable in a given time/space but appends lastly. It is this endless pursuit that forms a justifiable process. Good system environments are necessary to the survival of humans and other organisms. Ways of minimizing human impact are environmentally-friendly chemical engineering, environmental resources management and environmental protection.

Report is obtained from green chemistry, earth science, environmental science, conservation biology and Ecological. Economics studies the subjects of academic research that aim to address human economies and technological ecosystems [10].

Sustainability can be quite a malleable characterize.

Many people understand its purpose intuitively because it's hard to really pin down, it covers so many domains. The Brundtland Commission, made the best-known definitions of Environment and Development in The World Commission: Sustainable development is one that produce the needs of the people without bargaining the talent of future generations meeting their own needs sustainability manifests itself in industries at a variety of levels, including.

- Strategy –Some industries decide what to make as a result of sustainable business ideals. Stony field Farms has made social and environmental responsibility a key part of its business strategy since it began. Supply chain & value webs. Walmart requires its suppliers to evaluate and disclose the full impact of their goods. There continues to be demand attention to so-called company.
- Ecology, which evaluates the material and energy flows within whole industrial systems, often extending far beyond.
- Operations – The result of producing goods increasingly show ecological controls for floor casing industry Interface, the real justifiable business accomplishment was opening of the social and ecological effects of their processes because of problems, industries have inaugurated Environmental Management Systems (EMS), that have operationalized the tracking, documentation, and reporting of ecological certification, of ecological influences of the institution, There is a detailed of ISO standard (ISO 14001:2004) governing EMS. 5.

The level of sustainability concerns which is basically on product is where the majority of this guide will concentrate on,

nevertheless it is obliging to hold onto the fact that sustainability is not the sphere of just single portion of the business. Truthfully, an actual produce of sustainability should be able to merely be existent surrounded by the framework of an ample comprehensive scheme that gives a backing to its affirmative influence societies, globe, as well as proceeds [11]. The procedure involved in making simpler a topic which is composite or constituent to minor portions acceptable to achieve an enhanced appreciation for that [12]. Innovation is well-defined merely for instance an idea that is recent, scheme, or process. Conversely, innovation is over and over again similarly observed as the use of improved way out that encounter novel necessities, tacit desires, otherwise prevailing market desires. The accomplishment comes via extra operational produces, procedures, facilities, know-hows, or business mockups that are readily obtainable towards the markets, governments as well as society. The term "innovation" is well-defined as unique something as well as additional operative also, consequently, first-hand, that "enters into" the market place or the general public. It is interrelated to, however, not the alike as, invention. [13]

2. Literature Reviews

In the past, tools and implements were developed to perform tasks that were made to complement human physical strength. The discovery of the wheel is probably the most revolutionizing optimization tool made by humans. In today's complex business and industrial environments, the solution of operational problems cannot be achieved by technological advances alone.

The multitude of options available for implementing an operational plan has mandated the development of systematic procedures for selecting the options that best benefit of the industries as a whole [13].

There are few automated garri processing plants in the Nigerian market which have found to be performing well as regards to the quality of garri. As the outcome, some new products (developments have been made by Nigerian researchers to solve the problems associated with the models already in the market.

2.1. Traditional method

Throughout garri frying processes, the moisture content concerted and most of the small lumps established are broken down by continuous pressing and agitation, heat is then increased in order to further cook and dehydrate the production. formerly, design on garri production plants did not get the required and acceptable cassava production for the consumers The researchers of those plants did not take into account the specifications of the existing local technology [14].

The UNIBADAN (University of Ibadan) upgraded garri fryer (Igbeka J. E.) is, made of a fireplace oven Incorporated with a chimney and a frying pan. The frying pan which is 200cm x 60cm x 10cm is constructed to have trapezoidal shape with its side inclined at 60° to the horizontal. The clination of the sides allows for gradual gravitational flow of garri down the sides of the fryer. It is made from a 4mm thick black steel sheet, which is not easily corroded and does not turn black after heating [15].

2.2. Mechanized methods

Newell Dunford model

That equipment was a collaborative effort of Federal Institute of Industrial Research (FIRO), Oshodi, Nigeria. and Newell Dunford Company, London. It is a garri processing plant of which the fryer has a unique component. In the frying units, heat generated in the gas fire is measured and regulated by thermostats at several points in the process.

Brazilian model

This equipment's, designed and constructed in Brazil, appears to be better than the Newell Dunford models and the product obtained from it is comparable to Nigeria garri, even though it is not accurately the same. In this Design, frying was not distributed within a given batch and the process looked more like dried cassava mash than cooked and fried garri.

Fabrico model

This design which was constructed and manufactured by a Company, FABRICO, in Nigeria, produces an end-product. That is closer to one in the market(garri). The manufactured goods were not cooked but looked more like roasted garri. The University of Nigeria Nsukka, and the University of Ibadan improved they design.

The UNN model

The UNN (University of Nigeria, Nsukka) design was constructed by Odigboh and Ahmed (1982) to faithfully simulate the village manual frying operations (Odigboh, 1985). The fryer drives automatically produce continuous flow of well fried garri at 16% moisture content. An average through-put of 67kg of garri per hour has been re-counted for this equipment [16].

The Unibadan Model

The UNIBADAN model was designed, constructed and manufactured in the University of Ibadan (Igbeka and Akinbolade, 1986). It is a continuous flow fryer which is an upgrading and modification of the UNN model, hence a modified version [17].

The Fabrico Model.

The modification includes the paddles, the feeding device, and the heat source. The UNIBADAN model is constructed with a fryer plate, feeding hopper, power transmission devices, and shaft with paddles, pulverizes and an oven on which the fryer rests, The UNN model, Is incorporate with a semi-circular trough open at the top, both ends and a fryer plate. It is positioned at an angle of between 5 and 18 with a length of 2.45 m and diameter 0.67 m. The metering device is one of the basic innovations in the model, hopper and the rate of metering is very crucial to the quality of the final product. Another innovation in this model is in the paddles.

Like UNN model, paddles, the main shaft was design with 29 paddles and pulverizes fixed in such a way that they have a conveyor effect at the same time as they press scoop and agitate.

www.astesj.com

The pulverizes press the sieved cassava mash versus the hot pan surface while the paddles scoop and agitate it [18].

2.3. Design Innovation

Recently, innovative design has resultant models such as strategic design, design management and design thinking established speedily. The system of education, policy pertaining innovation and its support have not really measured up with the establishments. Companies that are inexperienced in design-mainly SMEs, low-technological companies, plus those not situated in predominantly design business areas - most times lack knowledge of where to find specialized assistance in design. Marketing of design businesses and its influencing powers are generally being affected by the size of the business.

The strength from adopting Innovation brings about solution to some environmental issues as climate change, and some social inequalities. It is a procedure (or a way of thinking) guiding the synthesis of creativity, technology, scientific and commercial restraints towards producing exceptional (and greater) products, services, and communications.

Good design is progressively vital avenue for businesses to grip theirs in global competition. Design has the power to make products and services more attractive to customers and users, so they are able to sell [19].

Innovation means ideas application towards creating fresh solutions. This solution, however, might be a new product, approach or even a new application of an old product or approach. Innovation is creation or acceptance, acclimatization, and exploitation of a important uniqueness in economic and social domains; regeneration and product increment, services, and markets; establishment of fresh production procedures; and production of new systems of management [20].

2.4. Design Analysis of the Fryer

Analysis is the progression method of making simpler a compound subject or material into slighter parts in order to acquire a good knowledge of it. The practice has been useful in the learning of mathematics and logical reasoning since before Aristotle in (384–322 B.C.), however *analysis* as a recognized concept is rather new development. While engineering analysts study necessities, constructions, machineries, schemes and measurements, electrical engineers investigate electronics systems. Systems Life span and down investigated by engineers, who also study various factors merged in the design. [20].

2.5. Design Analysis of the Mass of the Frying Chamber

Material – Stainless steel

Density, ρ – 8 g/cm³

Number – 1 unit

Length, l – 900 mm = 90 cm

Breadth, b – 900 mm = 90 cm

Height (thickness), $h = 3\text{mm} = 0.3\text{ cm}$

$$\text{Volume} = l \times b \times h \quad (1)$$

$$\text{Volume} = 90 \times 90 \times 0.3 = 2430\text{cm}^3$$

$$\rho = m / v \quad (2)$$

$$m = \rho V = 8 \times 2430 = 19440\text{ g} = 19.44\text{ kg}$$

The mass of the compartment = 19.44 kg

Volume of the Cylinder

Diameter, $d = 300\text{ mm} = 30\text{ cm}$

Height = 900 mm = 90

$$V = \pi r^2 h \quad (3)$$

$$= 22/7 \times 152 \times 90 = 63642.86\text{ cm}^3$$

However, one-quarter of the volume of the cylinder will be needed to have effective frying i.e. $14 \times 63642.86\text{ cm}^3 = 15910.7\text{ cm}^3$

Material – Mild steel

Density – 7.84 g/cm³

Number – 2 units

Length – 670 mm = 67 cm

Breadth – 50 mm = 5 cm

Height (thickness) = 3mm = 0.3 cm

$$\text{Volume} = l \times b \times h$$

$$\text{Volume} = 67 \times 5 \times 0.3 = 100.5\text{cm}^3$$

$$\rho = m/V$$

$$mf = \rho V = 7.84 \times 100.5 = 787.92\text{ g} = 0.788\text{ kg}$$

The mass of the frame = 0.788 kg

Material – Mild steel

Density – 7.84 g/cm³

Number – 2 units

Height – 470 mm = 47 cm

Diameter – 30 mm = 3 cm

$$\text{Volume, } V = \pi r^2 h = 22/7 \times 1.52 \times 47 = 332.35\text{ cm}^3$$

$$\rho = m/ V$$

$$mi = \rho V = 7.84 \times 332.85 = 2605.68\text{ g} = 2.61\text{ kg}$$

The mass of the inner cylinder= 2.61 kg

2.6. Design Analysis of the Outer Cylinder

Material – Mild steel

Density – 7.84 g/cm³

Number – 2 units

Height – 470 mm = 47 cm

Diameter – 30.4 mm = 3.04 cm

$$\text{Volume, } V = \pi r^2 h = 22/7 \times 1.522 \times 47 = 341.27\text{ cm}^3$$

$$\rho = m / V$$

$$m0 = \rho V = 7.84 \times 341.27 = 2675.6\text{ g} = 2.676\text{ kg}$$

The mass of the outer cylinder = 2.676 kg, Source [21]

2.7. Applying Computers to Design

Computers have positively affect engineering like no other. Various engineering areas regularly apply computer for calculation, analysis of information, design and simulation respectively. Computer is used to virtually perform various single tasks within the total design process. When these tasks variously made more proficient, the efficacy of the total process increment becomes pronounced. Computers are specifically located to perform out the areas in design corresponding to the recent mentioned stages of the overall design process. Computers function in the design process via geometric modeling abilities, analysis in engineering controls, programmed testing methods, and automatic drafting [22].

2.8. Environmentally Sustainable Design

People that take to designing initiate ideas regarding resource use, consumption manners and the lifespan of products and services. Globally sustainable design otherwise known as green design has the objective of caring for the environment by making sure that the use of resources that are not renewable are minimized, if not minimized entirely in the course of production, and provision of services. The significance is pronounced in architectural work, municipal designing and scheduling. Certain principles of environmentally sustainable design are stated hereunder as reduced- effect resources designing for use of non-toxic, sustainable product or reprocessed materials which need slight or artificial resources (like energy and water) to convey and process, and the use eco-friendly; Resource efficacy: designing industrial processes, services and products which make use of minimum non artificial resources as possible; Superiority and long-lasting: producing durable and improved operational products that are durable, or grow old in a way that have not effect on the value of the product, dropping the effect of creating another one; Re-claim, recycling and renewability: designing products that can be used again, reprocessed or use as compost after first usage.

2.9. Designs for Recyclability

The following guidelines will contribute to speeding up the disassembly process and recovering a larger proportion of system components: Avoid springs, pulleys, and harnesses which complicate the disassembly process. Minimize the use of adhesives and welds between separable components or between incompatible materials. Adhesives introduce contaminants, can detract from quality due to the potential for bond failure, and increase the costs associated with disassembly. If adhesives are required, try to use adhesives that are with the joined materials.

Use snap fits to join components where possible. Snap fits involve an undercut on one part, engaging a molded lip on a mating part to retain an assembly. Snap fits are relatively inexpensive to manufacture and have attractive mechanical properties. Avoid threaded fasteners (screws), if possible, because they increase assembly and disassembly costs. Use alternative bonding methods, such as solvent bonding or ultrasonic bonding. Such methods may be acceptable for bonding parts made from the same material and which will not be separated at end-of-life. Spring clips or speed clips can be an inexpensive and effective way of joining parts and materials. They permit easy assembly and disassembly, and do not introduce contaminants. Illustrated in the assemble drawing [22] below figure1.

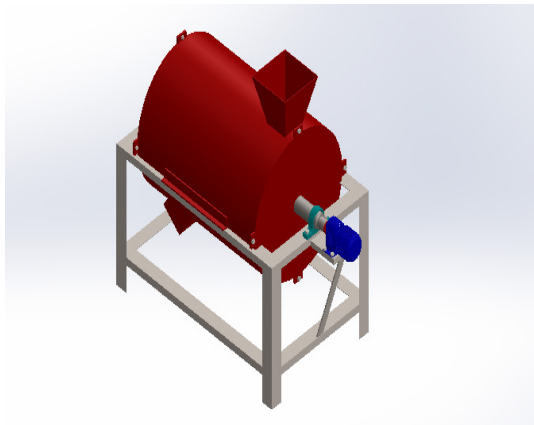


Figure 1: Final Design of Garri frying machine

2.10. Sustainability

Successful industrial processes formally have been remarkable by the application simulation technology aimed towards reducing costs, enhancing production and value, minimizing the time it takes to introduce fresh products to the market. Productions that are sustainable includes the incorporation of procedures, taking decision and ecological anxieties of vigorous production system for the growth of the economy, without affecting negatively environment. The application Sustainability to the whole lifespan of a creation is shown in Figure 1. It includes assortment of constituents, removal of those constituents, of portions, gathering approaches, retailing, produce usage, reutilizing, recapture, and removal ensue for success in simulation application to sustainability. Producers are required to concentrate especially on what initially there were not introduced to. Manufacturers will

need been concerned with before. [23] Illustrate in Figure 3 and Figure 4.

3. Sustainability Analysis of Garri Frying Machine

3.1. Catalogue of DFE Guidelines

The guidelines are divided into four principal strategies,

- Design for dematerialization seeks to reduce the required amount of material throughput, as well the corresponding energy requirements, for a product and its associated processes throughout their life cycle.
- Design for detoxification seeks to reduce or eliminate the toxic, hazardous, or otherwise harmful characteristics of a product and its associated processes, including waste streams that may adversely affect humans or the environment Illustrated in Figure 3.
- Design for revalorization seeks to recover, recycle, or otherwise reuse the residual materials and energy that are generated at each stage of the product life- cycle, thus eliminating waste and reducing virgin resource requirements [24].
- Design for Capital Protection and Renewal seeks to ensure the safety, integrity, vitality, productivity, and continuity of the human, natural, and economic resources that are needed to sustain the product life cycle. There is considerable overlap with other DFX disciplines such as Design for Manufacture and Assembly. Indeed, one strength of DFE is its synergy with other design disciplines. For example, reducing design complexity leads to fewer parts, lower assembly costs, and easier disassembly, resulting in reduced energy and material use as well as increased recyclability a principal strategy for improving sustainability is dematerialization, defined as the reduction of material throughput in an economic system. Dematerialization includes a variety of techniques, such as increasing material efficiency in operations; designing products with reduced mass, packaging, or life-cycle energy requirements; replacement of virgin materials with postindustrial or post-consumer wastes; reducing transportation requirements in the supply chain, thus reducing fuel and vehicle utilization; substitution of electronic services for material-intensive services; and substitution of services for products. These techniques are complemented by other DFE practices, such as recovering Value from obsolete or discarded products shown in figure 4

3.2. Design for Energy and Material Conservation

Reducing energy and material consumption is the most direct way to improve eco-efficiency, i.e., utilizing fewer resources to deliver equivalent or greater value. Decreasing resource intensity results in higher resource productivity, provides immediate reductions in operating costs, and, thus, is synergistic with business goals. In other words, the quantity and costs of purchased energy and materials are reduced by increasing operating efficiency.

Component Environmental Impact

Component	Carbon	Water	Air	Energy
mixer organ 1.	89	0.037	0.514	1100
mixer drum	89	0.037	0.510	1100
Mixer	62	0.206	0.325	660
mixer drum cover	24	0.081	0.127	260
electric motor	24	5.6E-3	0.171	300
pillow bearing	2.8	9.8E-4	0.035	29
hex screw am	0.040	1.7E-5	2.3E-4	0.478

Figure 2: Ten Components Contributing Most to the Four Areas of Environmental Impact

Carbon Footprint



Material:	1700 kg CO ₂ e
Manufacturing:	180 kg CO ₂ e
Use:	0.00 kg CO ₂ e
Transportation:	25 kg CO ₂ e
End of Life:	130 kg CO ₂ e

2000 kg CO₂

Total Energy Consumed



Material:	2.0E+4 MJ
Manufacturing:	1800 MJ
Use:	0.00 MJ
Transportation:	340 MJ
End of Life:	98 MJ

2.2E+4 MJ

Air Acidification



Material:	8.5 kg SO ₂ e
Manufacturing:	2.5 kg SO ₂ e
Use:	0.00 kg SO ₂ e
Transportation:	0.160 kg SO ₂ e
End of Life:	0.068 kg SO ₂ e

11 kg SO₂e

Water Eutrophication



Material:	3.3 kg PO ₄ e
Manufacturing:	0.096 kg PO ₄ e
Use:	0.00 kg PO ₄ e
Transportation:	0.029 kg PO ₄ e
End of Life:	0.167 kg PO ₄ e

3.6 kg PO₄e

Figure 3 Environmental Impact (calculated using CML impact assessment methodology)

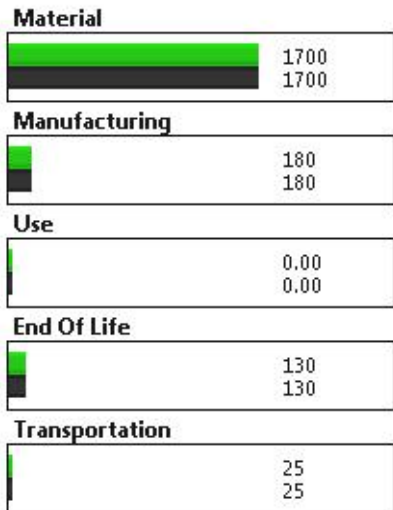
Moreover, energy reduces overall material consumption in the supply chain, since generating energy requires some type of fuel and/or equipment. Although energy management is often pursued as a separate program, energy and material resource conservation should, ideally, go hand-in-hand. Finally, energy conservation that reduces fossil fuel use will also reduce greenhouse gas emissions.

3.3. Life-Cycle Resource Intensity Reduction

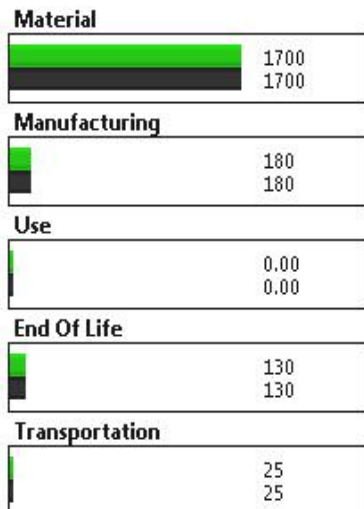
DFE needs to consider the full life cycle of a product, including all of the processes involved in sourcing, production, distribution, use, and recovery of the product. Thus, the investigation of opportunities for energy and material conservation should consider both supplier and customer processes. Depending on geographic locations and type of facilities, certain companies in the supply chain may have much better opportunities than others for energy and material conservation. The following types of opportunities should be explored, shown in Figure 2.

Baseline

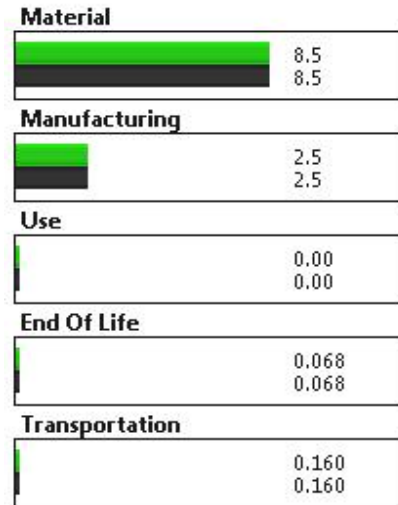
■ Better ■ Worse



Air Acidification – Comparison



Water Eutrophication - Comparison



Material Financial Impact Comparison



Figure 4: Environmental Impact Comparison

Many companies have begun to examine the environmental practices of their suppliers and encourage greater energy and material efficiency. This can reduce the life-cycle footprint of their own products and potentially lower their costs. The most prominent example is Wal-Mart, which has developed sustainability scorecards for packaging and energy use and is requesting environmental performance improvements from all of its suppliers.

3.5. Reduce the Operational Resource Footprint

Companies have found a great deal of “low-hanging fruit” by tightening up energy management practices, e.g., heating, cooling, and lighting systems, and materials management practices, e.g., maintenance, inventory, and waste management. Newer facilities are being designed with recycled materials and advanced energy-saving features, as interest in “green building” has mushroomed. But the largest gains in resource conservation come from redesigning production processes to reduce throughput requirements and install more efficient equipment. Example: From 2005 to 2007, General Electric (GE) conducted a “Lean and Energy” initiative that identified over \$100 million in potential energy savings through over 200 “energy treasure hunts” at GE facilities worldwide. This effort resulted in 5,000 related kaizen projects, most of which are funded and in various stages of implementation. GE was able to reduce greenhouse gas emissions by 250,000 metric tons and realized \$70 million in energy cost savings from implemented projects.

4. Sustainability Analysis in Garri Frying Machine Manufacturing in Nigeria: Design Innovation

In Nigeria, it appears that those agriculturists are not rich as to compare with other people in other sectors of the economy. That

is to say that their living standard is so low that acquiring facilities is a major problem that among others affect their agricultural development.

Studies confirmed that the little processing equipment that are small are small efficiency. This has actually affect negatively the production rate of farmers.

Productivity in Agriculture equals the ratio of outputs to inputs. Whereas discrete goods are typically stated using weight, various densities of them make difficult determining total agricultural output. As a result, measuring the output is dependent on market worth of last output, which omits middle crops use in meat industry as corn feed. The output worth is compared with yield from diverse categories of inputs labour and land. These are productivity partial actions. There is transformation basically on how we think of management of manufacturing as the four concepts develops. United States manufacturers are not ignorant of the fact that we need manufacturing theory that is recent. We are aware that mending ancient theories are not workable and in furtherance of it will really throw us backwards. Actually, these ideas offer us the basis for the requisite theory.

Immediately we describe manufacturing as the procedure that changes belongings for the satisfaction of the economy, obviously, production continues even when the produce originate from the company. Physical circulation of products and services which are portion of the producing procedure are combined with it, synchronized it, achieved organized with it. It is even now broadly documented that overhauling the produce is most important thought in the course of design and production. Customarily, manufacturing commerce have been prearranged "in series," with tasks as manufacturing, engineering, and advertising as sequential stages. Nowadays, that system is frequently supplemented by a similar crew body (Procter & Gamble's product management teams are a well-known example), which gives several tasks together as of the commencement of a novel produce or procedure scheme. If industrial activities are to be system, nevertheless, every single resolution in a manufacturing business turn out to be a manufacturing decision. Every single decision be duty-bound to meeting with the manufacturing's desires and requirements, and sequentially ought to exploit the powers and aptitudes of a company's exact engineering system. Undeniably, we have power to move ahead and involve executives all over the company to go round the factory projects during the course of their careers. Therefore, every single manufacturing boss ought to acquire as well as run-through a training that incorporates engineering, organization of people, as well as business economics into the manufacturing process. Moderately a small number of manufacturers presently work on that unknowingly. Hitherto such training has not been systematized as well as not imparted in engineering.

God so generous to Nigeria, nonetheless Nigerians have not been benevolent to Nigeria. Our country has been extraordinarily preferential vested upon with natural, mineral, and human resources. And the fertility of the lands is such that planting one

finger will grow to becoming a human being, but we lack leadership skills [25].

It is as a result that of the above that this work theme was resultant and the benefits are as ensuing [26].

Cost Effective: it has low as well as inexpensive cost within the wealthy. Besides, it needs slight or no recurrent looking after.

- Increase in Productivity- productivity increment brings about affluence increase of a Nation. Productivity of garri frying machine drive maximally in bringing about extent of increase in productivity level of a Nation which will rise worth totaling to GDP.
- Poverty Reduction and Job Creation: Its assistance in job creation is enormous as seen it its necessitating reduction in the migration between rural-urban.
- Usage and Observation by Local Fabricators. Ever since the machine is invented locally, the procedure drives local producers to modernize or improve in their construction. Therefore, cumulative the health rural people.
- Decrease in Joblessness. The decrease of joblessness are those machine fabricators as well as those using the machine will be the recipients.
- Inspire Direct as well as Indirect Investment, It drive hastily growth in industrial creation for garri manufacturing as well as further cassava processing projects. It will likewise intensify speedy growth and development of the Nation's economy.
- Economic Improvement and Industrial development. It will help in economic improvement and industrial development for the reason that it necessitates the formation of fresh asset culture, affluence establishment as well as enlarged economic plus social welfare. Entirely, these brings about enhanced application of cassava beforehand to avoid spoiled owing to small shelf life [27].

5. Conclusion/Recommendation

In accurately united method, DFE must be well-adjusted in contrast to other cost as well as superiority factors that sway design choices. The spot of an efficacious group is the capacity to revolutionize underneath stress, somewhat than conceding product excellence. [28] A "win-win" result is the bringing about of ecologically advantageous improvements that likewise advance the charge in addition to functionality of the invention as soon as is observed as portion of the total system [29]. If at all possible, lone design invention might donate to accomplishing quite a lot of dissimilar sorts of objectives. For instance, decreasing the quantity of a produce be able to cause in 1 energy in addition to measureable decrease, which gives to resource protection, as well as noxious waste release reduction, which gives to healthiness as well as security. Creation trade- off ideas is the greatest perplexing portion of the course for the reason that of the prerequisite to at once think through so numerous dissimilar standards. Base on this conversation the succeeding policy stay essential, hard work ought to be done to accept as well as disseminate the design-DFX, DFA, DFE illustrated in Figure1- 4 above ETC particularly for the profits of menfolk who create up

a pronounced proportion of the Nation's inhabitants. If, the usage of machine design innovations embraced, the difficulty in garri as well as supplementary agricultural handing out Tools will be lessened besides hunger as well as poverty will be exterminated.

Conflict of Interest

We declare no conflict of interest associated with this manuscript.

Acknowledgment

This Research was Sponsored by Tertiary Education Trust Fund(TETFUND) No6 Zambezi Crescent. off Agniyi Ironsi Street, Maitama, Abuja, Nigeria Tel: 070 98818818.

References

- [1] Akinyemi, J. O. and Akinlua. (1999). Design, Construction and Testing of Cassava Grater. *International Journal of Tropical Agriculture*. 17(1-4); pp103-108.
- [2] Olukunle, O. J. and Ademosun, O. C. (2006). Development of a double action self-fed cassava peeling machine. *Journal of food, Agriculture and Environment (JFAE)*, Accepted for publication
- [3] Olukunle O. J. and Oguntunde P. G. (2008). Analysis of Peeling Pattern in an Automated Cassava Peeling System. *Nigerian Journal of Technological Development* Vol.6 No. 1&2 41-52
- [4] Olukunle O. J. and Atere, A. O. (2009). Developments in Cassava Peeling Mechanisation. Proceedings of the International Conference the Nigerian Institution of Agricultural Engineers/West African Society of Engineers, held at Obafemi Awolowo University, Ile-Ife
- [5] Odigboh, E. U. (1983). Cassava Production, Processing and Utilization. In: Chan Jnr., H. T. (ed), *Handbook of Tropical Foods*. Marcel Decker Publisher, Inc.; 270, Madison Avenue, New York; pp145-200
- [6] Igbeka, J. C. (1995). Recent Developments in Cassava Frying Operation and Equipment used for Garri Production in Nigeria. *ORSTOM*: pp583-590
- [7] Jackson, A. T. and Lamb, J. (1981). *Calculations in Food and Chemical Engineering (Theory, Worked Examples and Problems)*. Mac Millan press Ltd, London. pp209.
- [8] Odigboh, E. U. and Ahmed, S. F. (1982). Design of the Continuous Process Garri Frying Machine. *Proceedings of the Nigeria Society of Engineers*; 6(2); 65-75
- [9] Gbasouzor Austin Ikechukwu, A. I. V. Maduabum (2012) Improved Mechanized Garri Frying Technology Sustainable Economic Development in Nigeria by proceeding s of international multiconference of Engineers and computer scientists
- [10] Chime .O Thompson, Inyama Fidelis Chidozie and Okonkwo Gloria Ngozi (2016) Design Innovation, Modelling and Simulation: Sustainability of Analysis of Bench Reactor for Kinetic Study of Hydro Carbon Removal Using Land Farming published by International Journal of Engineering Technology and Computer Research (IJETCR) www.ijetcr.org
- [11] Asheen Phanse, "Biomimicry," *Berkshire Encyclopedia of Sustainability: The Business of Sustainability* (New York: Berkshire Publishing Group, 2009), p. 37
- [12] Michael Beany (Summer 2012). *Analysis. The Inford Encyclopedia of Philosophy*. Michael Beey Retrieved 23 May 2012.
- [13] Our Common Future, Report of the World Commission on Environment and Development, World Commission on Environment and Development, 1987. Published as Annex to General Assembly document A/42/427, Development and International Co-operation: Environment August 2, 1987.
- [14] R. Chime, et al (2017) Improving Productivity In Hollow Impeller Palm Nut Cracking Machine Manufacturing In Nigeria published by Internal journal of Engineering technology and Computer Research (IJETCR) WWW.ijetcr.org March/April Edition
- [15] Okechuku (2012) Improved Mechanized Garri Frying Technology for Sustainable Economic Development in Nigeria Proceedings of the International Multi Conference of Engineers and Computer Scientists 2012 Vol II, IMECS 2012, March 14 - 16, 2012, Hong Kong
- [16] J.C. Igbeka, mechanization for cassava processing", *Journal of Agricultural Mechanization in Asia, Africa and Latin America (AMA)*, 22(1), pp. 45-50, 1972 M. Jory and D. Griffon, "Selective
- [17] E.U. Odigboh. (1985) "Prototype Machines for small- and medium-scale harvesting and processing of cassava, Yaounde, Cameroun. pp 323 – 338,
- [18] R .O.Chime ,et al (2017) Improving Productivity In Hollow Impeller Palm Nut Cracking Machine Manufacturing In Nigeria published by Internal journal of Engineering technology and Computer
- [19] Charles, M. *Manufacturing Simulation; The need for Standard Methodologies, Model, and Data Interfaces*, 2009.
- [20] J.A.Onuigbo and R.O.Chime (2018) Innovation in Feed Mixing Machine: Design for Manufacturing in Industry by World Journal of Engineering Research and Technology www.wjert.
- [21] B.O. Akinnuli, C.O. Osueke, P.P. Ikubanni O.O. Agboola and A.A. Adediran(2015)Design Concepts Towards Electric Powered Garri Frying Machine by International Journal of Scientific & Engineering Research, Volume 6, Issue 5, May-2015 1043 ISSN 2229-5518
- [22] The Danish government's 2007 white paper on design,
- [23] Design Issues," in *Tool and Manufacturing Engineers Handbook*, Dearborn, Mich., 1992
- [24] Joseph Fiksel: *Design for Environment: A Guide to Sustainable Product Development*, Second Edition. Design Rules and Guidelines, Chapter (McGraw-Hill Professional, 2009 1996), Access Engineering
- [25] R.O.Chime(2006)leadership Training published by Coal City Engineer The Nigerian Society of Engineer Enugu State Branch
- [26] J.A.Onuigbo and R.O.Chime (2018) Innovation in Feed Mixing Machine: Design for Manufacturing in Industry by World Journal of Engineering Research and Technology www.wjert
- [27] R .O.Chime ,et al (2017) Improving Productivity In Hollow Impeller Palm Nut Cracking Machine Manufacturing In Nigeria published by Internal journal of Engineering technology and Computer Research(IJETCR)WWW.ijetcr.org March/April Edition
- [28] Peter F. Drucker (1999) *The Emerging Theory Manufacturing*, May–June 1990 issue of *Harvard Business Review*.
- [29] Bulent Sezen, Sibel Yildiz Cankaya (2013) Effects of green manufacturing and eco-innovation on sustainability performance Published by Elsevier Ltd. Selection and peer-review under responsibility of the International Strategic Management Conference www.sciencedirect.com
- [30] *The Emerging Theory of Manufacturing Peter F. Drucker* from the May–June 1990 Issue
- [31] Design Issues," in *Tool and Manufacturing Engineers Handbook*, Dearborn, Mich., 1992
- [32] Douglas Harper (2001–2012). "Analysis(N)" Online. *Etymology Dictionary* . Douglas Harper. Retrieved 23 May 2012.

Two Degree-of-Freedom Vibration Control of a 3D, 2 Link Flexible Manipulator

Waweru Njeri*, Minoru Sasaki, Kojiro Matsushita

Mechanical Engineering, Mechanical Engineering, 1-1 Yanagido, Gifu, Japan

ARTICLE INFO

Article history:

Received: 24 October, 2018

Accepted: 2 December, 2018

Online: 15 December, 2018

Keywords:

Inverse control

Strain feedback

System stability

Two-degree-of-freedom

ABSTRACT

Considering link vibrations, the main limitation affecting flexible manipulators, this article seeks to make a contribution by presenting an enhanced two degree of freedom vibration controller. This controller uses a filtered right inverse controller in the feedforward and strain feedback controller in the feedback path. The Filtered inverse controller damps transient vibrations while preserving joint trajectories. On the other hand, strain feedback controller ensures a rapid decay of residue vibrations. Modeling of the manipulator was carried out in Maplesim, linearized and inverted in Matlab. Experiments were conducted in the dSPACE environment. Both the simulations and the experimental results showed that the two-degree-of-freedom controller yielded a superior performance over the two controllers individually.

1 Introduction

This journal article, an extension of our original work presented in the 2018 IEEE International conference on Applied System invention(ICASI) [1] seek to make a contribution by presenting a two degree of freedom controller comprising of a filtered inverse controller in the forward path and a Direct Strain Feedback controller(DSFB) in the feedback path. The strength of the proposed methods lies in the fact that it can suppress both the motion induced vibrations, as well as residue vibrations, and it is superior than the individual controllers separately.

Since the dawn of the industrial revolution, over and above making the work easier, man has been thinking of how to replace human labour altogether. This has brought to fruition many kinds of robots and manipulators to fulfil this dream. These robots came handy in carrying out repetitive chores, in hazardous work environment and in heavy industries to mention just but a few. Initially, robots comprised of bulky links driven by huge motors. Thus, they suffered from link inertia or were rather driven at low speeds and were expensive to operate considering the amount of power they consumed and the cost of maintenance. Challenges of rigid robots, except for heavy tasks and for application where low speeds is not a problem, were addressed by the introduction of flexible manipulators.

These Flexible manipulators come with lots of merits over rigid manipulators; like being light in weight hence, small actuators can be used. Also, the maintenance and running costs are low. However, their links are flexible, hence they tend to vibrate especially when operated at high speeds. These vibrations increases with both loading and additional links, due to coupling, which adversely affects the accuracy and duplicity of tasks. Also, link vibration leads to time wastage, hence having to wait for the residue vibrations to decay to healthy levels before the end-effector can be put to use. Link vibrations also raises a safety issue, as continued vibration can lead to mechanical failure due to fatigue, thereby reducing the life span of the manipulator, and posing a risk to the operators.

In a bid to reap all the benefits of the flexible manipulators, a lot research has been done to mitigate link vibrations and all its shortcoming. Researchers [2, 3] introduced a vibration control method for the flexible manipulators based on internal resonance. In their scheme, they designed an absorber comprising of a servomotor, sliding along the links of the manipulator and driving a branched link. At internal resonance, the energy stored in the links in form of vibration energy, is propagated back and forth between the different modes and is dissipated to the absorber. One challenge associated with their method includes the complication in establishing the internal resonance considering the

* Corresponding Author, Waweru Njeri, Mechanical Engineering Department, Gifu University, 1-1 Yanagido, Gifu, Japan
Email: v3812104@edu.gifu-u.ac.jp

different loading and trajectories. Another limitation is the feasibility of using this scheme with additional links, and the inertia introduced by the absorber.

In [4], the author proposed a novel filtered input shaping technique, implemented inside the position control loop of a single link flexible manipulator. The controller was developed based on the dominant vibration modes of the manipulator to mitigate the motion-induced vibrations from the high speed operations, which arose also from the flexible nature of the link. The controller was based on a lowpass and a bandpass elliptic filters to attenuate dominant modes in the input signal, and avoid exciting the manipulator at its natural frequencies. With fixed filters, however, the performance of their system would deteriorate with additional loading, since dominant modes are bound to change. The limitations of the fixed filters are addressed in [5]. There exist similar work involving input shaping for vibration control, for example [6–8].

In [9, 10], the authors noted that vibrations occur after the manipulator was brought to a sudden stop. They found that the severity of the induced vibration was largely dependent on the initial speed, speed prior to the stop and the duration of the motion. The higher the speed, the longer the vibrations will ensue, before the end pointer settles to the desired position. In their work, they developed trapezoidal and triangular velocity profiles to reduce joint velocities at the beginning of the operation and prior to stopping. In other words, the trajectory was portioned into acceleration period, constant speed period and deceleration period. The scheme resulted in minimal vibrations naturally as the manipulator decelerated to a stop.

Another popular active vibration control technique is the use of smart structures in form of piezoelectric actuators which are bonded to the root of the manipulator [11]. Authors in [12] made use of shunted piezoelectric transducers to mitigate vibrations. In this type of solution, the piezoelectric transducers are bonded onto the links of the manipulator. Link flexure is converted to corresponding voltages which are used as feedback. Under this broad scheme, alternative vibration suppression methods include: Position positive feedback [13], strain rate feedback control [14], quantitative feedback theory [15], and resonant control [16]. Current trends include using Neural Networks to tune PID gains [17], filtered inverse control [18], robust control, in particular H_∞ together with piezoelectric actuators [19], boundary control [20].

The rest of the article is organized as follows; Modelling of the two link 3D flexible manipulator is introduced in section 2. Development of the inverse model is highlighted in section 3. Application of the developed inverse is applied to the manipulator 4. Simulation and experimental results are presented and discussed in section 6 followed by conclusion in section 7.

2 Modeling and validation of the manipulator

The basic prerequisite for model based controller design is an accurate model of the plant to be controlled. This is an easy feat for simple systems which can easily be obtained from first principles by paper and pen observing the governing equations. However, for complex system especially if they are nonlinear and dynamic is not an easy task.

Advancements in the processing power of modern computers has brought forth another method of obtaining models of complex system by employing symbolic softwares. Examples of such softwares include but not limited to Maple/Maplesim[®], Mathematica[®], Matlab/Simscape[®] all based on Modelica[®] library. In symbolic modeling, mathematical models of common engineering parts such as joints, links, motors with possibilities of customization are dragged onto a workspace to form a complex system. Simulations are carried out considering a practical environment. The strength of this technique lies in its accuracy, simplicity and the possibility of including aspects which cannot be represented mathematically.

The plant reported in this article is a two link, 3D flexible manipulator, with technical description as tabulated in Appendix A.1. It has three rigid joints, driven by dc servomotors and fitted with harmonic drives to reduce joint velocities by a factor of one hundred. It has two flexible links with a variable load attached at the far end of the link number 2.

The control system includes a computer running the Matlab/Simulink and interfaced with dSPACE DSP board, which serves as Servomotors driver via DA converters and collection of data via AD converters. Operation of the system is carried out from the dSPACE control desk environment. The measurement of the joint angles and the velocities were achieved using encoders fitted at the bottom part of the servomotors. Strain data was obtained from strain gauges attached at the root of the links. Strain information was conditioned using wheatstone bridge, filtered and amplified before being transmitted to the computer via AD converters. The control system is configured as in Figure 1.

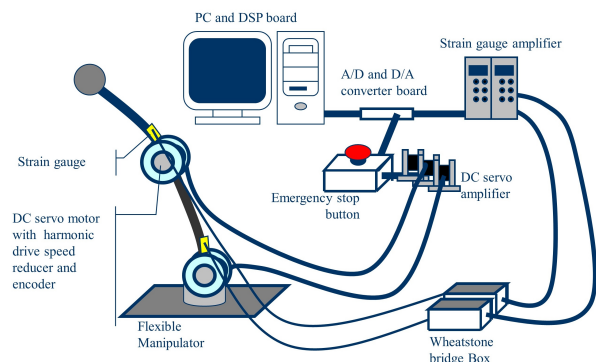


Figure 1: Control system setup

The manipulator and the control system were modelled and linearized in Maple/Maplesim and its

inverse model was developed in Matlab. Validation of the model against the actual manipulator was performed and a perfect agreement was observed between the nonlinear model, linearized model and the actual manipulator, as will be seen later in the results. State space matrices of the linearized model are posted in the Appendix A.3.

3 Development of the inverse system

To develop an inverse model, consider an Linear Time Invariant(LTI) continuous time square system $\Sigma(t)$, and let the triplet A,B and C be a minimal state-space representation. It is assumed that the system is stable or stabilized by negative feedback.

$$\dot{x}(t) = Ax(t)+Bu(t), \quad x(0)=x_0, \quad t \in \mathbb{R}^+ \quad (1)$$

$$y(t) = Cx(t) \quad (2)$$

where $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^p$, $y(t) \triangleq (y_1, y_2, \dots, y_p)^T \in \mathbb{R}^p$, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$ and $C \in \mathbb{R}^{p \times n}$.

Definition 1. Given $\Sigma(t)$, an LTI system defined above in equations 1 and 2, inversion involves the development of a model $\Sigma^{-1}(t)$ that yields the input control law $u_f(t)$ to reproduces $y(t)$ when used as the input to $\Sigma(t)$.

Definition 2. If C_i denotes the i_{th} row of the output matrix C, then the system is said to have a relative degree $r \triangleq (r_1, r_2, \dots, r_p)^T$ if $C_i A^l B = 0, \forall l < r_i - 1; 1 \leq i \leq p$ [21]. Further, if this holds true in the entire domain in the states, then we say the system has a well defined relative degree.

Following Definition 2 above and assuming that the system has a well-defined relative degree $r = (r_1, r_2, \dots, r_p)^T$, differentiating the i_{th} output r_i times w.r.t time yields

$$y^{(r_i)} = C_i A^{(r_i)} x + C_i A^{(r_i-1)} B u$$

where C_i is the i_{th} row of the output matrix C for $1 \leq i \leq p$ and the subscripts represent the Lagrange's notation of the r_{ith} derivative in time. Repeating this for all the rows and having the resulting expressions in vector form, we have

$$y^{(r)} = A_x x(t) + B_y u(t) \quad (3)$$

where

$$y^{(r)} \triangleq \begin{bmatrix} y_1^{(r_1)}(t) \\ y_2^{(r_2)}(t) \\ \vdots \\ y_p^{(r_p)}(t) \end{bmatrix}$$

$$A_x \triangleq \begin{bmatrix} C_1 A^{(r_1)} \\ C_2 A^{(r_2)} \\ \vdots \\ C_p A^{(r_p)} \end{bmatrix}$$

$$B_y \triangleq \begin{bmatrix} C_1 A^{(r_1-1)} B \\ C_2 A^{(r_2-1)} B \\ \vdots \\ C_p A^{(r_p-1)} B \end{bmatrix}$$

From equation (3), and the fact that B_y is invertible

because of the well defined relative degree assumption, the control law is

$$u(t) = B_y^{-1} [y_d^{(r)} - A_x x(t)] \quad \forall t \in (-\infty, \infty) \quad (4)$$

There exist a state transformation $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$

$$x(t) = T \begin{bmatrix} \zeta(t) \\ \sigma(t) \end{bmatrix}^T$$

which decomposes the states into internal dynamics(system states, which are not directly controlled by the input $u(t)$), $\sigma(t)$ and the external dynamics $\zeta(t)$, (i.e, the output and its derivatives in time up to $(r_i - 1)$) as

$$\zeta = [y_1, \dot{y}_1, \dots, y_1^{(r_1-1)}, \dots, y_p, \dot{y}_p, \dots, y_p^{(r_p-1)}]^T \quad (5)$$

The expression of the new system after coordinate transformation is

$$\dot{\zeta} = \hat{A}_1 \zeta + \hat{A}_2 \sigma + \hat{B}_1 u$$

$$\dot{\sigma} = \hat{A}_3 \zeta + \hat{A}_4 \sigma + \hat{B}_2 u$$

where $\hat{A} = \begin{bmatrix} \hat{A}_1 & \hat{A}_2 \\ \hat{A}_3 & \hat{A}_4 \end{bmatrix} = T^{-1} A T$ and $\hat{B} = \begin{bmatrix} \hat{B}_1 \\ \hat{B}_2 \end{bmatrix}$. Replacing

$x(t)$ in (4) with the transformed dynamics, the control law to maintain the exact tracking can be written as

$$u_f = B_y^{-1} [y_d^{(r)} - A_\zeta \zeta(t) - A_\sigma \sigma(t)] \quad (6)$$

where

$$[A_\zeta \quad A_\sigma] = A_x T$$

internal dynamics can now be expressed as

$$\dot{\sigma} = \hat{A}_3 \zeta + \hat{A}_4 \sigma + \hat{B}_2 B_y^{-1} [y_d^{(r)} - A_\zeta \zeta(t) - A_\sigma \sigma(t)]$$

$$= \hat{A}_\sigma \sigma(t) + \hat{B}_\sigma Y \quad (7)$$

where

$$\hat{A}_\sigma = \hat{A}_4 - \hat{B}_2 B_y^{-1} A_\sigma$$

$$\hat{B}_\sigma = [(\hat{A}_3 - \hat{B}_2 B_y^{-1} A_\zeta) \quad \hat{B}_2 B_y^{-1}] \text{ and}$$

$$Y = [\zeta^T \quad y_d^{(r)T}]^T$$

in the same respect, equation (4) can now be written as

$$u(t) = B_y^{-1} [y_d^{(r)} - A_\zeta \zeta(t) - A_\sigma \sigma(t)]$$

$$= -B_y^{-1} A_\sigma \sigma(t) - [B_y^{-1} A_\zeta \quad -B_y^{-1}] Y$$

$$= \hat{C}_\sigma \sigma(t) + \hat{D}_Y Y(t) \quad (8)$$

where

$$\hat{C}_\sigma = -B_y^{-1} A_\sigma \text{ and}$$

$$\hat{D}_Y = -[B_y^{-1} A_\zeta \quad -B_y^{-1}]$$

Equation (7) together with equation (8) form the inverse system and can be represented in state space form

$$\dot{\sigma}(t) = \hat{A}_\sigma \sigma(t) + \hat{B}_\sigma Y(t) \quad (9)$$

$$u(t) = \hat{C}_\sigma \sigma(t) + \hat{D}_Y Y(t) \quad (10)$$

and represented as in Figure 2(see Appendix A.4 for details of matrices $A_\sigma, B_\sigma, C_\sigma, D_Y$).

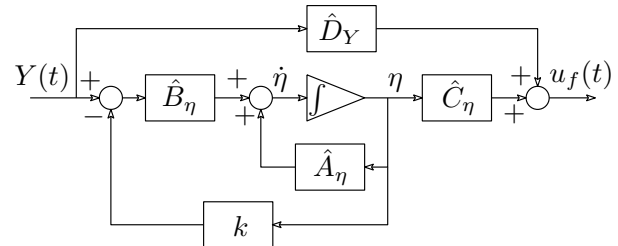


Figure 2: Block diagram of the inverse system

4 Inverting the manipulator

The linear model has 17 states distributed as:

- $x_1(t) = i_1(t)$ • $x_7(t) = \dot{w}_{21}(t)$ • $x_{13}(t) = \dot{\theta}_2(t)$
- $x_2(t) = w_{11}(t)$ • $x_8(t) = w_{22}(t)$ • $x_{14}(t) = \theta_3(t)$
- $x_3(t) = \dot{w}_{11}(t)$ • $x_9(t) = \dot{w}_{22}(t)$ • $x_{15}(t) = \dot{\theta}_3(t)$
- $x_4(t) = w_{12}(t)$ • $x_{10}(t) = \theta_1(t)$
- $x_5(t) = \dot{w}_{12}(t)$ • $x_{11}(t) = \dot{\theta}_1(t)$ • $x_{16}(t) = i_3(t)$
- $x_6(t) = w_{21}(t)$ • $x_{12}(t) = \theta_2(t)$ • $x_{17}(t) = i_2(t)$

where i_j denotes the armature current to the servomotor driving joint j ($j = 1, 2, 3$), θ_j and $\dot{\theta}_j$ are the instantaneous joint angles and joint velocities of joint ($j = 1, 2, 3$), respectively, whereas (w_{11}, w_{12}) , (w_{21}, w_{22}) and their derivatives denote the flexure variable for links 1 and 2 respectively.

Remark 1. In the modelling of the manipulator in Maplesim, the lengths of links 1 and 2 are broken into two to accommodate an instrument to measure the strain. In regard to this, in the linearized model, the flexure variable has two parts as w_{11}, w_{12} for link 1 and w_{21}, w_{22} for link 2. Except for having twice as many flexure variables as the number of links, breaking the links does not affect the performance of the model.

With a relative degree of $r = (3, 3, 3)$, the internal dynamics, $\sigma(t)$, were taken as the flexure variables, whereas the output variables and their derivatives ζ , were taken as the three joint angles, velocities and motor currents, i.e.

$$\sigma(t) = \begin{bmatrix} x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \\ x_6(t) \\ x_7(t) \\ x_8(t) \\ x_9(t) \end{bmatrix} = \begin{bmatrix} w_{11}(t) \\ \dot{w}_{11}(t) \\ w_{12}(t) \\ \dot{w}_{12}(t) \\ w_{21}(t) \\ \dot{w}_{21}(t) \\ w_{22}(t) \\ \dot{w}_{22}(t) \end{bmatrix}, \quad \zeta(t) = \begin{bmatrix} x_1(t) \\ x_{10}(t) \\ x_{11}(t) \\ x_{12}(t) \\ x_{13}(t) \\ x_{14}(t) \\ x_{15}(t) \\ x_{16}(t) \\ x_{17}(t) \end{bmatrix} = \begin{bmatrix} \theta_1(t) \\ \dot{\theta}_1(t) \\ \theta_2(t) \\ \dot{\theta}_2(t) \\ \theta_3(t) \\ \dot{\theta}_3(t) \\ i_2(t) \\ i_3(t) \\ i_1(t) \end{bmatrix}$$

Poles and zeros of the linear model and its inverse are as shown below

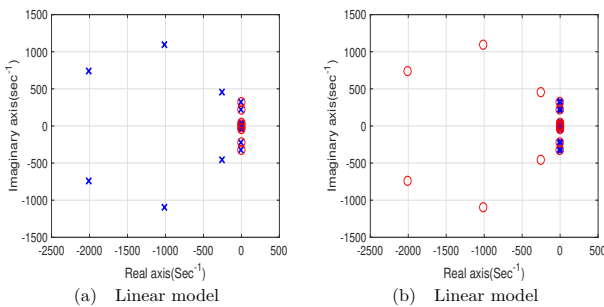


Figure 3: Poles and zeros of the plant and its inverse

Upon inversion, the eight poles corresponded to the internal dynamics, i.e. the flexure variables. These poles were found to be lying on the imaginary axis implying that these variables are marginally stable and that they will not decay to zero with time. This was addressed by applying pole placement technique to slightly shift these poles to the right. This is important

as it means that the stability of the internal dynamics and the resulting inverse is assured. Also the stable model still remains to be the inverse of the linear model. Consequently the stable solution of the internal dynamics follows from solving the ODE in equation 9 as

$$\sigma(t) = e^{A_\sigma t} \sigma(0) + \int_0^t e^{A_\sigma(t-\tau)} B_\sigma Y(\tau) d\tau \quad (11)$$

where τ is a dummy variable. The first term represent the zero-input response whereas the second term is the zero-state response. With negative eigenvalues of the matrix A_σ , which is enforced using pole placement technique, it can be deduced from this expression that the internal dynamics $\sigma(t)$ are bounded for bounded external dynamics $Y(t)$.

Driving the manipulator via an inverse controller implies that the joint angles will follow the desired trajectories exactly. For the high speed operations involving step or square wave trajectories; however, joint velocities during the rising and the falling edges would be too high thus not safe for the operators. It can also lead to mechanical failures. This was addressed by introducing second order bilinear low pass filter, before the inverse controller, of the form

$$f(s) = \frac{1}{(\lambda s + 1)^2}$$

where λ is an adjustable parameter for limiting the manipulator speeds to safe levels.

Remark 2. Operation without filter leads to very high speeds, hence exposing the operator and environment to risk, also risking the mechanical well being of the manipulator.

Remark 3. Operation with filters without the inverse controller means that the high frequency components of the trajectories are removed leading to a very high joint error.

Remark 4. The inverse controller ensures that the joint trajectories follows the desired trajectory, the filter ensure safe operation speeds.

5 Direct strain feedback control

The theory of Direct Strain Feedback (DSFB) was developed by Luo [22, 23] and experimented with a one-link flexible manipulator. In this control scheme, the strain measured at the root of the flexible link is multiplied by a constant gain k and the resultant signal is used to modify the control law as a negative feedback. The overall effect of direct strain feedback is to increase the system damping coefficient thereby leading to a rapid decay of transverse vibrations and torsional vibration as a result of coupling between the two types of vibrations. In [22], the author shows that the technique can satisfactorily dampen link vibrations. He also analytically derived the proof that the resulting closed loop is asymptotically stable. A block diagram showing the hybrid of the filtered inverse feedforward controller and the DSFB is shown in Figure 4.

From the figure, the new control law is expressed as $u(t) = \theta_f(t) - \theta(t) - k\varepsilon(0, t)$ where:

- $\theta_f(t)$ - Trajectory tracking signal generated by the filtered inverse controller
- $\theta(t)$ - Joint angle of the flexible manipulator
- $\varepsilon(0,t)$ - Strain at the root of the links
- k - Strain feedback controller gain

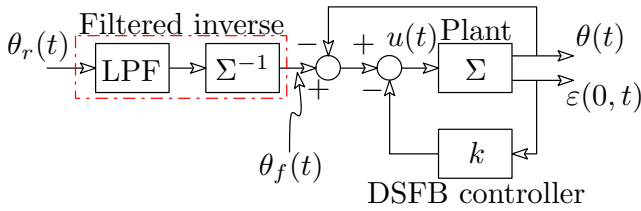
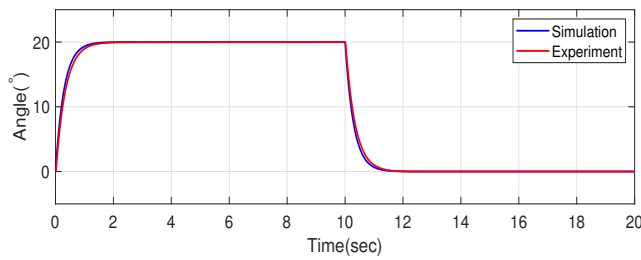


Figure 4: Hybrid control setup

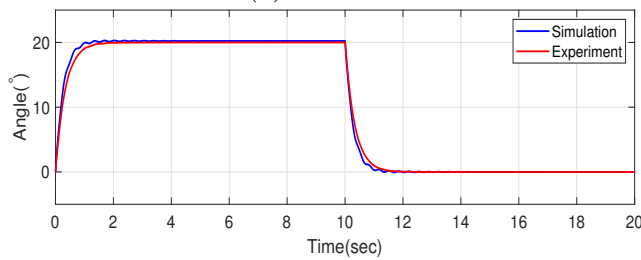
6 Results and Discussion

6.1 Modelling and validation results

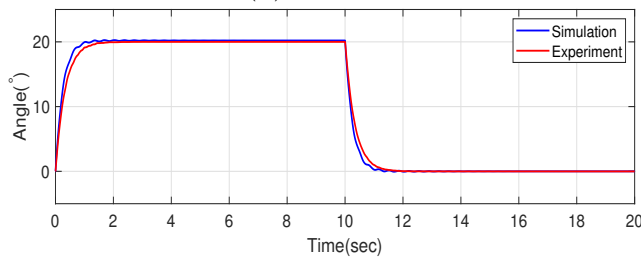
To validate the model developed and linearized in section 2, simulations were conducted in Matlab Simulink on the linearized models and the performance of the model compared with the existing flexible manipulator. The task involved moving the joints as in a simple pick and place task popular in industries e.g soldering, painting etc.



(a) Joint 1



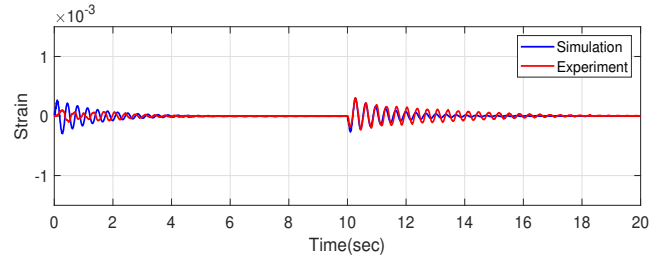
(b) Joint 2



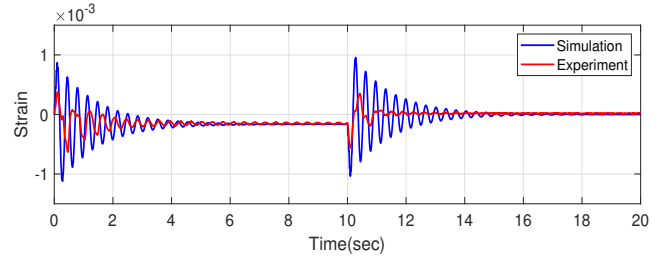
(c) Joint 3

Figure 5: Comparison of Joint angles between linear and the actual manipulator

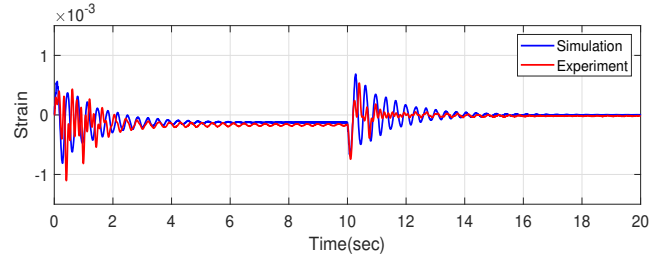
Figure 5 shows the joint trajectories for joint 1, 2 and 3 for the model and the manipulator. From the figures, we can see a perfect agreement of the model with the existing manipulator.



(a) Torsion



(b) Link 1



(c) Link 2

Figure 6: Comparison of strain between linear and the actual manipulator

Figure 6 validates the model employed in this work in terms of the vibrations excited. It shows the strain information of the linear model against the actual manipulator. Perfect agreement between joint angles in torsional and links strain in 6(a-c) can be observed. Observations in figures 5 and 6 leads to the deduction that the linear model represents an accurate model of the manipulator. In addition, the inverse derived from the linearized model represent an accurate inverse of the actual manipulator.

6.2 Simulation results

This section presents simulation results of the validated model subjected to the inverse controller, the strain feedback controller and a hybrid controller of the two. Simulations were conducted in Matlab/Simulink for a desired joint trajectory that involved moving the joints at an angle of 20 degrees for 10 seconds and back to the vertical position for 10 more seconds. Typical applications of such trajectories are in soldering and other pick and place related tasks. Simulink model of the simulation setup is as shown in Figure 7.

The servomotors used here were of the speed reference type. Thus, the angle feedback formed the outer feedback loop, having unity feedback gain, while strain feedback was in the inner loop with a feedback gain $k = 0.4$. To avoid errors due to self-weight, the gravitational effect is compensated for in the strain

signal. The compensation was done by removing the self-weight offset before being fed back through the DSFB controller.

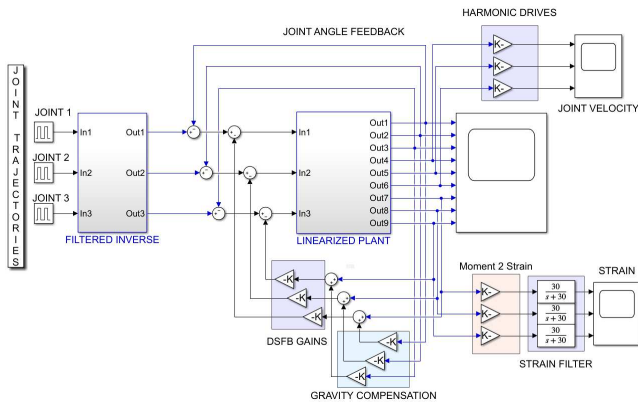
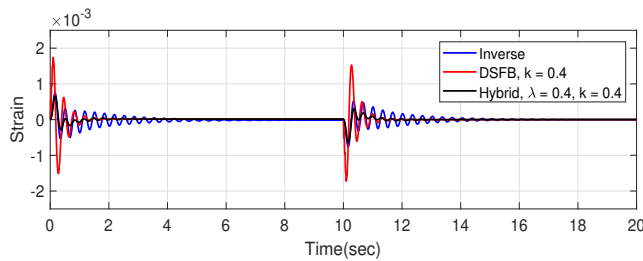
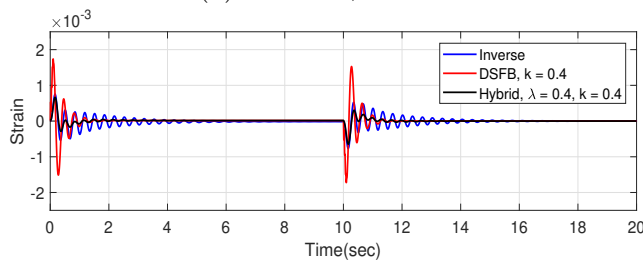


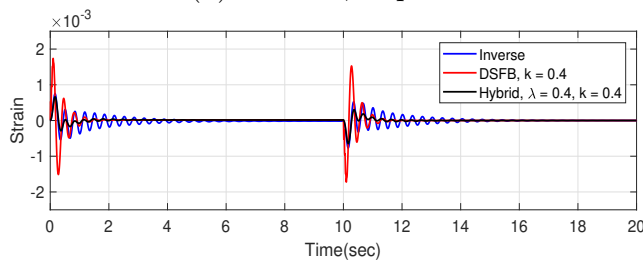
Figure 7: Simulation setup in MATLAB



(a) Link 1, torsion



(b) Link 1, in plane

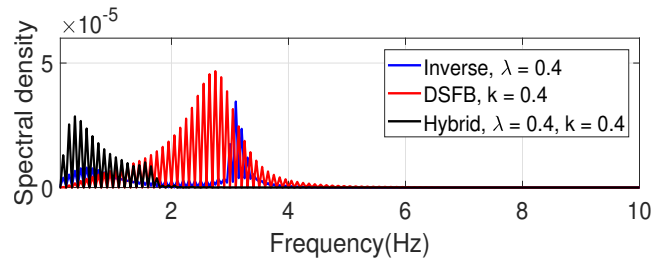


(c) Link 1, in plane

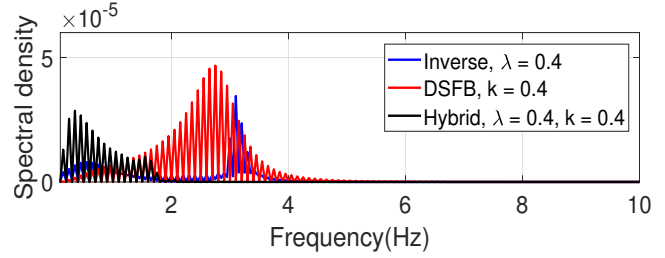
Figure 8: Strain information of the linear model

From the figures of torsional strain(Figure 8a), in plane strain for link 1(Figure 8b) and link 2(8c), it can be seen that the inverse controller had an upper hand in suppressing transient vibrations caused by the sudden starting and sudden stopping. However, these vibrations lasted for a relatively longer time. Interestingly, DSFB, though it had very poor transient response, it was very strong in dealing with residue vibrations. The hybrid of the two controller was inherently better in handling both the transient and the

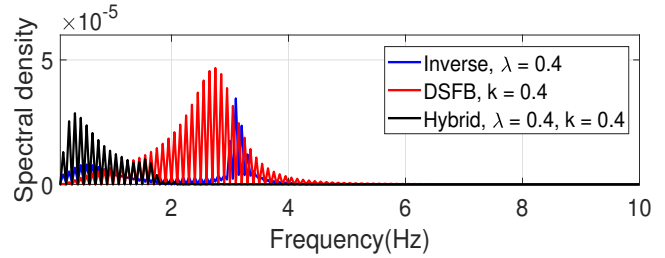
residue vibrations, hence, outperformed the individual controllers. Figure 9 gives a pictorial evidence of the of the comparison and the strengths of the individual controllers.



(a) Link 1, torsion



(b) Link 1, in plane



(c) Link 1, in plane

Figure 9: Strain power spectrum density

6.3 Experimental results

The experiment involved moving the three joints at an angle of 20 degrees, using a step signal lasting for 10 seconds, followed by returning to its original vertical position for another 10 seconds for a case without tip mass and with a tip mass of 100g. Strain measurement was achieved by attaching strain gauges at the root of respective links for torsional, link 1 in plane and link 2 in plane strain. Figure 10 shows the experimental setup of this work.



Figure 10: Experiment setup

Figure 11 shows the experimental results for the torsional strain, 11a, and strain for the two links, (11b, 11c) for a manipulator without any tip load. The figures show the comparison of the inverse controller, the DSFB controller and a hybrid of the two. It can be seen that the inverse controller had an upper hand in dealing with motion induced vibrations. On the other hand, the DSFB ensured a rapid decay of the residue vibrations. A combination of the two controllers as a two degree of freedom controller, yielded a system characterised by minimal motion-induced vibrations decaying very rapidly. The hybrid of the two yielded minimal strain amplitudes and shortened the duration of the vibrations. Actually, the overall performance of the hybrid was better than that of the controllers in their areas of strengths, individually. This is attributed to the combined strengths of the two controllers.

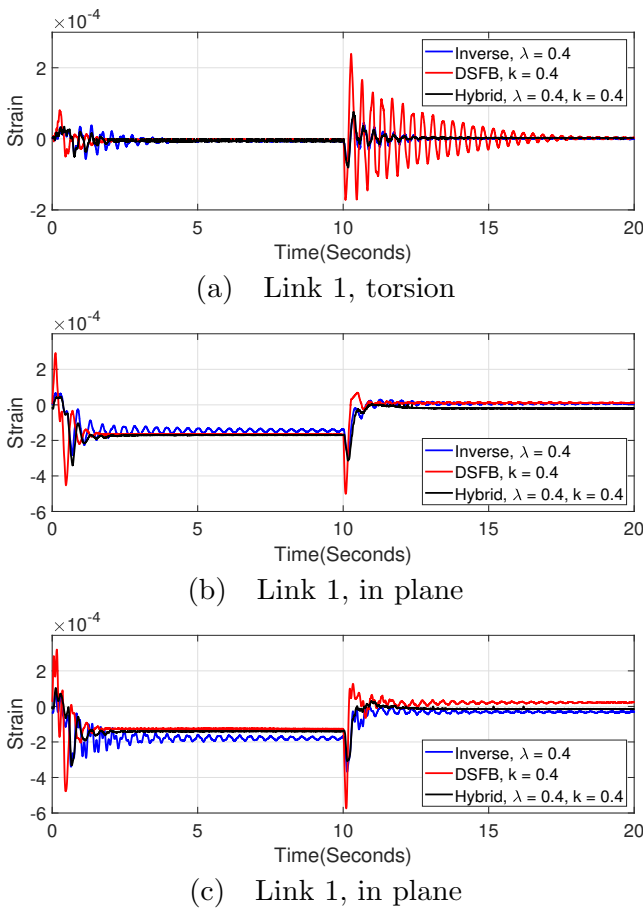


Figure 11: Torsional and lateral strain without any load

In plane strain in the first 10 seconds of the links 1 and 2(11b, 11c), an offset from zero strain can be seen. This error is associated with the fact that, during this time, links 1 and 2 are tilted and thus affected by gravity. Consequently, the bending strain at the root of the links did not converge to zero but remained to be a value of the distortion due to the self weight of the links for the entire period. However, this effect doesn't affect torsional strain.

Figure 12 shows the strain spectral power density without any load, where the improvement was very significant. It can be seen that individual controllers

suppress different frequencies while the hybrid inherits these capabilities, better yet outperforming the individual controllers. Again, the effect of self weight due to the tilted status of links 1 and 2 can be seen in the lower part of the spectrum (Figures 12b,12c), but it is absent in the spectrum for torsional strain (12a).

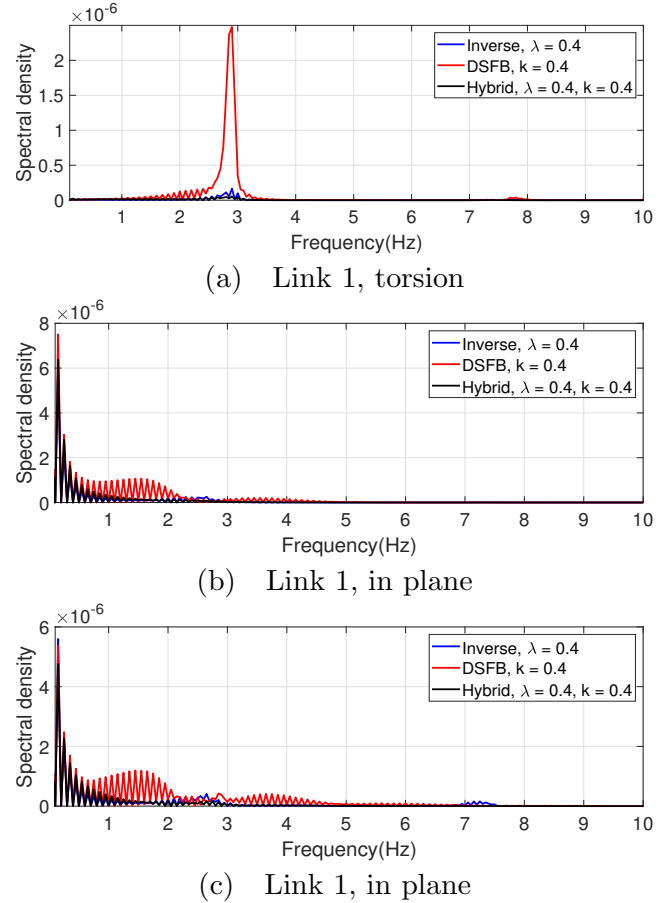


Figure 12: Strain spectral density without any load

With a load of 100g attached at the distal end of link 2, Figure 13 shows the torsional strain(13a) and in plane lateral strain for links 1 and 2(13b,13c). From the figure, the vibrations are a bit severe relative to those experience in a system without load. This is attributed to the fact that loading a flexible manipulator leads to excitation of more severe vibration at a relatively lower frequency when compared to a manipulator without any load. The offset due to self weight imposed by gravity during the first 10 seconds is also relatively higher.

The performance of the inverse controller is commendable in attenuating link vibration, however, complete mitigation is not possible and severe residues remains for the entire duration of operation. Strain feedback on the other hand, though its performance in eradicating the residues is very good, it is poor in dealing with trajectory induced vibrations. The hybrid of the two controllers, having inherited the complementary strengths of each controller is able to deal with both transient and residue vibrations.

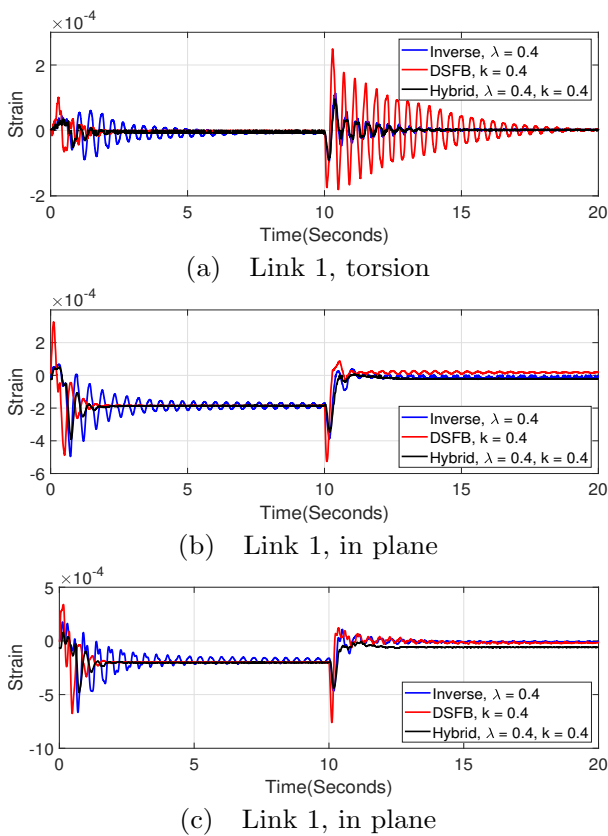


Figure 13: Torsional and lateral strain with a load of 100g

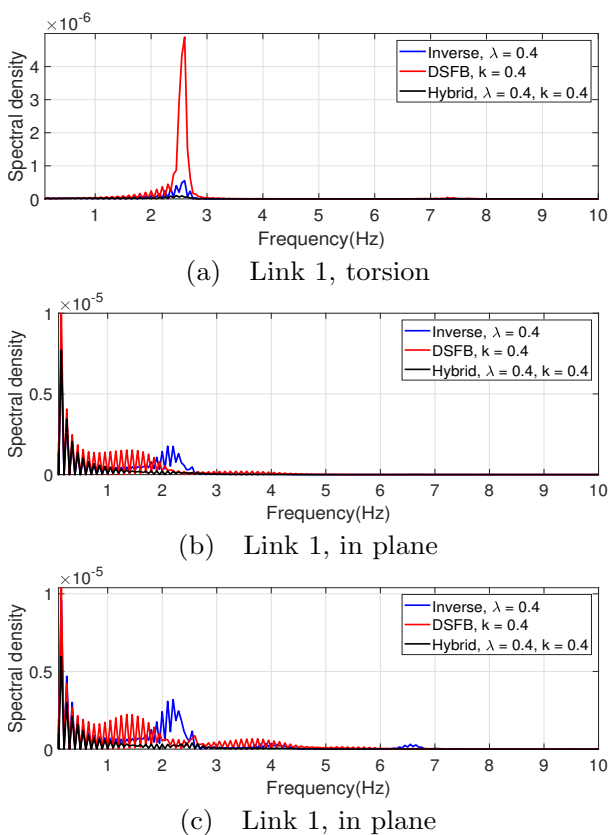


Figure 14: Strain spectral density with a load of 100g

To investigate the effect of loading on the strain spectral power density, Figure 14 illustrates the spectrum for torsional strain(14a) and in plane lateral strain(14b,

14c) for links 1 and 2 respectively. Comparing these frequency responses, we can observe that: 1). The peaks are higher than those seen in a system without any load, and 2). these peaks occurred at a slightly lower frequencies.

7 Conclusion

We successfully developed, linearized and validated a model of a 3D, two link, flexible manipulator. A stable right inverse of the linear model was developed, augmented with a low pass filter, and used as a feedforward controller. Together with a direct strain feedback controller, $k = 0.4$, in the feedback loop, the combination formed a two degree of freedom controller. From both simulations and experimental results, we found the inverse controller has an upper hand in handling motion induced vibrations which arose from sudden starting and stopping. Also, the strain feedback controller ensures rapid decay of the residue vibrations. Finally, the hybrid of the two controllers inherently has the strengths of the two controllers, exhibited a superior performance of suppressing both the transient and the residue vibrations.

References

- [1] Waweru Njeri, Minoru Sasaki, and Kojiro Matsushita, "Two-degree-of-freedom control of a multilink flexible manipulator using filtered inverse feedforward controller and strain feedback controller," in *2018 IEEE International Conference on Applied System Invention (ICASI)*, pp. 972–975, April 2018.
- [2] Y. Bian, Z. Gao, X. Lv, and M. Fan, "Theoretical and experimental study on vibration control of flexible manipulator based on internal resonance," *Journal of Vibration and Control*, vol. 0, no. 0, p. 1077546317704792, 2017. doi:10.1177/1077546317704792.
- [3] Y. Bian and Z. Gao, "Nonlinear vibration control for flexible manipulator using 1: 1 internal resonance absorber," *Journal of Low Frequency Noise, Vibration and Active Control*, vol. 0, no. 0, p. 1461348418765951, 2018. doi:10.1177/1461348418765951.
- [4] M. H. Shaheed and O. Tokhi, "Adaptive closed-loop control of a single-link flexible manipulator," *Journal of Vibration and Control*, vol. 19, no. 13, pp. 2068–2080, 2013. doi:10.1177/1077546312453066.
- [5] Y. Wang, Q. Zheng, H. Zhang, and L. Miao, "Adaptive control and predictive control for torsional vibration suppression in helicopter/engine system," *IEEE Access*, vol. 6, pp. 23896–23906, 2018.
- [6] B. Luo, H. Huang, J. Shan, and H. Nishimura, "Active vibration control of flexible manipulator using auto disturbance rejection and input shaping," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 228, no. 10, pp. 1909–1922, 2014. doi:10.1177/0954410013505951.
- [7] Z. Masoud, M. Nazzal, and K. Alhazza, "Multimode input shaping control of flexible robotic manipulators using frequency-modulation.," *Jordan Journal of Mechanical & Industrial Engineering*, vol. 10, no. 3, 2016.
- [8] Z. Mohamed and M. O. Tokhi, "Vibration control of a single-link flexible manipulator using command shaping techniques," *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 216, no. 2, pp. 191–210, 2002. doi:10.1243/0959651021541552.
- [9] L. Malgaca, ahin Yavuz, M. Akda, and H. Karaglle, "Residual vibration control of a single-link flexible curved manipulator," *Simulation Modelling Practice and Theory*, vol. 67, pp. 155 – 170, 2016.
- [10] . Yavuz, L. Malgaca, and H. Karaglle, "Vibration control of a single-link flexible composite manipulator," *Composite Structures*, vol. 140, pp. 684 – 691, 2016.

- [11] X. Zhang, J. K. Mills, and W. L. Cleghorn, "Experimental implementation on vibration mode control of a moving 3-prr flexible parallel manipulator with multiple pzt transducers," *Journal of Vibration and Control*, vol. 16, no. 13, pp. 2035–2054, 2010. doi:10.1177/1077546309339439.
- [12] S. O. Reza Moheimani, "A survey of recent innovations in vibration damping and control using shunted piezoelectric transducers," *IEEE Trans. Control Syst. Technol.*, pp. 482–494, 2003.
- [13] G. Song, S. P. Schmidt, and B. N. Agrawal, "Active vibration suppression of a flexible structure using smart material and a modular control patch," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, vol. 214, no. 4, pp. 217–229, 2000. doi:10.1243/0954410001532024.
- [14] Agrawal B. N. and Bang H., "Active vibration control of flexible space structures by using piezoelectric sensors and actuators," *In Proceedings of 14th Biennial ASME Conference*, Sept. 1993.
- [15] M. L. Kerr, S. Jayasuriya, and S. F. Asokanathan, "Qft based robust control of a single-link flexible manipulator," *Journal of Vibration and Control*, vol. 13, no. 1, pp. 3–27, 2007. doi:10.1177/1077546306064826.
- [16] D. Halim and S. O. R. Moheimani, "Spatial resonant control of flexible structures-application to a piezoelectric laminate beam," *IEEE Transactions on Control Systems Technology*, vol. 9, pp. 37–53, Jan 2001.
- [17] M. Sasaki, A. Asai, T. Shimizu, and S. Ito, "Self-tuning control of a two-link flexible manipulator using neural networks," in *2009 ICCAS-SICE*, pp. 2468–2473, Aug 2009.
- [18] Waweru Njeri, Minoru Sasaki, and Kojiro Matsushita, "Enhanced vibration control of a multilink flexible manipulator using filtered inverse controller," *ROBOMECH Journal*, vol. 5, p. 28, Nov. 2018.
- [19] J. Zhang, L. He, E. Wang, and R. Gao, "Active vibration control of flexible structures using piezoelectric materials," in *2009 International Conference on Advanced Computer Control*, pp. 540–545, Jan 2009.
- [20] F. Cao and J. Liu, "Vibration control for a rigid-flexible manipulator with full state constraints via barrier lyapunov function," *Journal of Sound and Vibration*, vol. 406, pp. 237 – 252, 2017.
- [21] D. Liberzon, A. S. Morse, and E. D. Sontag, "Output-input stability and minimum-phase nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 47, pp. 422–436, Mar 2002.
- [22] Z.-H. Luo, "Direct strain feedback control of flexible robot arms: new theoretical and experimental results," *IEEE Transactions on Automatic Control*, vol. 38, pp. 1610–1622, Nov 1993.
- [23] Z.-H. Luo and Y. Sakawa, "Gain adaptive direct strain feedback control of flexible robot arms," in *TENCON '93. Proceedings. Computer, Communication, Control and Power Engineering. 1993 IEEE Region 10 Conference on*, vol. 4, pp. 199–202 vol.4, Oct 1993.

A Appendix

A.1 Manipulator specifications

Table 1: Specifications of the flexible manipulator

Servo moter1	(Joint1)	Type	V850-012EL8
	Rated armature voltage	80	V
	Rated armature current	7.6	A
	Rated power	500	W
	Rated spindle speed	2500	rpm
	Rated torque	1.96	N.m
	Moment of inertia	6×10^{-4}	kg.m ²
	Mass	4.0	Kg
Servo moter2	(Joint2)	Type	T511-012EL8
	Rated armature voltage	75	V
	Rated armature current	2	A
	Rated power	100	W
	Rated spindle speed	3000	rpm
	Rated torque	0.34	N.m
	Moment of inertia	3.7×10^{-5}	kg.m ²
	Mass	0.95	Kg
Servo moter3	(Joint3)	Type	V404-012EL8
	Rated armature voltage	72	V
	Rated armature current	1	A
	Rated power	40	W
	Rated spindle speed	3000	rpm
	Rated torque	0.13	N.m
	Moment of inertia	8.4×10^{-6}	kg.m ²
	Mass	0.4	Kg
Encoder	Reduction ratio	1/100	P/R
	Spring constant	1.6×10^4	Nm/rad
Harmonic drive -joint1	Type	CSF-40-100-2A-R-SP	
	Reduction ratio	1/100	
	Spring constant	23	Nm/rad
	Moment of inertia	4.50×10^{-4}	kg.m ²
Harmonic drive -joint2	Type	CSF-17-100-2A-R-SP	
	Reduction ratio	1/100	
	Spring constant	1.6×10^{-4}	Nm/rad
	Moment of inertia	7.9×10^{-6}	kg.m ²
Harmonic drive -joint3	Type	CSF-14-100-2A-R-SP	
	Reduction ratio	1/100	
	Spring constant	7.1×10^{-5}	Nm/rad
	Moment of inertia	3.3×10^{-6}	kg.m ²
Link1	Material	Stainless steel	
	Length	0.44	m
	Radius	5×10^{-3}	m
Link2	Material	Aluminum	
	Length	0.44	m
	Radius	4×10^{-3}	m
Strain Gauge	Type	KGF-2-120-C1-23L1M2R	

A.2 Maplesim model

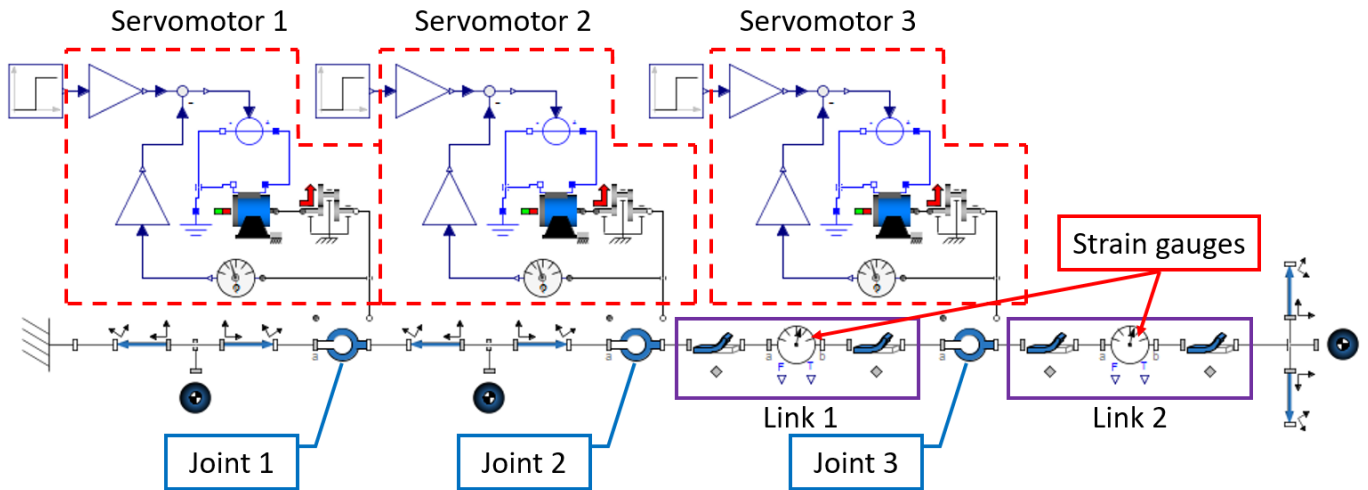
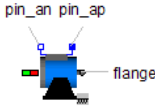

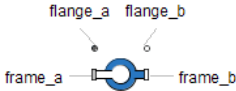

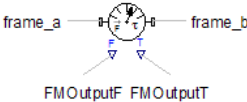
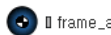
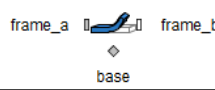
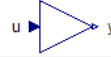




Figure 15: Maplesim model of the manipulator

Table 2: Building block of the arm in Maplesim

S.No.	Symbol	Component name	Short description.
1.		pmdc motor	Models a DC Machine with permanent magnets.
2.		Step	Generates a real step signal with variable height.
3.		Revolute	Joint allowing one rotational degree of freedom about a given axis.
4.		Angle Sensor	Measures the absolute flange angle.
5.		Force and Moment	Measures and outputs the forces and moments acting between two frames.
6.		Rigid body	Center of mass frame with associated mass and inertia matrix.
7.		Flexible Beam	A flexible beam with axial, lateral, and torsional deformations.
8.		Gain	Outputs the product of a gain value with the input signal.
9.		Lossy Gear	Gearbox with mesh efficiency and bearing friction. Represent the harmonic drive.
10.		Rigid Body Frame	Frame with a fixed displacement and orientation relative to a rigid body center of mass frame.

A.4 State, input, output and the transfer matrices of the internal dynamics of the inverse model

$$\hat{A}_\sigma = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -8877.8083 & 0 & -505.0084 & 0 & 6959.6426 & 0 & 963.5402 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -23691.8248 & 0 & -87127.7174 & 0 & 36338.2298 & 0 & 104574.9609 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 54312.6517 & 0 & 5955.6170 & 0 & -45939.3525 & 0 & -7980.4902 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 3910.2812 & 0 & 9021.0946 & 0 & -4191.1701 & 0 & -12659.5994 & 0 & 0 \end{bmatrix}$$

$$\hat{B}_\sigma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1.0225 & 0 & 0.2194 & 0 & 0 & 0 & 0.0057 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.1445 & 0 & 14.0596 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7.9491 & 0 & -1.7512 & 0 & 0 & 0 & -0.21900 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0158 & 0 & -1.2161 & 0 & 0 & 0 & 0 & 0 & 0.0035 & 0 & 0 & 0 \end{bmatrix}$$

$$\hat{C}_\sigma = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -9.6577 & -0.0048 & -0.0285 & 0 & 47.1468 & 0.0235 & 0.0364 & 0 & 0 \\ -0.1422 & 0 & 40.4355 & 0.0104 & 0.2636 & 0 & -5.6792 & -0.0014 & 0 \end{bmatrix}$$

$$\hat{D}_\sigma = \begin{bmatrix} 0 & -1 & 0 & 0 & -0.8887 & 0 & 0 & -0.0016 & 0 & 0 & 0 & 0 \\ -0.9999 & 0 & 0 & -0.7935 & 0 & 0 & -0.0007 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.9998 & 0 & 0 & -0.7660 & 0 & 0 & -0.0006 & 0 & 0 & 0 \end{bmatrix}$$

Adaptation of Electronic Book Publishing Technology by The Publishers in Southeast Nigeria

Godson Emeka Ani*, Chike Ogbob

Department of Printing Technology, Institute of Management and Technology, Enugu

ARTICLE INFO

Article history:

Received: 25 September, 2018

Accepted: 23 November, 2018

Online: 13 December, 2018

Keywords:

Electronic publishing

E-book technology

Digital technology

New technology

ABSTRACT

This paper investigates the electronic book production and distribution practices adopted by the Publishers in Southeast Nigeria because of the global growing interest in writing new book titles in digital form as well as converting paper titles to digital content. To investigate this study, the researcher formulated three research questions to guide the study. A number of relevant literatures were reviewed to establish the gap anchored on the Technological Determinism Theory. Descriptive Survey design and In-depth interviews were used to elicit data for the study. Fifteen publishing firms were randomly selected from Enugu (Enugu State) and Onitsha (Anambra States). Data from the survey were analyzed using the Statistical Package for Social Sciences (SPSS). The key findings of the study include that: the rate of adaptation of e-book technology by educational publishers in the southeast Nigeria is relatively low. However, books are now distributed and sold in digital forms as online books and print on demand titles by publishers in southeast Nigeria through the Internet; Lack of adequate regulations for the protection of intellectual resources, high cost of digital equipment, incessant power failure and high cost of using alternative power supply for e-book distribution, are some of the challenges undermining the adaptation of e-book publishing technology in Southeast Nigeria. It was recommended that the government should provide an enabling environment through dependable infrastructure for the promotion, distribution and marketing of e-books in the Southeast Nigeria. The Book Publishers Association of Nigeria should also make commitment to the development of indigenous book market.

1. Introduction

The advances in Information Communication Technologies have structurally changed the way books are produced, distributed and sold. Book publishers all over the world are diversifying into electronic publishing – digital publication of e-book, digital magazines, and the development of digital libraries and catalogues. Strongly associated with electronic publishing are many non-network electronic publications such as encyclopedias on CD and DVD as well as technical and reference publications relied on by mobile users. The entry of core technology companies into the publishing space and the introduction of portable electronic reading devices have made digital books gain wider interest. Digital technology has changed the face of all media, including print. In order to survive, all parties involved must adapt and focus on the consumers' changing needs [1]. Publishers in Nigeria who fail to digitalize their future/current book titles may lose out in the emerging publishing market [2].

The print products (print media) can be used by the customers directly (the information is available on paper), whereas electronic media (e.g. CD-ROM or network as a medium) requires special equipment for its visualization (monitor, computer, connection to the network).

Electronic books (e-books) are book titles that are available online. They can be read as e-mail, retrieved by a portable electronic reading device or as a file and can be downloaded on a computer [3]. The most popular method of getting an e-book is to purchase a downloaded file of e-Book from a website, such as Barnes and Noble or through on-line bookshops, such as Amazon.com. Electronic book publishing is available in Nigeria through Bookstores, Amazon, Lulu, Smashword, Createspace, Okadabooks, Nook and African Books Collective (ABC) for the print-on-demand titles.

Roger fidler asserts that "When new forms of communication media emerge, the older forms usually do not die - they

*Corresponding Author: Godson Emeka Ani, Email: godsonani@imt.edu.ng

www.astesj.com

<https://dx.doi.org/10.25046/aj030650>

continue to evolve and adapt. In this way, the different media compete for the public's attention and jockey for positions of dominance but no media has yet disappeared. It is obvious that each medium contributes to the development of its successors"[4].

E-book publishing is enabling new authors to release books that would be unlikely to be profitable for traditional publishers. It makes a wider range of books available, including books that would not have been available in book retailers' shop due to insufficient demand. This technology also responds quickly to changing market demand, reduces production and distribution costs and gets more books to readers faster than the current publishing business model. This fundamental shift in the media scope is driving on how consumers engage with brands and how marketers have to go where the customers are, in order to communicate effectively and efficiently [1]. There are indications that electronic publishing is becoming a major mode of publishing in many countries [5].

This paper investigates the electronic book production and distribution practices adopted by the Publishers in Southeast Nigeria because of the global growing interest in writing new book titles in digital form as well as converting paper titles to digital content.

1.1. Problem Statement

The emergence of new communication technology shrinks the world into a global village and made authors to market their digital contents directly to their consumers. The business of book publishing has advanced from hand setting of movable metal type to Linotype setting, Monotype setting, Lithographic offset printing, Word processing, Desktop publishing, online books and now portable electronic books [2]. But today, print channel is limited to a certain field of applications.

The books being produced by the indigenous publishers could not compare with the established Publishing Houses abroad and many publishers are no longer in the business. Publishing houses have closed down because there was no capital to continue, while the rest have been struggling to survive [6]. It has not been easy for many indigenous publishers in recent years.

The computer literacy level in the society is low and majority of the populace do not find any comfort in reading from a computer screen. The high cost of personal computer in the country has also caused a considerable low computer ownership density and limited access to Internet facilities.

It is really worrisome to imagine how the publishers in the Southeast Nigeria could survive the changing global publishing business if they fail to adapt e-book technology now. The need to find answers to this worry is what informed this study.

1.2. Justification

It is expected that this study will help the publishers to integrate electronic book (e-book) publishing technology in their present

and future book titles in order to be relevant in the emerging publishing market. It is also expected that the rapidity in the publishing industry will create opportunities for the publishers in the global competition.

1.3. Objectives

The general objective is to assess the adaptation of electronic book publishing technology by the publishers in Southeast Nigeria.

The specific objectives are as listed below:

- To determine the extent in which e-book technology is adopted by educational publishers in Southeast, Nigeria.
- To examine the benefits accrued by educational publishers who adopt e-book technology in Southeast, Nigeria.
- To ascertain the challenges posed by the adaptation of e-book technology in the publishing business in Southeast, Nigeria.

1.4. Research Questions

- To what extent has Digital Book Technology been adopted by Educational Publishers in Southeast Nigeria?
- What are the benefits of the adaptation of digital book technology to educational publishers in Southeast Nigeria?
- What are the challenges posed by the adaptation of e-book technology in the publishing business in Southeast Nigeria?

1.5. Scope of the Study

The scope of this study was narrowed down to the Educational Publishers in Onitsha, Anambra State and Enugu, Enugu State, both in Southeast zone Nigeria. Electronic book publishing technology as operationalized in this study refers to books distributed by any form of electronic device, including books on CD-ROM, Books in Cassette, Paper books with CD-ROM, Audio CD Books and Print-on-Demand.

1.6. Limitations of the Study

The limitations of this study include the poor attitude of some of the educational publishers to research work. They were reluctant to accept appointment for oral interviews but would rather prefer to complete the copy of questionnaire.

The research also observed that a reasonable number of the Educational Publishers in the Southeast Nigeria print and market only the Traditional Hardcopy Book Titles, unlike what happens in Lagos and Ibadan (Southwest Nigeria).

2. Literature Review

Technology plays a dominant role in changing almost all facets of the publishing industry, from the writing to the distribution of books to final consumers. Even the process of distributing books to the final consumers has been made easier with electronic mail, inventory management programs and other tools. According to

former Times Mirror group Vice-President, Jerome Rubin, whereas Gutenberg's moveable type enabled such mass distribution of identical copies upon its implementation in the fifteenth century, the twentieth century's "electronic technologies allow the creation of infinite variations" [7].

Digital communication technologies and processes for designing and producing information as print media or electronic media are penetrating the market. This means that the production processes for both groups of media are closely interconnected by a common basis, the pre-media production section. The pre-media area prepares the content and produces a digital document that may, to a large extent, be used in both sectors of production; "cross-media publishing" has become possible and is in place in the printing, publishing and communication industries [8].

Basically, new technology has enabled the consumer to assume some of the traditional publisher's role in the choice and format of information, leaving the consumer with more control [7]. The Transfer of Information, documents and content via electronic media (CD-ROM or Internet) creates a varied array of interesting and useful applications. Electronic media can provide innovative alternatives to print media. In contrast to the simple, flexible handling of the common book, however, digital playback of reading material can be relatively complicated and awkward [8]. Suitable software tools and interactive user interfaces will enable the text passage to be marked with circling, highlighting or page marking. These are measures that will improve the acceptance of electronic books. Several websites such as www.fictionwise.com, www.gutenberg.org and www.memoware.com, offer e-books specifically for PDAs and cell phones. Existing models such as the Rocket e-book (NuvoMedia), Softbook (Softbook Press) and EB Study Model (Everybook) have made it easier for readers to peruse an e-book. Additionally, by mastering a few key functions, the user can operate this e-book in a variety of places, thus making network-independent reading simple.

Convergence is altering almost all aspects of the book industry. Most obviously, the Internet is changing the way books are distributed and sold. But this new technology, in the form of e-publishing, the publication of books initially or exclusively online, offers a new way for writers' ideas to be published. E-publishing can take the form of d-books and print on demand (POD) [9]. Digital books content could also be selected and downloaded from the catalogue of books in Publishers' Database and instantly printed as Print-on-Demand (POD) titles in a bookstore that has the appropriate technology. The printed and bound copies are produced on demand with the possibility of buying chapter-by-chapter. Examples of POD Publishers are Xlibris, Authors House, and Toby Press.

E-books have potential in enhancing distance education. In particular, e-books are able to enhance the interaction between educators and students when dealing with teaching and learning

materials [3]. Computers and computerized editorial systems are now used to write texts, process images and produce books.

The Internet and World Wide Web (www) are facilitating the publishing of books in the areas of typesetting, layout and editing, on-line distribution, on-line ordering, marketing, advertising, pricing, payments and hiring [2, 10, 11, 12]. Most digital publishers now provide a full range of services such as copy editing, online publishing, securing or commissioning artwork, jacket design, promotion, and in some cases, even hard copy distribution to brick-and-mortar book stores, based on a variable royalty or fee arrangement.

The content of e-book can be delivered to multiple platforms and operating systems to reach broad audience at a reduced cost. The use of CD-ROM and Internet publishing has circumvented the constraints of paper to a reasonable extent if extensively adopted.

The advantages of e-publishing for readers are in time and money. D-books are down loaded at a very low cost as compared with the cost of the hardcopy and an electronic bookstore never closes. No matter what time of the day or night, readers can download their textbooks and begin reading immediately. Print-on-Demand (POD) reduces production and distribution cost, by getting books faster and cheaper to readers, than the current publishing business model.

The physical form of books is changing using this new technology - many of today's books are no longer composed of paper pages snuck between two covers [13]. The Internet is changing the way books are The Internet made it possible for an author to write a book, puts it on a website and offers it for sale, without the aid of a traditional publisher or retailer. An author could as well place a manuscript on a website run by a company that collects and distributes e-books. Distributed and sold

The challenges of book distribution can be remedied by making books more accessible through alternative publishing models; electronic publishing and through print-on-demand technologies [2]. He concluded by saying that e-publishing in Nigeria is simply a novelty and the appropriate reading habit a mirage. Digital content is ever-harder to control. In one sense, everything is Open-access. Even though some publishers are fighting the pirated titles with law suits and authentication technology, others embrace open access [14].

Digital books are now distributed over several delivery channels, primarily a combination of print and Internet. The Internet helps larger media organisation to have more publication channels. It has opened new frontiers for expansion, taste and distribution of print information thus, sustaining industrialization in Nigeria. Books could also be made more accessible through co-publishing; a model that allows publishers from different parts of the world to publish a book jointly so that a title is made available to different markets by publishers in the respective territories [15].

Some authors now take advantage of e-publishing to publish their e-book on the Internet by themselves [2]. Authors who publish in

their own websites keep 100 percent of the income while authors who distribute their works through an established e-publisher. Additionally, he said that they usually get royalties of 40 percent to 70 percent, compared to the 5 percent to 10 percent offered by traditional publishers.

Generally underdeveloped market, a weak developed reading culture, lack of distribution hubs such as commercial outlets and short print runs are part of the challenges facing book publishers today [16]. The publication of e-book is technology driven, cheaper, and there is no distribution or warehousing cost.

2.1. Theoretical Framework

The study was anchored on Technological Determinism Theory whereby technology is viewed as a driver of social change. It explains that when new systems of technology are developed, society will immediately change and adapts to that technology. Hence, “the medium is the message”.

Media do not only extend our reach and increase our efficiency, they act as filter to organize and interpret our social existence [17]. The message that is produced on the traditional print media on paper formats is unique in many ways with the same message that is produced on the electronic board of a computer. The innovation of electronic book publishing technology opens a new frontier for production, expansion, taste and distribution of books in Nigeria.

A change in technology has a multiplier effect on other factors. Hence, the phrase, “we shape our tools and they in turn shape us” explains how one aspect of change affects all others. The digitalization and convergence of computer technology have greatly influenced the way information is produced and disseminated in the Book Publishing Industry. A consumer that obtained gratification in reading various book publications online, as against reading only the traditional paper books will not hesitate to adopt the e-books. Similarly, a publisher is likely to be interested in the direction of the market force and the profit that will be realized from every venture. He is likely to adopt the new technology as a result of the market shift to satisfy customers’ demand and taste. The electronic book technology (*shaping of our tools*) compels educational publishers to adjust and reposition themselves in order to use the modern tools (*they in turn shape us*).

The application of Technological Determinism Theory in this study is justified since the change in technology made book publishers to acquire more sophisticated skills that help print information to appreciate in value.

3. Methodology

The survey research method and semi structured interviews were used to gather data from practicing publishers in Enugu, Enugu State and Onitsha in Anambra State, Nigeria, where about 75 percent of all the publishing houses in the Southeast zone, Nigeria are based.

Fifteen (15) copies of questionnaire were distributed to practicing educational book publishers in Onitsha, Anambra State, and Enugu in Enugu State, Nigeria. The questionnaire consists of thirteen (13) close-ended questions. Fifteen (15) copies of questionnaire were successfully completed; giving a return rate of 100 percent. In order to validate the data from the survey, semi structured interviews were conducted with two publishers based in Enugu and Onitsha respectively. Notes from the in-depth interviews were transcribed and analyzed using hermeneutic interpretation. Data from the survey was analyzed using the Statistical Package for Social Sciences (SPSS).

The following publishers were randomly selected from the list of printing and publishing companies in Nigeria [18] and Enugu state [19]:

- (i) DeRafelo Ltd, Enugu; (ii) Fourth Dimension Publishing Company Ltd, Enugu; (iii) Rocana Enugu; (iv) New Generation Educare Ltd, Enugu; (v) Kongrat Press, Enugu; (vi) Executive Press Resources Ltd, Enugu; (vii) Jibalo Publishers, Enugu; (viii) Rhyce Kerex Publishers, Enugu; (ix) Keny and Brothers Ventures, Enugu; (x) Dulacs Publishers, Enugu; (xi) K.C.Graphics and Design, Enugu; (xii) Brand Wise Advertising/Publishing Ltd, Enugu; (xiii) Pacific Publishers, Onitsha; (xiv) University Publishing Company Ltd, Onitsha; (xv)Diamond-Pen Publishing Company Ltd, Onitsha.

3.1. Data Presentation

The data in table 1 shows that 3 publishers or 20 percent of their sampled publishers engage in book publishing only while 12 publishers or 80 percent combine printing and publishing. The above data shows that the majority of the respondents were educational book publishers.

Table 1: Scope of the Publisher

S/N	Specialty	Respondents/Publishers	Percentage
(a)	Publishing only	3	20
(b)	Printing & Publishing	12	80

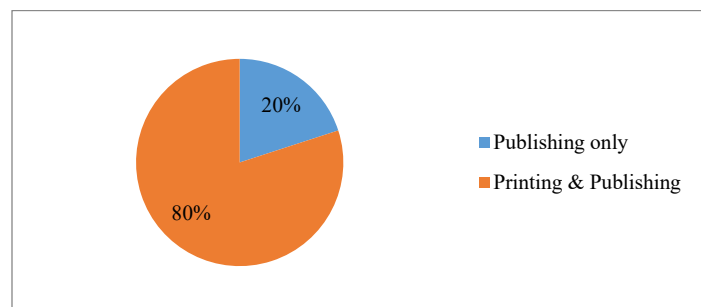


Table 2 shows that 7 publishers, representing 47 percent of the sampled book publisher either produce and market their traditional books with CD-ROM, 11 publishers or 73 percent have digitized their books and market them online as print-on-demand

titles 7 publishers or 47 percent market their books as CD-ROM only, while 2 publishers, representing 13 percent of the respondents, market their e-books in Cassette and CD. The analyzed data above indicated that majority of the publishers adapt the Print-on-demand technology for e-book publishing.

Table 2: Extent of adoption of digital book technology (available variety)

Format		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Audio CD Books	1	20.0	20.0	20.0
	Books in Cassette	1	20.0	20.0	40.0
	Books in CD-ROM	1	20.0	20.0	60.0
	Paper Books with CD-ROM	1	20.0	20.0	80.0
	Prints-on-demand titles	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

The data in table 3 above shows that all the sampled publishers agree that new technologies have brought impressive improvement in the quality of their colour design, while 14 publishers representing 93 percent indicated that new technologies also improved the quality of book content. 10 publishers, representing 67 percent admitted an improvement in inside illustrations, 8 publishers or 53 percent admitted improvement in layout, 7 publishers or 47 percent indicated improvement on cover lamination, binding and finishing. The analyzed data indicated that the quality of the cover design and content of books published today are better because of the adaptation of digital technology.

Table 3: Areas of improved quality on output due to newly adopted technology

Format		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Binding and Finishing	1	16.7	16.7	16.7
	Cover Design	1	16.7	16.7	33.3
	Cover Lamination	1	16.7	16.7	50.0
	Inside Illustrations	1	16.7	16.7	66.7
	Layout	1	16.7	16.7	83.3
	Quality of Book Content	1	16.7	16.7	100.0
	Total	6	100.0	100.0	

Statistics		
Percentage		
N	Valid	6
	Missing	0
Mean		67.83
Median		60.00
Std. Deviation		23.481
Variance		551.367

The data in table 4 above shows that 13 publishers, representing 87 percent of the entire population admit that new technologies improved the prepress in the areas of reformatting, editing and

easy input of information. 11 publishers or 73 percent admitted that there is improvement in redesigning and printing, while 9 publishers or 60 percent and 8 publishers or 53 percent identified improvement in the rewriting and spell-checking of typeset matters. The analyzed data indicated that electronic editing has made it easy for publishers to input, reformat, edit, reformat and print books.

Table 4: Aspect of improvement in prepress due to the adaptation of Digital Technology

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Easy to Input	1	20.0	20.0	20.0
	Redesign & Print	1	20.0	20.0	40.0
	Reform & Edit	1	20.0	20.0	60.0
	Rewrite	1	20.0	20.0	80.0
	Spell-check	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

Statistics		
Percentage		
N	Valid	5
	Missing	0
Mean		70.60
Median		73.00
Std. Deviation		14.011
Variance		196.300

Table 5: Skill development strategies adopted for e-book publishing.

Training		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		2	40.0	40.0	40.0
	In-service Course	1	20.0	20.0	60.0
	On-the-Job Training	1	20.0	20.0	80.0
	Short-service Course	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

Table 5 shows that 10 publishers, representing 67 percent train staff on digital publishing through the in-service-course, 11 publishers, representing 73 percent train staff on-the-job, while 3 publishers, representing 20 percent train the staff through the short-service-course. The analyzed data indicated that the publishers adopted “In-service-Training” strategies to embrace e-book technology.

Table 6 shows that 15 respondents, representing 100 percent are of the opinion that Incessant power failure and fluctuation of electricity is a factor undermining the adoption of e-book technology in the Southeast Nigeria, while 12 respondents, representing 80 percent of the sampled publishers indicated that cyber copyright and piracy are militating against the flourishing of electronic commerce in Southeast Nigeria. 11 respondents, representing 73 percent opined that they encountered the challenge of learning a new skill (computer technology) in the

adaptation of e-book publishing technology. Eight (8) respondents, representing 53 percent admitted that the adaptation of new technologies has rendered some publishing equipment redundant, while 6 publishers agreed that staff were retrenched and most of the publishing organizational settings were restructured as a result of the adaptation of e-book technologies. The analyzed data indicated that incessant power failure, cyber copyright and piracy constitute a major obstacle to the adaptation of e-book technology.

Table 6: Challenges encountered in adaptation of e-book publishing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cyber Copyright and Piracy	1	16.7	16.7	16.7
	Difficulty of Learning New Skill	1	16.7	16.7	33.3
	Fluctuation of Electricity	1	16.7	16.7	50.0
	Redundant Publishing Equipment	1	16.7	16.7	66.7
	Restructured Organizational Setting	1	16.7	16.7	83.3
	Retrenchment of Staff	1	16.7	16.7	100.0
	Total	6	100.0	100.0	

Statistics Percentage		
N	Valid	6
	Missing	0
Mean	64.33	
Median	63.00	
Std. Deviation	24.105	
Variance	581.067	

4. Discussions/ Analysis:

4.1. Research Question One: To what extent has Digital Book Technology been adopted by Educational Publishers in Southeast Nigeria?

The survey shows that educational publishers in Southeast Nigeria adopted more of the print-on-demand titles (73%) than other variety of e-book products on Table 3, and closely followed by the adoption of CD-ROM attachment to the traditional book (47%).

4.2. Research Question Two: What are the benefits of the adaptation of digital book technology to educational publishers in Southeast Nigeria?

The adoption of digital books has significantly contributed to the improvement in the quality of published books in the areas of cover design (100%), book content (93%), inside illustrations (67%), and layout (53%), as indicated in table 4.

The adoption of new communication technologies enabled publishers in Southeast to reap the benefits of digital book reformat and editing (87%), ease of input (87%), redesign and printing (73%), as indicated on table five. The conduit that is often used for e-book publishing in the Southeast is e-mail address (87%), while conduits such as website and other internet facilities or access are not yet prevalent (table 6). The survey also shows that the marketing and distribution of e-book is chiefly through “e-publishing partnership (73%). In order to ensure the preparedness of the local workforce for the shift in technology, the type of the training adopted for the purpose of e-book technology was investigated. The survey shows that the most prominent skill development strategies adopted for e-book publishing in the Southeast is “on-the-job training (73%) and in-service course (67%).

4.3. Research Question Three: What are the Challenges posed by the adaptation of e-book technology in the publishing business in Southeast Nigeria?

The survey shows that there is no security for the protection of intellectual property that is in digital form such as e-book (80%), and copies are distributed over the internet or via pirated CD-ROMs at a very low cost. The survey collaborated the opinion of Mr. James Opara’s view that the rate of piracy is discouraging a lot of publishers from publishing their e-books in the southeast Nigeria, since piracy has adverse effect on job turnover. The incessant power failure/fluctuation of electricity is a major hindrance to the adaptation of e-book technology in the Southeast Nigeria (100%). The e-book technology involves the learning of new skill (computer technology), and some of the publishers sampled indicated that it constituted a great challenge (73%). Some analogue publishing equipment and staff were rendered redundant due to the adaptation of e-book technology by educational publishers (table 9).

5. Conclusion and Recommendations

5.1. Conclusion

The study shows, in empirical term, that the rate of adaptation of e-book technology by educational publishers in the southeast Nigeria is relatively low. However, books are now distributed and sold in digital forms as online books and print on demand titles by publishers in southeast Nigeria through the Internet. E-book technology has greatly improved book publishing in the area of pre-press; formatting, layout, editing, spell checking and design.

This agrees with the suggestion [1] that this fundamental shift in the media scope is driving on how consumers engage with brands and how marketers have to go where the customers are, in order to communicate effectively and efficiently. It also agrees with the argument [20] that the book publishing industry will be profoundly affected by the shift to digital. We further discovered that educational publishers who adopted e-book technology in the southeast Nigeria used the e-publishing technology to release books that consumers would not have been able to find in standard

retailers' bookshops due to insufficient demand for the traditional 'print-run'.

This attribute was characterized by underdeveloped market and a weak developed reading culture [16]. Accordingly, some authors now take advantage of e-book to publish their work on the Internet. This innovation has empirically opened new frontier for production and expansion via distribution of books in Nigeria.

E-publishing has responded to the changing market demand by making a wide range of books available and enabling the traditional book publishers digitize their current book titles. However, Lack of adequate regulations for the protection of intellectual resources, high cost of digital equipment, incessant power failure and high cost of using alternative power supply for e-book distribution, are some of the challenges undermining the adaptation of e-book publishing technology in Southeast Nigeria.

Despite these challenges, e-book distribution and sales are now accessible through alternative publishing model: electronic publishing and through print-on-demand technologies [2]. Generally, the adaptation of e-book technology is gaining reasonable ground among educational publishers in Southeast Nigeria.

5.2. Recommendations

The government should provide an enabling environment through dependable infrastructure for the publication and marketing of e-books in the Southeast Nigeria.

Efforts must be intensified to ensure that the publishing houses in Nigeria are comparable with their counterparts in developed countries in order to avoid print flight. Indigenous book publishers should go beyond the appreciation of imported printed items and strategies on the possibility of replicating the same in Southeast Nigeria.

Policy makers should put relevant law in place that will discourage pirates from exploiting the intellectual resources of electronic book publishers.

Institutions and corporate organizations should be encouraged to have their e-libraries and e-learning platforms so as to encourage e-book technology and electronic book consumers to be able to read at their own pace anytime and anywhere.

There should be relevant regulatory bodies in e-publishing and traditional book publishing business that will monitor the standard of published books and ensure that it meets the approved standard. This body should observe ethics, rules and regulations for engaging in the publishing industry.

The Federal and State Governments should give financial assistance to educational Book publishing industry, to enable them to transit from the traditional book publishing into electronic book publishing. They should also enhance power generation to the industry.

The Book Publishers Association of Nigeria should make serious commitment to develop the indigenous book market.

Acknowledgements

The authors gratefully acknowledge the Tertiary Education Trust Fund (TETFund), Nigeria for the funding of the research project from which this paper is drawn. The views expressed in the paper are those of the authors, and neither the TETFund nor the authors' institutional affiliation (IMT, Enugu) bears responsibility for them.

References

- [1] G. Stamp, T. Hodson, The need for transformation within the UK printing sector. London: Dotgain.org white paper, 2010.
- [2] E. Ifeduba, "Digital publishing in Nigeria: evidence of adoption: evidence of adoption and implications for sustainable development in journal of research, in journal of research in national development". Vol.8, No.1 pp.1-8, June, 2010.
- [3] N. Shiratuddin, M. Landon, F. Gibb, S. Hassan, "E-book technology and its potential- applications in distance education", Texas Digital Library, Vol. 3, No 4, 2003.
- [4] S. Biagi, Media Impact: An Introduction to Mass Media, Canada: Thomson Wadsworth, 2003.
- [5] M. Tiamiyu, Prospects of Nigerian Book Publishing in the Electronic Age in Adesanoye and Ojeniyi (eds), Issues in Book Publishing in Nigeria, Ibadan: Heinemann Educational Books. pp.143-157, 2005.
- [6] A. Alhassan, Our Challenges, By Indigenous Publishers, Daily Trust, Saturday, April 14, 2013, 2013.
- [7] H. Tetkeh, "Evolution of the book publishing Industry: structural changes and strategic implication in journal of management history" vol.4 No.2, pp104 – 123, 1998.
- [8] H. Kipphen, Handbook of print media, Heidelberg Germany. SPIN: 10764981, 2001.
- [9] S. J. Baran, Introduction to Mass Communication: Media Literacy and Culture: Fifth Edition. Boston: McGraw-Hill, 2009.
- [10] G. O. Abegunde, Quality Book Production in Nigeria, in Adejuwon (ed.) Quality Book Production, Ibadan: Codat publications, 2003.
- [11] C. O. Adejuwon, Quality in Book production in Adejuwon (ed.) Quality Book production. Ibadan: Codat Publications, 2003.
- [12] J. J. Iwu, Problems of the book publishing industry in Nigeria: the Onibonje publishers experience after 50 years. PNLQ quarterly, 75 (3). Pp. 105-111, 2011.
- [13] S. J. Baran, Introduction to Mass Communication: Media Literacy and Culture: Second Edition. Boston: McGraw-Hill, 2002.
- [14] R. Schonfeld, Book Publishing: University Presses Adapt, in Springer Nature (Online): Macmillan Publishers Limited, 2017.
- [15] A. Veglis, "Cross-media publishing by U.S. Newspapers, in Journal of Electronic Publishing". Vol.10, Issue 2, 2007. Doi: <http://dx.doi.org/10.3998/3336451.0010.211>
- [16] S. Ngbeni, Scholarly publishing: The Challenges Facing the African University Press, The Netherlands. African Studies Centre, 2012.
- [17] H. E. Ikpe, S. S. Ibekwe, "The Print Media and the Consumer in the Cyber", 2007.
- [18] "Nigeria Printing and Publishing Companies", Finelib.com
- [19] "List of Printing and Publishing Companies in Enugu State", Galleria Media Limited.
- [20] J. R. Dominick, "The dynamics of mass communication media in the digital age", 7th Edition. New York, McGraw-Hill, 2002.

An Integrated & Secure System for Wearable Devices

Callum Owen-Bridge¹, Stewart Blakeway^{1,2}, Emanuele Lindo Secco^{1,*}

¹Robotic Laboratory, Department of Mathematics and Computer Science, Liverpool Hope University

²Faculty of Arts, Science and Technology, Wrexham Glyndŵr University

ARTICLE INFO

Article history:

Received: 13 September, 2018

Accepted: 27 November, 2018

Online: 19 December, 2018

Keywords:

E-Health

Wearable sensors

Human Health Monitoring

ABSTRACT

Health services are under increasing pressure to reduce large waiting times for appointments. The delayed diagnosis of a human illnesses could have profound consequences for the patient. The large waiting times may be attributed to lack of data, the accuracy of the data or the timeliness of the availability of the data. In addition to this people often priorities their busy lives over their health. Other factors that affect delayed diagnosis of patients could be attributed to ill-health and the difficulty in the patient visiting the health-care practitioner, this is more evident in elderly patients that may have suffered a deterioration of their health. To overcome these obstacles eHealth systems could be put in place to facilitate the transmission of real-time and accurate data regarding a patient's health directly to secure database that can be accessed by authorized practitioners. This paper outlines the use of eHealth (electronic health) to improve health care services which makes use of a wearable set of secure sensors and a secure database that is accessible using a web application. The proposed wearable system records and stores encrypted data which is related to the body's physical parameters. Then, the encrypted data are transmitted using wireless technologies from the wearable device to a secure relational multi-user database using a protected protocol. The data can be accessed by the patient and professionals, such as general practitioners, using a web interface once authentication has taken place. This secure wearable architecture alleviates the problems of the lack of data, the inaccuracy of data and the timeliness of data by recording vital body parameters throughout the day and by sending real-time live data to a system that can be immediately accessed by the practitioner. The system also allows for the automatic analysis of data and presentation of the data using graphs which could aid the practitioner in determining patterns in the patients' health statistics.

1. Introduction

Health care services are struggling to diagnose people accurately and quickly, due to lack of data. Humans are becoming less proactive and unaware of their own health. Many people struggle to allocate time to see a health care professional due to their lifestyle and/or physical health.

This research presents a secure medical wearable device for patients with health issues, to help support in the monitoring of physical parameters in the daily life and in the Ambient Assisted Living (AAL) context. This paper focuses on the creation and implementation of a hardware device with sensors that captures

patients health data and the development of secure software for data storage which can be accessed remotely by multiple authorised users.

The research undertaken details how the hardware was implemented, the software used to control the device, how data is secured before being transmitted over unsecured wireless media to a secure multiple-access relational database. This papers also discusses eHealth, health of the population, security, privacy and legal issues surrounding medical wearable devices. The result of this research shows that there are high expectations and standards for eHealth strategies. Providing that security and privacy mechanisms are put in place, there seems high motivation in the adoption of such systems.

*Emanuele Lindo Secco, Robotic Laboratory, Department of Mathematics & Computer Science, Liverpool Hope University, L16 9JD, UK, +44 (0) 151 291 3641, seccoe@hope.ac.uk

www.astesj.com

<https://dx.doi.org/10.25046/aj030651>

The construction of the secure medical wearable device and development of the secure database with the front-end website application is outlined in the materials and methods section. The results of this research show that secure wearable devices for this purpose can be successful and the results of the application of these devices is discussed in the discussion and conclusion sections.

1.1. Population Health

It is reported that across the globe the causes of death in 2015, are related to cardiovascular and respiratory systems, these figures have risen since the year 2000 [1], for example, ischaemia heart disease had risen by 2 million in 2015 [1]. In 2015, diabetes mellitus caused around 1.7 million deaths and became one of many new causes of death and diseases related to the respiratory system caused around 1.8 million deaths [1].

A medical device which could record body parameters especially those parameters involving the cardiovascular system and respiratory system would be beneficial in monitoring the health of the population and improving healthcare. The device could provide crucial data to help reduce illness and diseases. The system could also be used to identify deterioration of health so that preventative measure can be put in place.

1.2. eHealth

The *eHealth* field involves public health and medical informatics; it describes the enhancement of health services through use of technology [2].

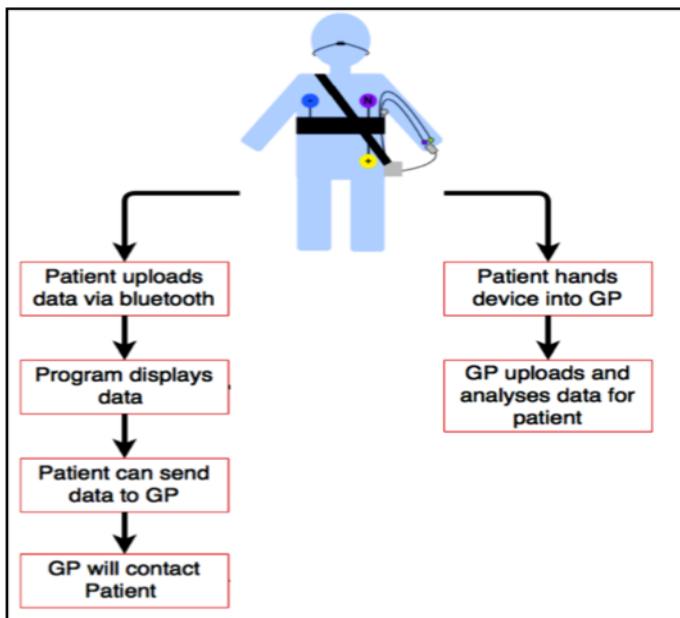


Figure 1: Functional overview of the Health Monitoring System

Research has shown that *eHealth* has provided improved support for patients whilst reducing the overall financial cost and improving the efficiency of healthcare. Portable health monitors have shown to increase self-care, improve quality of life and assist in providing better medical care [3]. An *eHealth* system can provide multiple benefits to the patient. The reasons for this are:

- Increasing the efficiency of health care by improving communication between the healthcare professionals and the

patients. Prevention of data redundancy, increased availability of data and improved accuracy of the data play an important role in the correct diagnosis of patients.

- *eHealth* strategies should be evaluated scientifically for its effectiveness and efficiency [4] which allow for improved services to the patient.
- *eHealth* improves time efficiency of health care, services such as telehealth save time because appointments can be made online or consultations could take place using video streaming services over the internet.
- *eHealth* also allows people to be more aware of their health and allows individuals to easily share data with their health care provider. Which provides for more effective and efficient health care because patients can be diagnosed more quickly and accurately.
- An *eHealth* System can also reduce administration as the health data can be digitally securely before being shared. This reduces hard copies (paper) which could be stored insecurely or even lost.
- Moreover, electronic systems could have validation built in which may reduce some human administrative errors resulting in more accurate data.
- Patients that are concerned with the security or the privacy of their data can be assured that the data is governed by The EU General Data Protection Regulation.
- Software for such systems is becoming easier to use. The technology should be simple to use and user-friendly [5] for accessibility to the patient.

eHealth is a successful mechanism in solving many health care issues. These innovative solutions could also be invaluable for developing countries by reducing the administrative burden involved in providing health care.

The *eHealth* platform has shown to be a widely used health care scheme by many countries and the use of *eHealth* is increasing as more countries are adopting the scheme. Since 1990 there has been a steady adoption of *eHealth* [6], more recently application of *eHealth* has seen a dramatic increase. Health care professionals usually assess a patient's condition at regular intervals, which results in discreet data. *eHealth* devices provide continuous data which gives a more complete medical history, this could help health care professionals make better informed decisions regarding treatments for patients. This is crucial as the patient could form new symptoms between the intervals the health care professional assesses the patient [7].

1.3. Security, Privacy and Legal considerations surrounding Medical Wearable Devices

There are security, privacy and legal issues surrounding *eHealth* wearable devices which must be addressed for the service to be viable. It is essential that the devices are kept secure because the recorded data hold personal information of patients. If security were compromised there could be catastrophic consequences, for example, the patient being incorrectly diagnosed or receiving incorrect treatment, which may result in the death of the patient.

Of course, this is true of all systems that store patient data and procedures must be in place to secure these systems.

The proposed medical wearable device in this research does transmit data over an insecure wireless medium and this connection link could be eavesdropped [8], but this is unlikely because of the proximity of the transmitter and receiver and the likely distance the signal will travel from the low powered transmitter (theoretically the signal should be confined to about 10m). However, even with such a low risk of interception of data, steps are in place to ensure that the data is encrypted prior to transmission. The data are also stored in a secure database in encrypted form and can only be accessed by authorised personal using the authentication scheme.

Listed below are some security and privacy concerns with medical wearable devices:

- The device’s system is vulnerable to eavesdropping, denial of service, unauthorized access to the device and data mining [9].
- In line with GDPR (General Data Protection Regulation), only required data should be stored. In addition to this the data stored should not allow an adversary to determine the identity of the patient.

The use of authentication mechanisms prevents unauthorized access to the data, examples of these are usernames, digital signatures and challenge-response protocols [10].

unique identification numbers which hides identifiable patient data and prevents patient identification. There is an authentication scheme and an access control list policy which prevents anyone without the correct authentication credentials from accessing the database.

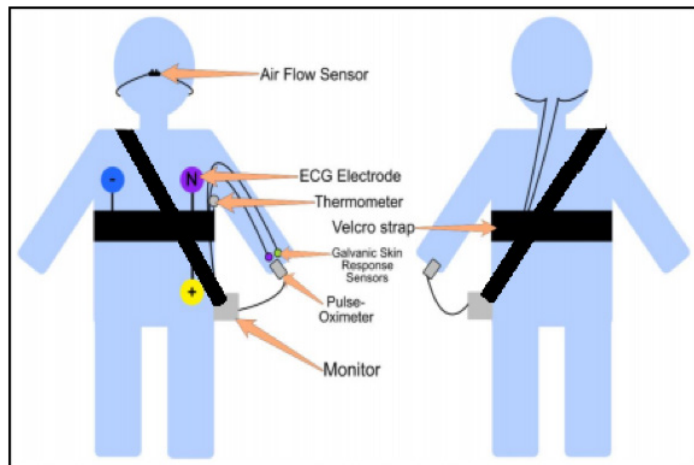


Figure 2 – Sensor configuration of the Health Monitoring System

2. Materials & Methods

The overall system process shown in Figure 1 and the creation of the secure medical wearable device shown in Figure 2 and web application shown in Figure 4, will be outlined within this chapter. The goal was to create a secure medical wearable device for use

on humans, to allow for independent health monitoring, health awareness, and improved health care, through use of eHealth.

2.1. Secure Medical Wearable Device

The wearable device was made up of an Arduino UNO Board, eHealth shield [11], data-logger shield and Bluetooth shield. The eHealth shield was used as it allows the Arduino to record and measure data of the human body parameters.

The sensors used with the eHealth shield are reported in Figure 2 and 3, which are:

- Electro cardiogram (ECG) and Electromyograph (EMG): A three electrode sensor. The electrodes can be used to either record data for an ECG or EMG. The ECG records data for heart health and the EMG records data for muscle health.
- Pulse-Oximeter: A non-invasive small battery-operated device which attaches to the index finger. Records blood oxygen levels and heart rate.
- Glucometer: Records blood glucose in Millimoles per litre.
- Body Thermometer: placed at the axillary point of the body. Records body temperature.
- Air Flow: Measures the breathing rate and is placed on the upper lip.
- Galvanic Skin Response: measures the electrical conductance of the skin. This consists of two sensors placed on two separate fingers.

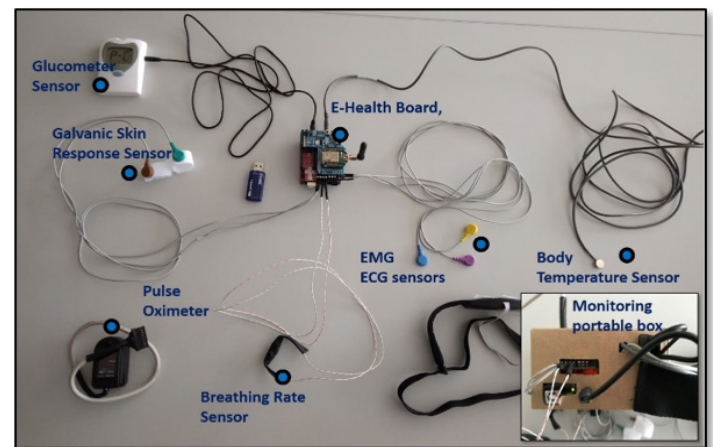


Figure 3 – The sensors and monitoring box set-up of the Health Monitoring System

The data logger shield allows the Arduino to store all the measured data onto an SD Card, with the correct date due to an

onboard real-time clock. This is programmed to store data every minute. Each log contains the date, time and the sensor readings. An authentication method was added to the SD card. To ensure the SD card was correct for the patient, so the data stored on the SD card had to match the data on the device before sensor recording would take place. The Bluetooth shield allows the Arduino to transfer data via a different medium. This shield uses a Bluetooth communication protocol (IEEE 802.15.1, Bluetooth 2.1 +EDR Class 2) and Asynchronous Encryption Standard with data transfers.

The Arduino board was used as it allows for these shields to be used and stacked together.

The Arduino Boards and shields was then encased within a box for protection of the user and the device. The box allowed access to ports and cables. A Velcro strap was attached to the box, so it can be worn over the shoulder and keep the cables tidy and managed. The strap also went across the chest supporting the ECG and thermometer cables. The device was powered by a 1200 mAh battery, with an output of 0.8 A and 5V. The power bank included LEDs to indicate charging and power.

2.2. Communication with the Device

An application was developed using Processing IDE, to communicate with the health monitor via Bluetooth or USB cable. This application was programmed to search for the Arduino board, once found it would display a window asking to name a file for which the data will be stored in. Once entered, three coloured steps were added to retrieve data, close the file of the stored data and reset the device.

2.3. Web application mask

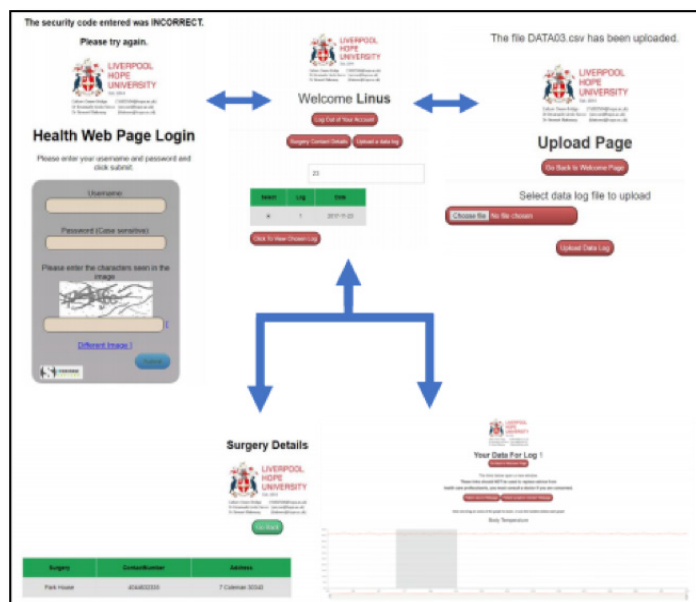


Figure 4 - The Web Application Masks for the Patient

A database was also designed to store the patient information, the health care professional information and the data output of the health monitoring system. The database was stored on a WampServer version 3.1.0 [12]. This server allowed storage of a MySQL database, web application and the communication between them. This database held login information using SHA2 at 512 bits to hash passwords for the web application.

A website was also designed. A login page was added, containing field box for username, password and secure image text. Depending on the username, the user is directed to either the health care professional web page or patient web page. Moreover, a secure image capture was added to the login page to reduce brute force attacks. Text entered in the password field box is hidden and data transmitted between pages are hidden.

A patient welcome webpage was developed. This displays a search box for a table containing data logs, buttons for logging out, surgery details webpage, upload webpage and to continue to graphical webpage once a log is selected. Figure 4 reports an overall summary of the different webpages which are presented to the patient, a similar view is provided for the health care professional.

The upload webpage allows the user to upload files of data saved when importing data from the health monitor. The surgery details webpage, displays details of the surgery the patient is attending, this is shown in Figure 4.

When directed to the graphical webpage, the user can view all their sensor data in graphical form and are able to zoom into specific time points. Also, the user can go back to the previous page or go to webpages for a symptom checker or patient advice, this is shown in Figure 4.

The health care professional webpage, allows health care professionals to search for patients they have access to and view their data logs in graphical form.

3. Results

3.1. Secure Medical Wearable Device

The secure medical wearable device, shown in Figure 3, can measure vital body parameters and store the data with the date and time they are measured on the SD card. The communication with the device was successfully tested between the monitoring system and a personal computer; the data were also uploaded successfully to the database and accessed via the web application. The AES Bluetooth encryption algorithm properly encoded the data. The storage medium provides a mechanism for user authentication; therefore, data is stored in encrypted form before transmission via a public network or internet.

The device has been successful in these preliminary tests (i.e. sensing): the data are encrypted at the patients end before being transmitted to the secure database and the SD card is protected using authentication mechanisms.

3.2 Web Application

The successful implementation of the front-end web application is shown in Figure 4. This application is driven and populated using a secure relational database. The system allows for

multiple types of users with different security policies, for example a patient user and a medical professional.

The system is designed such that only the patient can upload, and view data associated with them. The medical professional can only access data that is associated with their patients. The graphs were able to show all the data recorded, allowing for a more accurate diagnosis.

To ensure a secure system many security methods were implemented, from a simply masking of the password and the one-way hashing of passwords which guards against the possibility that someone who gains unauthorised access to the database can retrieve the passwords of the users in the system. Also, used were functions that removed unauthorised characters to prevent SQL code injection type attacks. In addition, the system prevents brute force attacks by using a secure image capture facility. The system would also use transport layer security and security certificates, by hiding passwords entered, using hashed passwords to hide what has been typed in the field box, HTML special chars was used to prevent the webpages from being exploited with code injections and the secure image capture was added to reduce brute force attacks.

3.3 Ethical Issues with the System

As with any system that collects data about individuals there must be consideration of the ethical issues related to this system.

The collected data could be used against the patient by health insurances or wellness programs run by employers. This data could be shared with of health care clinics without the patient knowing. This data could be misused to harm the patient. This health monitor could be misused to keep track of people. For example, a business company or insurance company could force employees or customers to wear this device to keep track of them, instead of using the device for its purpose of wellness improvement and health awareness. It would be imperative that the system complies with all GDPR and that users of the system are trained and authorised.

4. Conclusion and Discussion

In this paper we presented a novel integrated system for the real-time monitoring of physiological parameters. The system provides an SD card reader for saving the data locally, encryption algorithms to encrypt the data, authentication services, and other security procedures. The wireless IEEE 802.15.1 secure communication protocol for transferring the data and a web service was used for the patient and the health care professional.

The system outlined in this paper was implemented successfully. It records physiological parameters, transfer data securely to a computer where it was stored in a secure database and was accessible by different users of different security policies (i.e. patient and health care professional).

The system automatically performs some analysis of the data and presents the information in graphical form. for both the patient and health care professional in graphical form using the web application.

Despite the preliminary performed tests, the system will have to be validated vs. golden standard instrumentations, namely

medical and certified equipment [13-15]. Moreover, further laboratory trials should be performed to optimize the portability of the system and to make the interface more user-friendly in a daily medical and hospital scenario/environment. On this matter, an assessment on the reliability and accuracy of the system was not performed at this stage. Therefore, a preliminary validation of this system vs. golden standard instrumentation should be performed in order to validate the system. Such a validation may be based on comparing the measurements of the system with the same parameters as obtained from a polygraph or similar devices and by processing this data with a similar approach to [16, 17], namely Bland Altman technique or similar ones.

The device developed is a prototype and currently we are looking into ways to reduce the size of the energy source and increase the overall lifetime of the system before a recharge is required, as the current system will only last up to 10 minutes. We are also investigating a better, smaller design for the housing of the hardware, so that it is less invasive for the wearer. It is believed however, that point of concept has been proven.

In future these devices could be medically certified which would result in trust of the devices, the processes, and with the accuracy of the data.

Acknowledgment

This work was presented in thesis form in fulfilment of the requirements for the BSc in Computer Science and Electronic Engineering for the student Callum Owen-Bridge under the supervision of Dr. S. Blakeway and Dr. E.L. Secco from the Robotics Laboratory, Department of Mathematics & Computer Science, Liverpool Hope University

References

- [1] World Health Organization. (2017). The top 10 causes of death. [online] Available at: [Accessed 6 Dec. 2017].
- [2] Eisenach G. (2001). What is e-health?. Journal of Medical Internet Research, [online] Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1761894/> [Accessed 6 Dec. 2017].
- [3] Alnosayan N, Lee E, Alluhaidan A, et al (2014). MyHeart: An intelligent mHealth home monitoring system supporting heart failure self-care. 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services.
- [4] Government.nl. (2018). Benefits of eHealth. [online] Available at: <https://www.government.nl/topics/ehealth/benefits-of-ehealth> [Accessed 1 Mar. 2018].
- [5] de Grood C, Raissi A, Kwon Y and Santana M (2016). Adoption of ehealth technology by physicians: a scoping review. Journal of Multidisciplinary Healthcare, 9, 335-344.
- [6] World Health Organization. (2017). Global Diffusion of Ehealth. Geneva: World Health Organization.
- [7] Virone G, Wood A, Selavo L, et al (2006). An Advanced Wireless Sensor Network for Health Monitoring. Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare
- [8] Li M, Lou W and Ren K (2010). Data security and privacy in wireless body area networks. IEEE Wireless Communications, 17(1), 51- 58..
- [9] Omoogun M, Seeam P, Ramsurrin V, Bellekens X and Seeam A (2017). When eHealth meets the internet of things: Pervasive security and privacy challenges, 2017 International Conference on Cyber Security And Protection Of Digital Services.
- [10] Meingast M, Roosta T and Sastry S (2006). Security and Privacy Issues with Health Care Information Technology. 2006 International Conference of the IEEE Engineering in Medicine and Biology Society.
- [11] Cooking-hacks.com.(2018).e-HealthSensorPlatformV2.0forArduinoandRaspberry Pi [Biometric / Medical Applications]. [online] Available at: <https://www.cooking-hacks.com/documentation/tutorials/ehealth->

biometric- sensor-platform-arduino-raspberry-pi-medical [Accessed 22 Mar. 2018].

- [12] Bourdon R (2018). WampServer. Alter Way.
- [13] Secco EL, Curone D et al (2012). Validation of Smart Garments for Physiological and Activity-Related Monitoring of Humans in Harsh Environment, *American Journal of Biomedical Engineering*, 2(4), 189-196
- [14] Magenes G, Curone D, Caldani L, Secco EL (2010). Fire fighters and rescuers monitoring through wearable sensors: The ProeTEX project, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*
- [15] Curone D, Secco EL, et al (2012). Assessment of sensing fire fighters uniforms for physiological parameter measurement in harsh environment, 16(3), 501-511, *IEEE Transactions on Information Technology in Biomedicine*
- [16] Secco EL, Curone D, Tognetti A, Bonfiglio A, Magenes G (2012). Validation of Smart Garments for Physiological and Activity-Related Monitoring of Humans in Harsh Environment, *American Journal of Biomedical Engineering*, Vol.2, No.4, 189-196
- [17] Curone D, Secco EL, Caldani L et al (2012). Assessment of Sensing Fire Fighters Uniforms for Physiological parameter Measurement in Harsh Environment, *IEEE Trans on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 501-511

Study on CD ROADM Contention Blocking

Guangzhi Li^{*1}, Kerong Yan², Li Huang², Bin Xia², Fanhua Kong², Yang Li²

¹Futurewei Technologies, Inc., 400 crossing road, Bridgewater, NJ 08807, USA

²Huawei Technologies Co., Ltd., Shenzhen, 518129, China

ARTICLE INFO

Article history:

Received: 18 October, 2018

Accepted: 05 December, 2018

Online: 19 December, 2018

Keywords:

CD ROADM

Contention Blocking

Contention Avoidance

ABSTRACT

Service providers are transferring their static optical transport networks from semi-permanent connections to agile automatic switched optical networks (ASON) with dynamic optical connection provisioning and restoration. To achieve this goal, service providers are looking for flexible optical network ROADMs with CDC capabilities. Although many contention scenarios during network connection provisioning and restoration have been illustrated, surprisingly academic simulations have showed that the blocking probability improvement of CDC ROADM comparing with CD ROADM is not significant. This is good news for service providers since most deployed optical networks are only CD ROADM capable, instead of CDC ROADM capable. How to make use of existing CD ROADMs to achieve network automation becomes an urgent challenge. In this paper, we present two research results to attack this challenge: (1) first, we built an analytical mode to estimate the CD ROADM contention blocking probability and show that when a CD ROADM add/drop local direction capacity occupation ratio is low or moderate, the contention blocking probability is not significant. From this model, we estimate that one can use a CD add/drop local direction capacity occupation ratio up to 75% before installing another CD ROADM add/drop local direction or installing a CDC add/drop local direction when available. Simulation results on real network topologies and traffic matrices verified our recommendation; (2) second, we observed that most deployed optical networks are usually providing 100G or 200G per wavelength while majority applications are still requesting much smaller bandwidths and service providers often provide OTN (Optical Transport Network) over ROADM architecture for transport services. Since OTN provides electronic switching capability, in this paper, we present a new algorithm and methodology to make use of both OTN switch and CD ROADM to avoid service contention without using CDC ROADM.

1. Introduction

Color-less, Direction-less, and Contention-less Reconfigurable Add/Drop Multiplex (CDC-ROADM) architectures have recently generated considerable interests among service providers and optical transport vendors [1~7]. There are many technical papers and industry white papers to describe the benefits and applications of CDC ROADM. Comparing with CD ROADMs (Colorless and Directionless only), CDC ROADMs are able to offer additional flexibility and simplicity for optical wavelength planning and operation, especially for dynamic wavelength traffic as well as wavelength dynamic restoration. However CDC ROADM brings extra

components and complexity, which increases ROADM cost. In field network deployments, the majority deployed optical networks are still CD ROADM only. Then following questions come to our mind when service providers upgrade their optical ROADM networks: (1) where does the CD ROADM contention come from? (2) How much is CD ROADM contention blocking? (3) Can we use an analytical formula to estimate CD ROADM contention blocking? (4) Most importantly, what should we do when CD ROADM contention blocking occurs?

In this paper, we present our two research results to answer above questions: (1) first, we built an analytical mode to estimate the CD ROADM contention blocking probability and show that when a CD add/drop local direction capacity occupation ratio is low or moderate, the contention blocking probability is not

^{*}Guangzhi Li, Futurewei Technologies, Inc., 400 crossing road, Bridgewater, NJ 08807, USA, Guangzhi.li@huawei.com

significant [8]. From this model, we estimate that one can use a CD add/drop local direction capacity occupation ratio up to 75% before installing another CD add/drop local direction or installing a CDC add/drop local direction when available. Simulation results on real network topologies and traffic matrices verified our findings; (2) second, we observed that currently most optical networks are usually providing 100G or 200G per wavelength while majority applications are still requesting much smaller bandwidths and service providers often design OTN (Optical Transport Network) over ROADM architecture for transport services. Since OTN provides electronic switching capability, in this paper, we present a new algorithm and methodology to make use of both OTN switch and CD ROADM to avoid service contention without using CDC ROADM.

2. ROADM Architecture Comparison

Early stage of optical transport network is typical point-to-point wavelength division multiplex (WDM) system which consists of two terminals connected by a pair of fibers. Each terminal contains an optical wavelength multiplexer and de-multiplexer, amplifier, and transponders that interface client signals. The number of inline amplifiers placed between the terminals depends on the length and quality of fiber. Each transponder re-transmits its incoming client signal onto a particular wavelength of the optical grid (also called a channel) and the optical multiplexer combines these signals at different wavelengths together and transmits the combined signal over a fiber to the de-multiplexer at the other end. The de-multiplexer decomposes the multiplexed signal back into original signals at their respective wavelengths, which are re-transmitted by each receiving transponder into its client signal.

When a point-to-point demand (also called connection) is transported over two WDM systems, two OTs are needed: the first OT converts the first WDM system wavelength into the common short-reach wavelength (λ_0), and the second OT converts the λ_0 short reach wavelength into the second WDM system wavelength. However, when the two WDM systems are of the same technology from the same vendor, one regenerator can be used to replace the two back-to-back OTs, avoiding the conversion to the common short-reach wavelength, and thus reducing the component costs.

Later on, Reconfigurable Optical Add Drop Multiplexer (ROADM) technologies for optical transport network have been deployed due to their high capacity and capital savings. A ROADM network typically includes a set of multi-degree nodes connected via fibers to form a mesh topology. Traffic may be added or dropped, regenerated, or expressed through at ROADM nodes. Inside ROADM node, each network side WSS (wavelength selective switch) is called one line direction, each add/drop side WSS is called add/drop local direction. At the add/drop local direction, the classic ROADM was designed with fixed wavelength transponders and directed tributary for each line direction, and it was called colored and directed ROADM. Each transponder only is allowed to transmit signal on a fixed wavelength and to a fixed direction, see Figure 1 for an example. This ROADM design is acceptable for static optical connections. Network planners could plan the connections carefully and deploy them as planned for quite a long time. As network traffic grows and become more and more dynamic, service providers prefer to

automate optical connection provisioning without manual intervention. In such an operation scenarios, this colored and directed ROADM architecture is no longer able to satisfy this requirement. For example, to reroute an existing wavelength to a different direction or to reuse an existing transponder for a different wavelength, all require new ROADM add/drop architecture, which was colorless and directionless ROADM, the so called CD ROADM.

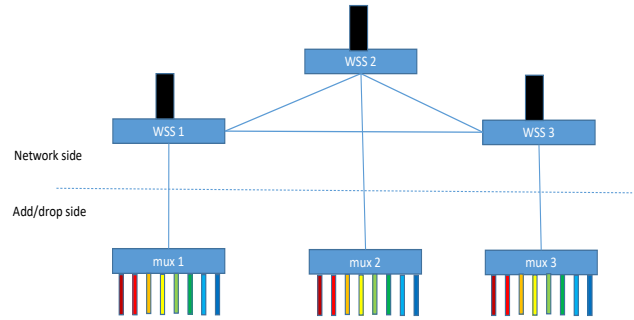


Figure 1: Example of colored and directed ROADM

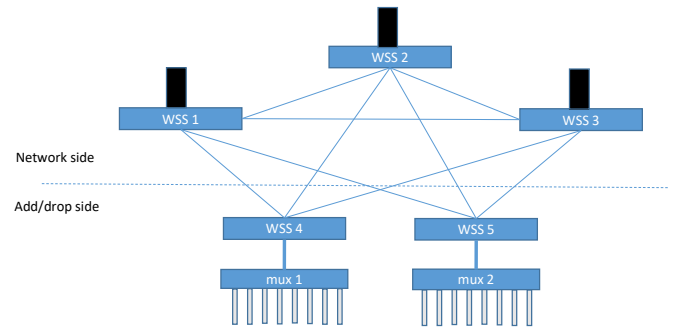


Figure 2: Example of CD ROADM

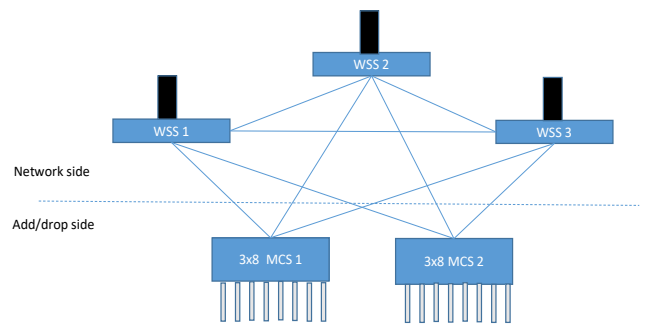


Figure 3: Example of CDC ROADM

Figure 2 shows an example of CD ROADM architecture with 3 line directions and 2 add/drop local directions. There are 8 tunable transponders, which means each transponder can be tuned to any wavelength. However, no more two transponders at the same add/drop local direction can be tuned to the same wavelength since all tuned wavelengths from transponders will be multiplexed into a single fiber pair. This limitation is called add/drop local direction contention. If two clients are connected to two transponders at a common add/drop local direction. Each client has a wavelength request at the same time, and routing wavelength assignment algorithm finds the same wavelength for the two requests. Due to add/drop local direction contention, the

two transponders could not use the same wavelength. Then either one of them has to find a different wavelength or may be blocked. This limitation of CD ROADM architecture leads to another advanced colorless directionless, contention-less ROADM architecture, the so called CDC ROADM.

Figure 3 shows an example of CDC ROADM architecture with 3 line directions and 2 local directions. In this architecture, the WSS (wavelength selective switch) and Multiplexer are replaced by a 3x8 multicast optical switch (MCS). Without the fiber bottleneck of CD architecture, the transponders in same add/drop local direction can be tuned to the same wavelength, i.e., there is no wavelength contention in CDC ROADM architecture. This wavelength assignment flexibility is brought in by the relatively expensive component of MCS. In Figure 3, one can tune 3 transponders at a single add/drop local direction into a single wavelength since there are total 3 line directions. Of course, it is unnecessary to tune more than 3 transponders into a single wavelength since there are at most 3 common wavelengths can be supported at the network side. This extra flexibility should be able to improve network capacity utilization and reduce network blocking probability.

3. CD ROADM contention

It is widely accepted that CD ROADM architecture causes wavelength contention during wavelength provisioning when two connections with the same wavelength need to be added/dropped at the same add/drop local direction [9~12]; during wavelength restoration when the network available wavelength has been used by other connection at the same add/drop local direction, and/or during regeneration when the connection wavelength is free at most one add/drop local direction only. To overcome CD ROADM contention issue, one may add as many local directions as the number of ROADM line directions. Then if there is one free wavelength at any line direction, there must exist one add/drop local direction with the same wavelength free. However during wavelength restoration, contention could still occur: during wavelength restoration, the transponder at the client side could be reused, and the restoration wavelength at network side is available while the restoration wavelength may not be available at add/drop local direction of the client transponder. To solve this contention, client-side optical cross connect architecture is recommended [10]. Problem seems solved, the issue is the extra cost of cross-connect. At each ROADM site, the number of connection requests usually much less than the value of $M*W$, where M is the number of line directions and W is the number of wavelengths per fiber. Large number of add/drop local directions will reduce the available number of potential line directions for the same size of WSS, which makes network expansion difficult. In real deployed optical networks, the number of add/drop local directions usually is much smaller than the number of line directions. In this case, an expensive CDC add/drop local direction could be used to avoid wavelength contention.

All above analysis is quantitative with extreme case assumptions. How severe is the CD ROADM contention during optical network planning and operation? Academic studies show that under reasonable assumption on network topologies and demands, the blocking probability improvement using CDC ROADM comparing CD ROADM with the same number of add/drop local directions at each node is not significant [13~15].

This surprising observations brought us interests for a further investigation at CD ROADM contention issue, including theoretical analysis and simulation verification, as well as how to resolve CD ROADM contention.

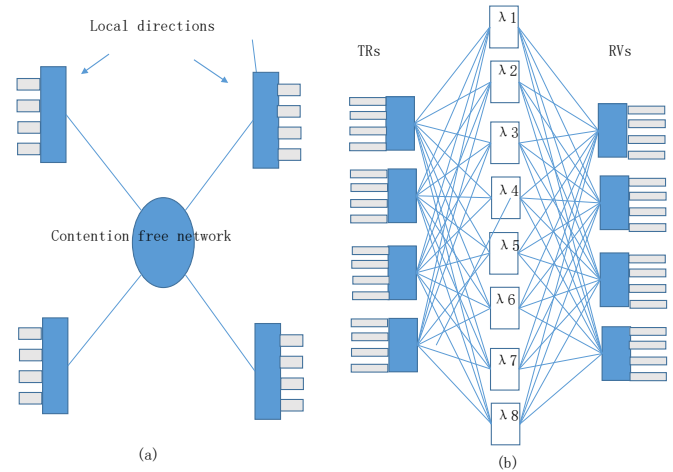


Figure 4: Clos model of CD ROADM

A connection in a CD ROADM network is blocked when the network could not find a free wavelength, or the wavelength is not available at either side of add/drop local directions. The first one is due to optical network wavelength continuity constraint and the second is due to CD wavelength contention. Without considering network wavelength continuity constraint, we could simply model the CD ROADM network as a star topology, see figure 4(a): the network is a full contention free optical switch at the center, each CD add/drop local direction is a star terminal. Assume there are n tunable transponders at each terminal, and each star branch fiber supports m wavelengths. Each transponder can be tuned to any wavelength, but no more two transponders can be tuned to a same wavelength. Then two terminals are able to establish a connection if and only if the two terminals have two transponders and the two branch fibers have a common free wavelength. If there are total r add/drop local directions, the simplified star model can be viewed as a (r,n,m) Clos network [16], see figure 4(b), where a single transponder is separated into one transmitter (TR) and one receiver (RV). Thus the non-blocking condition would be $m \geq 2n - 1$, i.e., $n \leq (m+1)/2$. So when an add/drop local direction is half filled, in order to avoid contention, it would be the time to add new CD add/drop local directions, or upgrade CDC add/drop local directions. Of course this simplified model has drawbacks since the optical network is not contention free at all. Even if two add/drop local directions have free transponders and same available wavelengths, but the network may not have the same wavelength between the associated two nodes; on the other hand, if the associated two node has an available wavelength along a path between them, the two add/drop local directions may not have the same common wavelength available. Thus our model is under estimate the contention blocking. However since there could be large number of paths between any two network nodes in real networks, we think this estimation may not be too off from reality. In next sections, we will verify our simplified model via simulation comparing CD ROADM and CDC ROADM of real networks with real demand matrix.

Academic papers show that even if the CD add/drop local direction traffic is more than half-filled, the blocking probability

is still relatively low. When we want to create a connection between two CD add/drop local directions, again we look at the clos network model, which means we want to establish a connection from one TR bank I of the first stage to one RV bank J of the third stage. We define clos network state $\{u, v\}$ of TR bank I and RV bank J as u TRs are busy and v RVs are also busy, and define $B(u, v)$ as the contention blocking probability in state $\{u, v\}$ independent of other TR banks and RV banks.

According to Jacobaeus [17], the contention blocking probability $B(u, v)$ in state $\{u, v\}$ is:

$$B(u, v) = \begin{cases} 0 & \text{when } u + v < m \\ \frac{u! v!}{m! (2n - m)!} & \text{when } u + v \geq m \end{cases}$$

Simple proof: when $u+v < m$, there is at least one wavelength available in the second stage, which is reachable for both TR bank I and RV bank J , so no blocking for a new connection. When $u+v \geq m$, depending on how u and v connect to the second stage wavelengths, there could be no more wavelength available for next connection if and only if RV bank J all available wavelengths ($m-v$) resides in TR bank I busy wavelengths. So the blocking probability could be calculated as:

$$B(u, v) = \frac{\text{choosing } m - v \text{ wavelengths from } u \text{ wavelengths}}{\text{choosing } m - v \text{ wavelengths from } m \text{ wavelengths}} = \binom{m-v}{u} / \binom{m-v}{m} = \frac{u! v!}{m! (u+v-m)!}$$

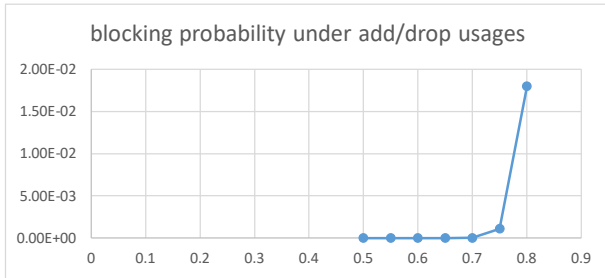


Figure 5: CD ROADM estimated blocking probability

Now we let $u=v=x*m$, where $m = 80$, and calculate the blocking probability of $x=0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8$. See Figure 5, where x-axis is the value of x for number of existing connections in CD add/drop local directions, and y-axis is the estimated blocking probability. It is easy to see that even if the add/drop local direction is 80 percent full, the contention blocking probability is still under 2% without considering network wavelength continuity constraint. When the CD add/drop dimension is 75% full or less, the contention blocking probability is still negligible. Based on this analytical modeling, it explains why all academic simulations reported very small blocking probability using CD ROADM architecture. In next sections, we will use real network topology and network traffic matrix to

simulate CD ROADM blocking probability. Our results confirm this analysis with very small variance.

4. An example interpretation of clos model

In section 3, we model CD ROADM as a clos network model and claim that when the CD ROADM local directions are not half filled, the CD ROADM blocking probability will be similar to CDC ROADM blocking probability. This claim may not be easy to interpret. In this section, we give one simple example to show that when CD ROADM local direction is half-filled, some CD ROADM connections will be blocked while the same network configuration with CDC ROADM will not block those connections.

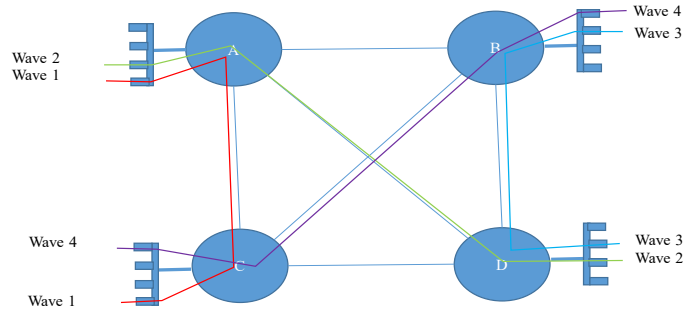


Figure 6: Example of CD ROADM half-filled blocking

Figure 6 shows a 4 node ROADM network and each node is deployed with a single CD add/drop local direction. Without loss of generality, we assume that each fiber supports 4 wavelengths. Due to network dynamic operation and wavelength connections arrive and leave. At some network stage, we assume there are only 4 connections: connection 1 from node A to node C with wavelength 1; connection 2 from node A to node D with wavelength 2; connection 3 from node B to node D with wavelength 3; and connection 4 from node B to node C with wavelength 4. If there is a new connection request from node A to node B, the request will be blocked due to CD ROADM local contention, while CDC architecture will be able to provision such connection request; similarly another connection request from node C to node D will be blocked also in CD architecture while CDC architecture will not block it. In this simple example, it is easy to see that CD ROADM local direction half-filled is a critical point to cause contention blocking.

5. CD/CDC ROADM network simulation

We first simulate a few planned or planning CD ROADM optical networks under preplanned demand matrixes with and without rerouting restoration. Table 1 shows the parameters of these networks, where #degree is calculated as $2*(\#links)/(\#nodes)$, where #links means the number of network links, and #nodes means the number of network nodes. The #add/drop direction distribution $X(Y)$ means y network nodes with x add/drop local directions. We first assume all add/drop local directions are CD architecture, then using a multipath routing and first-fit wavelength assignment algorithm to provision the connections as many as possible.

Table 1: Simulated network parameters

networks	#node	#link	#degree	#connections	#add/drop direction distribution X(Y)
Net 1	30	71	4.73	324	2(2),1(18),0(10)
Net 2	28	43	3.07	404	2(5),1(23)
Net 3	55	81	2.95	723	3(2),2(4),1(47)

Table 2: Simulation results of CD vs CDC

networks	CD (W) blocking	CDC (W) blocking	R(CD)	R(CDC)	Extra CD for working	Extra CD for rerouting	Max CD fill ratio before extra CD	Max CD fill ratio after extra CD
Net 1	0	0	0.43%	0	0	8	86.25%	43.50%
Net 2	7.4%	1.7%	16.5%	16.9%	3	3	85%	56.70%
Net 3	3.87%	1.93%	6%	3.75%	12	12	88.80%	53.00%

For those provisioned connections, we simulate single failure rerouting and calculate the maximal success rerouting ratio:

$$R(CD) = \frac{\text{total rerouting success connections over all failures}}{\text{total failed connections over all failures}}$$

Then we assume all add/drop local directions are CDC architecture, then using the same algorithm to provision the connections as many as possible. For all provisioned connections, we also simulate single failure rerouting and calculate the maximal success rerouting ratio, R(CDC).

To investigate the CD contention impact, we try to add extra CD add/drop local directions to critical nodes such that the CD blocked demands equal to CDC blocked demands. The CD add/drop local direction adding policy is based on decreasing ratio of $(\#add/drop_connections)/(80*\#add/drop_direction)$ at each network node. We simulated both with and without failure rerouting cases. Table 2 shows our simulation results, where “Extra CD for working” means total extra CD local directions for working only with the same blocking ratio as CDC ROADM architecture, “Extra CD for rerouting” means total extra CD local directions for working and restoration with the same blocking ratio as CDC architecture, “Max CD fill ratio before extra CD” means maximal CD add/drop local direction fill ratio before adding extra CD add/drop local directions, and “Max CD fill ratio after extra CD” means maximal CD add/drop local direction fill ratio after adding extra CD add/drop local direction for rerouting. We observed that for all tested real networks, the differences between CD and CDC are relatively small. Especially when node CD fill ratio (total #add/drop demands divided by link capacity times #add/drop directions) is less than 50%, CD and CDC outputs are almost the same, which means that CDC does not provide any extra benefits in this case; on the other hand, under the same network topology and demand matrix, in order to achieve the same throughput and rerouting success ratio, CDC did require relatively fewer add/drop local directions, but not significant.

All these tests are based on off-line planning algorithms with static traffic demands. In order to simulate dynamic traffic, we also created a hypothetical network with 50 nodes, 123 links with average node line directions about 4. Link capacity is 80, and each node has one add/drop local direction. In this simulation, we assume paths between any two nodes are all reachable and no regenerator is used. Demands are random generated with poisson arrival and expectation holding time. We measure the Erlang

value when blocking probability is under 1%. We observed that CDC architecture could improve the throughput of Erlangs about 4% comparing with CD architecture, see Figure 7. Again CDC is able to improve dynamic traffic throughput, but the improvement is not significant.

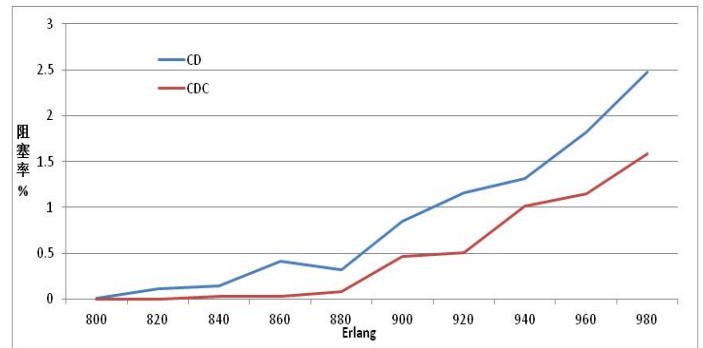


Figure 7: Throughput improvement of CDC vs CD

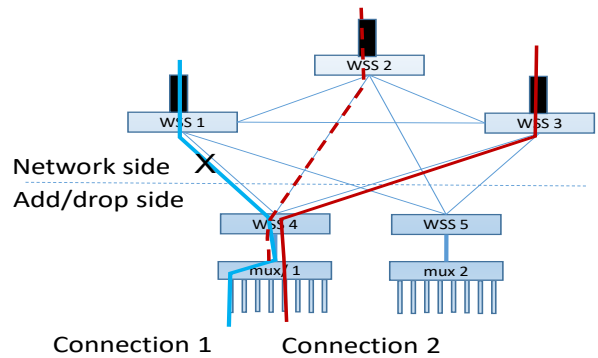


Figure 8: Example of CD ROADM contention

6. A Contention Solution for CD ROADM

Although CD ROADM contention blocking is small when add/drop local direction utilization is low or medium, there is still some possibility of contention blocking during dynamic optical network operation, such as ASON (automatic switch optical network) connection provisioning or restoration. For example in Figure 8, assuming local direction WSS 4 has two connections of 1 and 2, connection 1 was assigned wavelength 1 and routed along WSS 1 line direction while connection 2 was assigned wavelength 2 and routed along WSS 3 line direction. If connection 1 fails and required rerouting restoration, and the source node finds out that

only wavelength 2 along WSS 2 line direction is available and other wavelengths along other WSS line directions are not available. But due to add/drop local direction WSS 4 wavelength contention, wavelength 2 could not be used for connection 1 rerouting, which could lead to connection 1 rerouting failure. If this ROADM is CDC architecture, connection 1 is able to be routed to WSS 2 line direction with free wavelength 2.

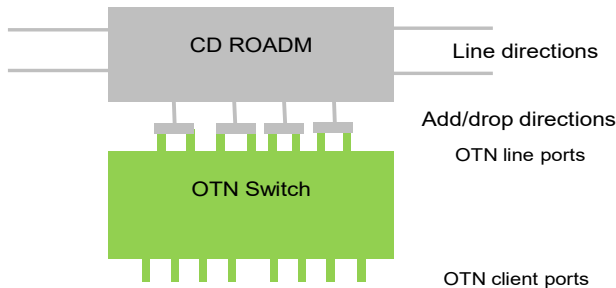


Figure 9: OTN over ROADM architecture

As we mentioned before, CDC ROADM relies on MCS module to multicast each wavelength to each direction, which increases hardware cost and system complexity. In this paper, we present a new solution to avoid CD ROADM wavelength contention during wavelength provisioning and rerouting using existing mature technology without increasing hardware cost. We noticed that most of ROADM node deployments are armed with OTN electronic switching node to form an OTN over ROADM architecture, see Figure 9, where OTN switch is used for traffic grooming and ROADM is used for wavelength switching. However such OTN over CD ROADM architecture still could not be equivalent to CDC ROADM architecture. For example, assume a CD ROADM has 4 line directions, 4 add/drop local directions, and each fiber has 4 wavelengths. Since OTN line ports are pre-installation in field operation, assuming 50% connections add/drop per node, the service provider pre-installed 8 OTN line ports and their usage is based on average policy as recommended by paper [10]. If connection wavelength sequence order is 1,1,1,2,3,3,3,2, the last connection would be blocked due to wavelength contention, however CDC ROADM won't have this contention. However under OTN over ROADM architecture, we have following observations:

Observation 1: wavelength contention blocking condition is that for a given wavelength, all add/drop local directions with free OTN line ports have used the given wavelength.

Observation 2: if CD ROADM has equal number of add/drop local directions as line directions, and one line direction has a free wavelength, then at least one add/drop local direction has the same free wavelength.

Observation 3: given a connection routing wavelength, if the add/drop local direction with the free given wavelength has free OTN line ports connected, the connection can be deployed without contention;

Observation 4: given a connection routing wavelength, assuming equal number of add/drop local directions and line directions, then according to observation 2, at least one add/drop local direction has the wavelength free, say WSS X, and X has no free OTN line port, but add/drop local direction WSS Y has free OTN line port. If OTN line ports are equally connected to CD ROADM add/drop local directions, then WSS X at most has one OTN line port less

than WSS Y. So WSS X at least has one wavelength occupied but WSS Y is still free. This is because if WSS Y occupied all wavelengths of WSS X used, and WSS Y also used the given new wavelength, and free OTN line port, then WSS Y would have at least 2 more OTN line ports than WSS X, which contradicts the assumption of OTN line ports equal distribution to ROADM local directions.

Observation 5: if we can change the free OTN line port (say y) of WSS Y into the wavelength that WSS X used but WSS Y is free, say λ_1 , identify wavelength λ_1' ROADM line direction (say d), cross-connect the OTN line port y to ROADM line direction d , switch the service OTN client port to OTN line port y , free WSS X wavelength λ_1 occupied OTN line port (say z), then WSS X has free wavelength λ , and connection free OTN line port z , thus the new connection can be provisioned without contention.

Observation 6: to achieve hitless switching, one could use make-before-break technology; however traditional make-before-break usually is used on the same layer of technology, such as in optical layer or electronic layer, in this proposal, the make-before-break is used in cross-layer between electronic layer and optical layer. We have discussed with both optical system engineers and electronic system engineers and was confirmed that it can be done using existing technologies.

7. Two Use Cases

In this section, we provide two use cases on how to apply the method of section 6 to avoid CD ROADM contention.

7.1. Use case 1: service provisioning contention blocking avoidance

Figure 10 shows an OTN over CD ROADM service provisioning contention blocking avoidance process. Assume the OTN, has 8 client ports numbered from 1 to 8, and 8 line ports connecting to 4 add/drop local directions of a CD ROADM, numbered as x,y , where x is the CD ROADM add/drop local direction number, y is the line port sequential order number at the local direction. For example 2.1 means the first line port at the second local direction of the ROADM.

We further assume that each fiber supports 4 wavelengths, OTN client port 1 cross-connects to line port 1.1 with wavelength 1, client port 2 cross-connects line port 2.1 with wavelength 1, client port 3 cross-connects line port 3.1 with wavelength 2, client port 4 cross-connects to line port 4.1 with wavelength 2, client port 5 cross-connects to line port 1.2 with wavelength 3, client port 6 cross-connects line port 2.2 with wavelength 3, client port 7 cross-connects to line port 3.3 with wavelength 3. At this time, there is still one free client port 8 and one free line port 4.2. If a new service finds available wavelength 2 over one line direction, the service cannot be deployed since CD ROADM local direction 4 already occupied wavelength 2 (see figure 10.a). Contention blocking occurs and following procedure could resolve this contention.

- ✓ Identify free line port: 4.2 on local direction 4
- ✓ Identify one local direction with free wavelength 2: local direction 1.
- ✓ Identify one OTN line port with occupied wavelength from local direction 1 which is free at local direction 4: line port

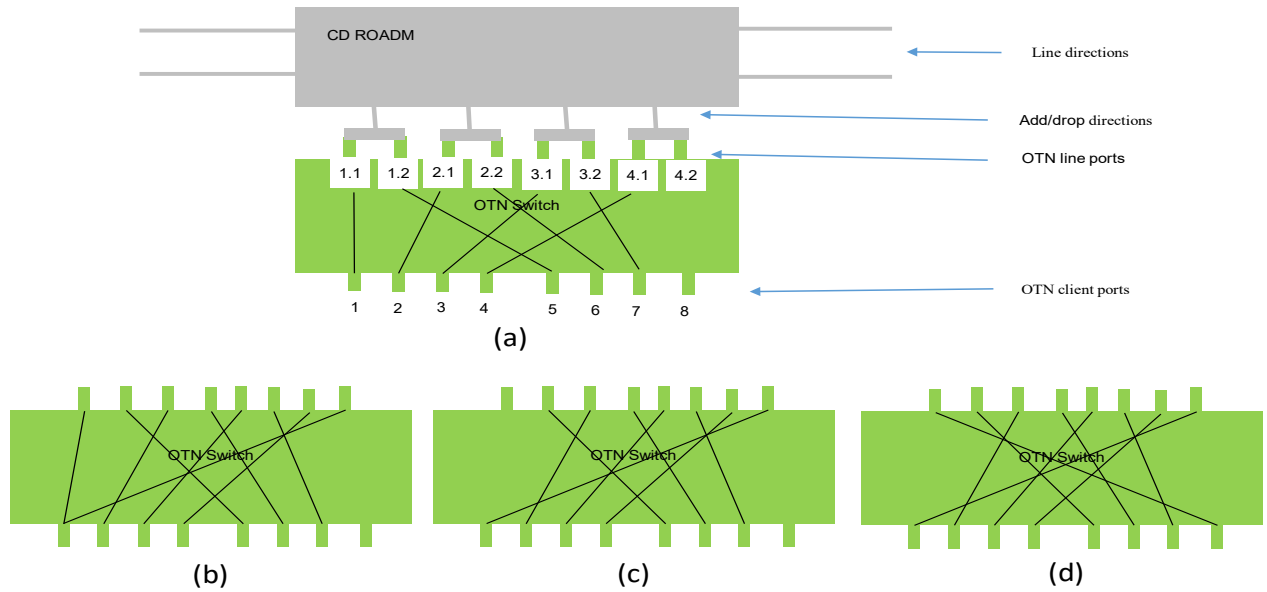


Figure 10: OTN over ROADM provisioning contention avoidance

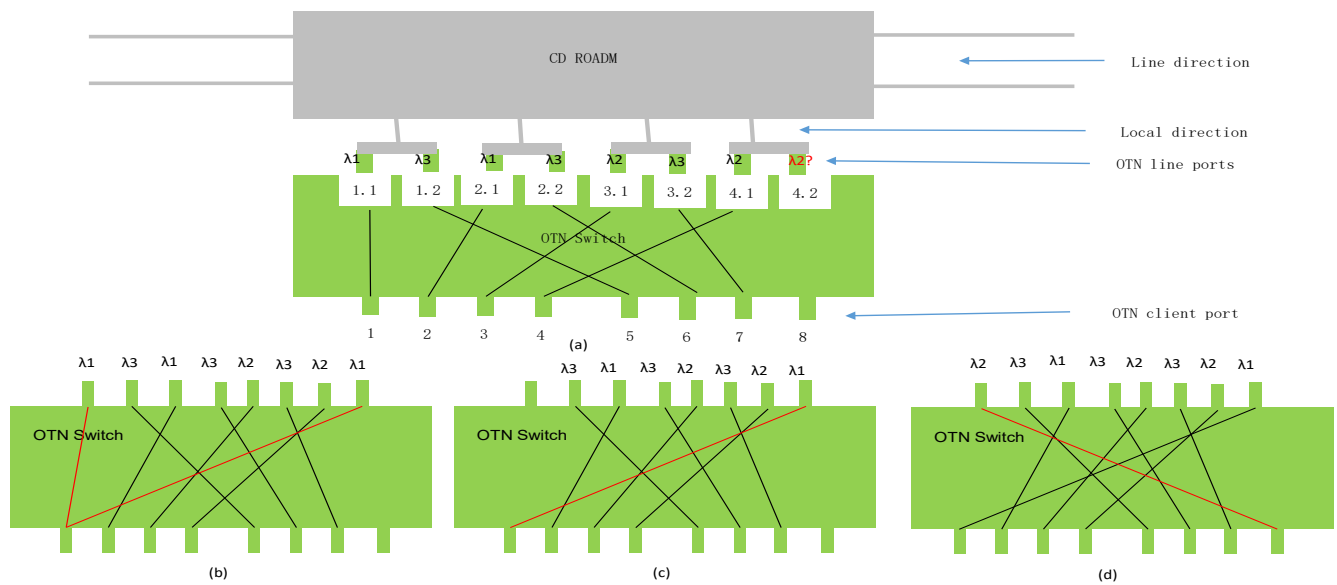


Figure 11: OTN over ROADM rerouting contention avoidance

1.1 with wavelength 1. This can be done according to section 6 observation 4

- ✓ Identify line port 1.1 associated client port: client port 1.
- ✓ Client port dual-cast to line ports 1.1 and 4.2, see figure 10.b
- ✓ CD-ROADM cross-connects 4.2 signal to the same ROADM line direction of line port 1.1
- ✓ ROADM line WSS of line port 1.1 selects signal from line port 4.2
- ✓ Client port cancels signal to line port 1.1, then line port 1.1 is free, see figure 10.c
- ✓ Client port 8 new service cross-connects to line port 1.1 with wavelength 2. See figure 10.d

7.2. Use case 2: failure rerouting contention blocking avoidance

Figure 11 shows the use case of service failure rerouting contention avoidance solution. Figure 11.a shows service OTN

internal configuration and wavelength assignment. When a service fails, wavelength rerouting restoration is activated. If rerouting wavelength is the same as original working service wavelength, then OTN configuration has no change, the only change is CD ROADM to cross-connect line port signal from original working line direction to rerouting line direction; If rerouting wavelength is different from original working wavelength, and the local direction of working wavelength has free rerouting wavelength, then the OTN configuration does not need to change, the only change is to tune the working line port to rerouting wavelength and ROADM to cross-connect the rerouting wavelength to rerouting line direction. However if rerouting wavelength is different from working wavelength, and the local direction of working wavelength has occupied the rerouting

wavelength, then we need to find out all local directions with free rerouting wavelength. If one of them has free line port, then we could tune this line port to the rerouting wavelength, ROADM cross-connect the rerouting wavelength to rerouting line direction, and cross-connects the service client port to the rerouting wavelength line port. The difficult case is when all local directions with free rerouting wavelength have no free line ports, we need to adjust at most one existing service configuration. Assume OTN configuration as in figure 11 (a) and client port 7 service fails and rerouting wavelength is wavelength 2. Then working wavelength local direction 3 has no free wavelength 2. After identifying all local directions having free wavelength 2, we find local direction 1 and local direction 2. However both local directions have no free line ports. We choose one local direction with free rerouting wavelength and this local direction at least has one occupied wavelength which is not used by the local direction of failed working wavelength, this can be done according to section 6 observation 4. In Figure 11.a, we choose local direction 2, where wavelength 3 is not used by local direction 3 (failed working wavelength is not used anymore). We reuse OTN line port 3.2, which is free currently and tune it into wavelength 3, dual-cast client port 6 service to both line port 2.2 and line port 3.2, (figure 11.b) cross-connect line port 3.2 signal to ROADM line direction of line port 2.2 connected line direction, and let this line direction WSS selects signal from line port 3.2, cancel signal cast from client port 6 to line port 2.2 and free line port 2.2 (figure 11.c), at last tune line port 2.2 to rerouting wavelength 2 and cross-connects client port 7 signal to line port 2.2, ROADM cross-connects line port 2.2 signal to rerouting line direction (figure 11.d)

8. Conclusion

In this paper, we investigated CD ROADM contention blocking problem. Since ROADM network operation is moving from static to dynamic, from 1+1 failure protection to dynamic rerouting restoration, CDC ROADM architecture has attracted lots of attention. However most deployed ROADM networks are still CD architecture. Is CD ROADM a show stop for dynamic operation? How much contention blocking will CD ROADM produce? If CD ROADM contention happens, is there any solution to solve this contention? We have answered all above questions using analytical model, simulation, and configuration method of procedures. From mathematic model and simulation results, we showed that CD ROADM did produce some contention during network operation, but the contention blocking is not as significant as we originally thought. If the service fill ratio at each local direction is 50% or less, the contention blocking is almost negligible. When contention did occurs during network operation, if service providers are using OTN as grooming and ROADM as wavelength transmission, as the often used architecture in field deployments, we provided a method to reconfigure at most one existing service to avoid the contention blocking. Based on our study, we concluded that a service provider with CD ROADM optical network can move forward to ASON (automatic

switched optical network) dynamic operation without worrying too much of CD ROADM contention blocking.

References

- [1] Mina Paik, "4 Things we learned at #OFC17", March 30, 2017, <http://www.ciena.com/insights/articles/4-Things-We-Learned-at-OFC17.html>
- [2] Fujitsu Network Communications Inc., "CDC ROADM applications and cost comparison", whitepaper, 2014, <https://www.ofcconference.org/getattachment/188d14da-88ba-4a63-91d6-1cc14b335d8b/CDC-ROADM-Applications-and-Cost-Comparison.aspx>
- [3] Nokia, "Benefits of CDC-F ROADMs", whitepaper, 2017, <https://knect365.com/ngon/article/e081ea00-9bc9-433d-ae21-94617299be41/whitepaper-benefits-of-cdc-f-roadms>
- [4] Huawei Technologies Co., Ltd., "white paper on technological developments of optical networks", whitepaper, 2016, <http://www-file.huawei.com/-/media/CORPORATE/PDF/white%20paper/White-Paper-on-Technological-Developments-of-Optical-Networks.pdf>
- [5] Sterling Perrin, "Building a fully flexible optical layer with next-generation ROADMs", heavy reading whitepaper, 2011, <http://www-file.huawei.com/-/media/CORPORATE/PDF/white%20paper/White-Paper-on-Technological-Developments-of-Optical-Networks.pdf>
- [6] B. C. Collings, "Advanced ROADM technologies and architectures," Los Angeles, CA, Paper Tu3D.3, OFC 2015.
- [7] S. Poole, S. Frisken, M. Roelens, and C. Cameron, "Bandwidth-flexible ROADMs as network elements," Los Angeles, CA, Paper OTuE1, OFC 2011.
- [8] Guangzhi Li, Kerong Yan, Li Huang, Bin Xia, Fanhua Kong, Yang Li, "How much is CD ROADM contention blocking?", San Diego, CA, OFC 2018.
- [9] L.Zong, H. Zhao, Z. Feng, and S. Chao, "Demonstration of ultra-compact contentionless Roadm based on flexible wavelength router," ECOC 2014.
- [10] M. Feuer, S. Woodward, P. Palacharla, X. Wang, I. Kim, and D. Bihon, "Intra-node contention in dynamic photonic networks", JOLT, 29(4), Feb. 2011.
- [11] Y. Li, L. Gao, G. Shen, and L. Peng, "Impact of ROADM colorless, directionless, and contentionless (CDC) features on optical network performance," JOCN, 4(11), Nov. 2012.
- [12] J. Simmons, "a Closer look at ROADM contention," IEEE Comm. Mag, 55(2), pp 160-166, Feb 2017.
- [13] F. Naruse, Y. Yamada, H. Hasegawa, and K. Sato, "Evaluations of OXC hardware scale and network resource requirements of different optical path add/drop ratio restriction schemes," J. Opt. Commun. and Netw., 4(11), pp B26-B34, Nov. 2012.
- [14] T. Zami, P. Jenneve, and H. Bissessur, "Fair comparison of the contentionless property in OXC," Asia Commun. and Photonics Conf., Hong Kong, AM3G.3, 2015.
- [15] J. Pedro and S. Pato, "Impact of add/drop port utilization flexibility in DWDM networks," JOCN, 4(11), Nov 2012.
- [16] C. Clos, "A study of non-blocking switching networks," BSTJ, 32(2), pp 406-424, March 1953.
- [17] C. Rigault, "Clos networks: a correction of the Jacobaeus result," Analysis of telecommunications, 57(11-12), pp 1244-1252, Nov 2002.

Real Time Eye Tracking and Detection- A Driving Assistance System

Sherif Said^{1,2,*}, Samer AlKork¹, Taha Beyrouthy¹, Murtaza Hassan¹, OE Abdellatif², M Fayek Abdraboo²

¹College of Engineering and Technology, American University of the Middle East, Kuwait

²Shoubra Faculty of Engineering, Benha University, Shoubra, Cairo, Egypt

ARTICLE INFO

Article history:

Received: 31 October, 2018

Accepted: 05 December, 2018

Online: 19 December, 2018

Keywords:

Eye Tracking System

Viola-Jones

Haar Classifiers

Smart systems

Driver's safety

ABSTRACT

Distraction, drowsiness, and fatigue are the main factors of car accidents recently. To solve such problems, an Eye-tracking system based on camera is proposed in this paper. The system detects the driver's Distraction or sleepiness and gives an alert to the driver as an assistance system. The camera best position is chosen to be on the dashboard without distracting the driver. The system will detect the driver's face and eyes by using Viola-Jones Algorithm that includes Haar Classifiers that showed significant advantages regarding processing time and correct detection algorithms. A prepared scenario is tested in a designed simulator that is used to simulate real driving conditions in an indoor environment. The system is added in real-vehicle and tested in an outdoor environment. Whenever the system detects the distraction or sleepiness of the driver, the driver will be alerted through a displayed message on a screen and an audible sound for more attention. The results show the accuracy of the system with a correct detection rate of 82% for indoor tests and 72.8 % for the outdoor environment.

1. Introduction

The major factor of driving on roads is the driver's attention, and once this attention is lost, major accidents could happen. According to the National Highway Traffic Safety Administration (NHTSA) [1], 153297 car crashes happened due to drowsiness in the period from 2011 to 2015. Another study by Kuwait times [2] stated that for more than 80,000 accidents, 95% of them happened due to lack of attention. The attention of the driver can be diverted through many things; using mobile phones, changing radio stations, eating and drinking, and daydreaming. In addition to that, sleepiness due to stress or fatigue; when the driver is sleepy or tired, his reaction will be slower than the normal driver which leads to accidents. There are many symptoms that can help detect sleepiness or distraction of the driver, the main symptom is the eyes of the driver. One of the ways that will assist the driver to pay attention while driving is to add an eye-tracking system that uses a camera to detect sleepiness due to stress, fatigue or any distraction. The system will alert the driver when his attention is distracted. This system is to be added to the wearable bracelet, which is used to monitor the physiological parameters of the driver [3].

2. Objectives

Real-time eye-tracking system is used to track the driver's eye. When the driver is drowsy or distracted his response time to

react in different driving situations is slow. Therefore, there will be higher possibilities of accidents. There are three ways of detecting driver's drowsiness. The first one is the physiological changes in the body like pulse rate, brain signals and heart activity which can be detected by a wearable bracelet system. The second way is behavioural measures for example sudden head nods, eye closure, blinking, and yawning which is achieved by the proposed eye tracking system. The third way is vehicle based like lane position and steering wheel movements. Based on literature study conducted, the eye tracking system is the most accurate and precise way to detect drowsiness and fatigue. [4] In addition to that, it is used to detect the driver's attention on road which might happen due to texting on mobile phone, changing radio station or chatting with passengers. The paper revolved around the design of the eye tracking system. The system consists of a camera to track the driver face and detect the eyes and interactive screen for user interface with the system. Different locations are studied, and the best location is chosen and tested, the system is tested on a designed simulator. The simulator is consisting of a steering wheel and pedals as a cockpit. A robotics rover is controlled via RF signals as a vehicle in which is controlled manually using the simulator wheels and pedals. In case of eyes closing detection, the interactive screen alerts the drivers by a message and hearable sound. The robustness of the system is studied, and false detection rate is detected in in-door and out-door testing environments.

*Sherif Said, American University of the Middle East, +965 2225 1400 Ext. 2168, sherif.said@aum.edu.kw

www.astesj.com

<https://dx.doi.org/10.25046/aj030653>

The objective of the designed system aims the following five points:

- **Affordable:** The systems must be affordable as the price is one of the main factors that kept on mind during design phase.
- **Portable:** The systems to be portable and easy to install in different vehicles models.
- **Safe:** The safety of the system is achieved by choosing the appropriate location for each component.
- **Fast:** The response and processing time to react in case of driver's emergency is one of the keys factors since the accident happens in few seconds.
- **Accurate:** The system must be accurate; therefore, the most accurate algorithms have been chosen.

3. Literature Review

Distracted driving is a serious and growing threat to road safety [5]. Collisions caused by distracted driving have opened an investigation of the US Government and professional medical organizations [6] during the last years. There is not an exact figure regarding statistics about accidents caused by inattention (and its subtypes) since studies are made in different places, different time frames and therefore, different conditions. The distraction and inattention account for somewhere between 25% and 75% of all crashes and near crashes.

The use of in-vehicle information systems (IVISs) trend is critical [7] due to the fact they manual, induce visual, and cognitive distraction may have an effect on performance of driving in qualitatively distinct ways. Moreover, the advancement and incidence of personal communicate gadgets has exacerbated the problem during these remaining year's [8]. Some of these elements can result in the increment of the wide variety of obligations subordinate to riding pastime. Those obligations, particularly secondary tasks, which may additionally cause distraction [9], include drinking, eating, tuning the radio or the act of taking something or the use of cell phones and other technologies.

The secondary duties that take drivers' eyes off the forward roadway lessen visual test [10] and growth cognitive load can be specifically risky. For instance, the use of mobile phones even as driving, consistent with naturalistic studies [11], causes heaps of fatalities inside the US each year [12].

To tune the irradiation of IR illuminators many developments have been done recently. To let IR illuminators, operate in multiple glass reflection, distinct lighting conditions, and varying direction of gaze must be tuned. Research has been conducted to combine appearance-based methods with active IR methods. The advantage of this combination will be that, this method can do eye tracking even when the pupils are not bright due to different interferences from external illumination. Along this model, an appearance model is also incorporated with the use of shift mean tracking and vector machine support in both eye tracking and detection [13].

There are two types in which eye tracking and detection can be classified namely, Active (infrared) IR based method and

method of passive appearance. A bright pupil effect is utilizing by active IR illumination method. This method is simple and effective for easy tracking and detection of eyes. The principle of working is a differential infra-red scheme [14]. This method utilizes two infrared sources of frequencies. A distinct glow in the pupil is produced when the first image is captured at 850 nm by these infrared lights. 950 nm infrared source is used by the second image for illumination that displays a dark pupil's image. The first and the second image both are synchronous with the camera and the only difference exists on the pupil region brightness. Blobs are identified after post processing the pupil and works for eyes tracking [15]. Several factors affect the success rate. Pupil's size and brightness face orientation, interference of external light (lights from street lights and other vehicles) driver distance from camera. External light intensity should be limited. Another problem is the glints and reflection from glasses.

There are two steps involved in appearance-based methods which are: detecting face to extract eye region and after that detection of eye from eye windows. In order to overcome face detection method different approaches like: principal and independent components, neural network, and method of skin colour based. There are some constraints in each method: images without expression, frontal view, short changes in conditions of light, background that is uniform and so on. To track and locate eyes of the driver Papanikolopoulos and Eriksson present a system. The approach used was a symmetry-based approach. It is used to locate face in grayscale image and after that eyes are detected and tracked.

If the eyes are closed or open, this can be determined by template matching [16]. For the identification of driver fatigue, a system of non-intrusive vision-based system was proposed by Papanikolopoulos and Singh [17]. A graphics video camera is used by the system that focus on face of driver and continuously observe eyes of driver to detect micro sleeps.

Haar like feature can be used to detect face and Michael Jones and Paul Viola proposed this method. Using integral images Haar like features will be quickly computed. AdaBoost method is used to train this algorithm. Combination of different weak classifiers are used to form a strong classifier. The advantage of this technique is that, the combination will let detector work in cascaded manner and the first stage classifier faces like regions will be more intensively processed. The rate of detection with this method was above 95%. In a few processes authors used colour, facet, and binary records to come across eye pair candidate regions from input photograph, then extract face candidate location with the detected eye pair. SVM (Support Vector Machines) are used to detect region of candidate face and eye pair [18]. There are approaches in which eye's dimensions and appearance are used. Mostly tracking and detection of eye's pupil or iris is required by eye tracking applications. Iris and pupil can be modelled by five shape parameters, as they both appears to be elliptical depending on the viewing angle. In order to detect iris or pupil, Hough transform is the effective way, but it requires detection of explicit features.

There are four stages in Viola-Jones algorithm, consisting of integral image, Haar features, cascade and AdaBoost [19]. To begin with, Haar features detection can be summarized in three steps. First, convert the colored, RGB, images to grey-scale images. After converting the image to grey-scale, integral images

are generated. The integral image is a summation of the pixel values of the original image. The summation of the pixels are calculated by choosing a coordinate (x,y) and summing all the values to the top-left of the point including the point itself. Second, from different levels of integral images the Haar wavelets are obtained. Haar-like features are a combination of two or more rectangles that have different contrast [20]; they can be two-rectangle, three-rectangle, or four-rectangle, and so on. The third step is the AdaBoost, which is an algorithm that looks at relevant features and is used to detect objects, in this case the face and eye. When the features are extracted, a stage comparator will sum all the features and compares them with the threshold. After these steps are done, there is the cascade step; and in this step the stages are cascaded to eliminate any non-face candidates to improve the computation time an accuracy of the detection.

One of the algorithms that were used to detect the pupil of the driver's eye is the Circular Hough Transform [21]. This algorithm detects any circular object in the images, in these cases the pupil of the eyes. The first step is defining the radius or the average radius of the human's pupil, two circle equations are needed, one for each eye. The algorithm will first filter the image to reduce the noise, and then the image will be converted into a grey-scale image. The edge of the eyelids will be detected to look at the circles between the two eye lids, which are the pupils.

A correlation matching algorithm, which is used only on raw images, is used to identify the targeted position, which is the eye, and then tracking the motion of the detected eyes. The reason for using only raw images is because all the information of the images is kept, therefore, making the process easier. First, there is a source image which is the original image of the driver. Second, the template image is the image that will be compared to the source image. To compare the source and template images, an equation is used where the matrices of the two images are compared together to find the similarities between them by using the Normalized correlation equation

Another method that is used and based on Digital Image Processing algorithms is the opened and closed eye template. The eyes of the driver are continuously scanned to determine whether the driver is drowsy or not. If three consequent frames were captured and the eyes were found to be closed, the system will state the driver is drowsy.

In [22], the author that uses DIP algorithms and continuous eyes monitoring to detect drowsiness of driver. The best indicator of fatigue state is the Micro sleeps as they are for the short period of almost 2 to 3 seconds of sleep. This presented system contains two cameras namely: narrow and wide angle camera. Narrow angle camera focuses on the eyes and monitors gaze and eyelid movements, while wide angle camera focuses on face and monitors facial expression and head movement. If some abnormal actions are detected by the system, it gives alarm warnings and reduces the speed of the system. This system not only detect the drowsiness of the driver but also detect the distance of the car from objects or other cars using Ultrasonic sensor. When these sensors detect objects they warn the driver and reduces speed.

The above algorithms are implemented using several software's including GUI (MATLAB), and Python (OpenCV). So briefly, the algorithms will work as follow: the camera will capture an image, after that, the algorithm will scale and extract features from the image, and then the algorithm will classify features.

Many driver-monitoring systems have been developed while working on driver fatigue detection. All these systems focus on providing information to drivers that will facilitate their driving and increase traffic safety. These can be divided into systems, DAISY (Driver Assisting System) and DAS (Driver Assistance System). Driver assistant system DAISY is used in German motorways to warn and monitor driver for lateral and longitudinal control [23]. The second system DAS is developed by a group at Australian National University [24][25]. To monitor the driver, a dashboard mounted face LAB head and eye tracking system is used. Driver performance is detected by algorithm known as distillation algorithm. Driver is provided with the feedback on deviation in lane tracking by using steering wheel force feedback that is related to offset lateral estimation by the tracker of lane.

4. Methodology

The driver's distraction and sleepiness while driving can be detected by adding a camera on the vehicle dashboard to provide a real-time tracking of the driver's face. In that case, the task of detecting the driver's drowsiness, fatigue and distraction will be much easy and more accurate.

After the literature study about the systems, algorithms and methods used for eye-tracking system. Some important factors were noticed that help in the development of the proposed system in this paper. The average response time between the analysis and the alert process to be 50 milliseconds. The alarm should be loud enough to alert the driver. The system should operate even when the driver is using sunglasses and at night without flash to ensure non-intrusiveness. Different methods of detection are mentioned in Table 1 Methods of eye tracking systems.

Table 1. Different systems with accuracy

Method	TN	FP	FN	TP	Overall Accuracy
Blinking	91.72%	8.18%	10.95%	89.05%	90.74%
Lateral Position	89.39%	10.61%	20.95%	79.05%	85.37%
Steering Angle	88.48%	11.52%	14.77%	85.23%	87.22%

Where:

- TN: Eyes closed system, detected closed
- FP: Eyes closed, system detected open
- FN: Eyes open, system detected closed
- TP: Eyes open, system detected open

This means that available systems are most accurate in eye tracking with detection of blinking rate, and the least accurate method was the lateral position. Therefore, based on these results, the best method to detect drowsiness is by eye-detection.

To sum up, the systems were able to detect sleepiness and distraction for almost every eye-type and for both genders. Also, the most accurate method to detect driver's distraction is through

face and eye-detection. By using software, these researches were tested then implemented successfully, and by using the display screen and the buzzer, the driver is alerted and this will help in reducing the number of car accidents and increase the safety on roads.

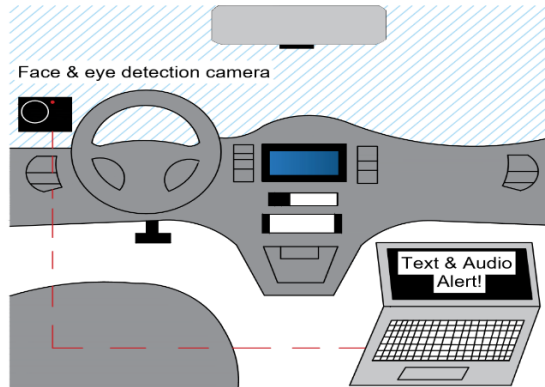


Figure 1. Proposed Eye-tracking System Sketch

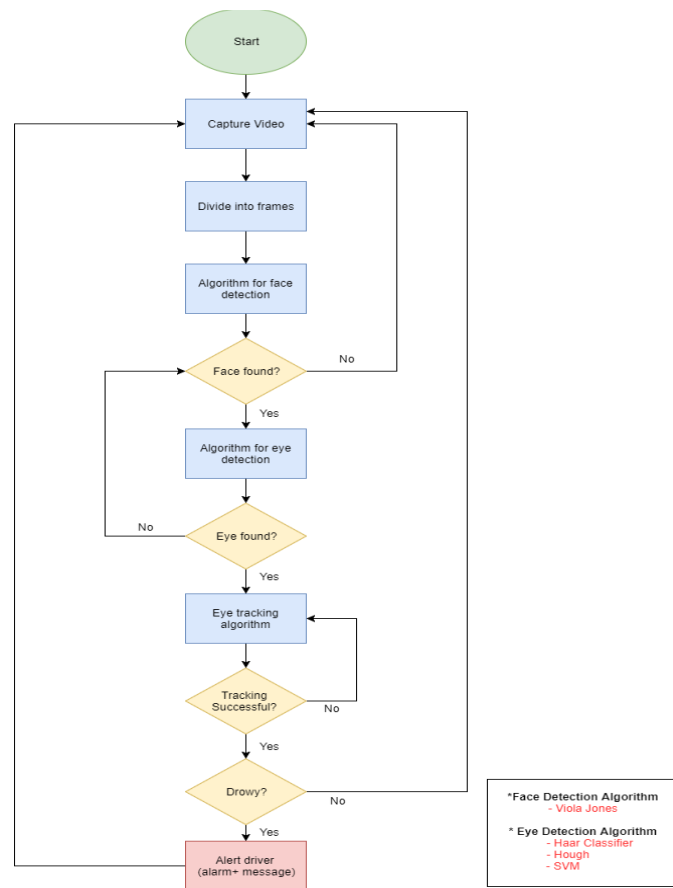


Figure 2. Flow-chart of the eye-tracking process

5. System Components

The proposed system consists of a high-resolution camera, an LCD, a speaker, and a microprocessor. The screen to be placed on the dashboard to the right-hand side of the driver. The camera will be positioned on the dashboard to the left-hand side of the driver and will be fixed by a camera holder. To start the system, the camera will be continuously recording a video of the driver and this video will be analyzed on PC initially then replaced by

microprocessor later. The analysis will be done through many algorithms to detect the eyes of the driver. When the driver seems drowsy or distracted, the system will show an alert on the display screen among with a sound alert. Figure 2 shows the proposed design in a form of sketch for the system. The system flow chart that need to be implemented is explained in the flow chart. The real-time camera will stream an online video. An eye-tracking algorithm can be used to extract the features of the eyes. The classifier will play a role in the detection of open/close of the eye.

6. Design Specifications

The design of the real-time eye tracking is implemented to be an added-on solution that can fit in any vehicle. The purpose of the system is to alert the driver in case of drowsiness and distraction. The system will alert the driver as precaution event. In emergency cases, the system should interact with the vehicle and take over the control. This system is an added value to the wearable bracelet that couldn't detect the sleepiness and distraction.

6.1. High Level Design

A simulator cockpit is designed and implemented in a professional way for testing the system. A steering wheel among with pedals are developed in an ergonomic way to simulate driving scenarios on road.

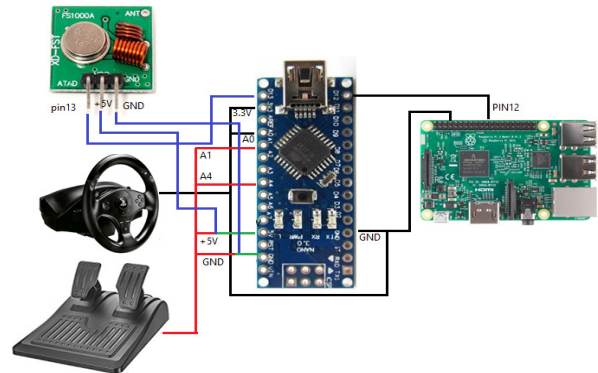


Figure 3. Connection of Raspberry Pi and Arduino Nano (Steering wheel)

6.2. Low Level Design

For the Low-Level design, the connections are made with Arduino and Raspberry Pi (see **Error! Reference source not found.**). The Raspberry Pi pin 12 is connected to the input pin 12 of the Arduino Nano that is attached to the steering wheel. In addition, the RF transmitter, the steering wheel and the pedals are all connected to the Arduino Nano as shown. The RF communication module is used to overcome the delay problem that happened when the control signals sent via Bluetooth. An Arduino Uno is attached to the rover (see **Error! Reference source not found.**), which shows the connection of the RF receiver, the LED, used to show that the rover is moving in autonomous mode, and the push button that the user must press to return to manual driving.

7. System Hardware Specifications

7.1. Camera

The Camera used in detection of the eye open/close is Logitech C290. The specifications of the camera are presented in

Table 2. The most important factor in selecting the camera is the number of frames that the camera can capture in one second. For this camera is 30 frames per second (fps).

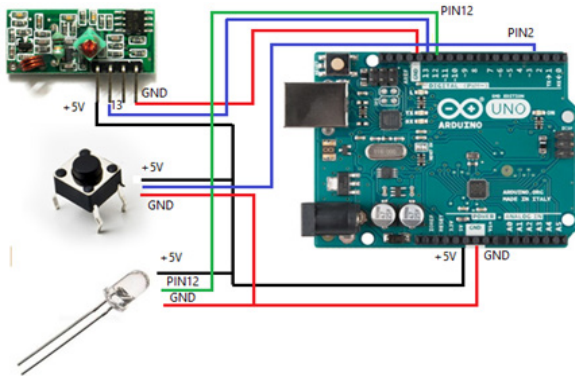


Figure 4. Connections of Arduino Uno (Rover)

Table 2. Specifications of the Camera

Device Type	Web camera
Connectivity Technology	Wired
Digital Video Format	Colour
Max Digital Video resolution	H.264
Features	1080p Full HD movie recording,
Battery	None
Audio support	Yes
Computer interface	USB 2.0
OS required	Microsoft Windows 7, Microsoft Windows Vista, Microsoft Windows XP SP3 or later
Connector type	4 pin USB Type A

Table 3. Raspberry Pi3 Specs

SoC	BCM2837
CPU	Quad Cortex A53 @ 1.2GHZ
Instruction set	ARMv8-A
GPU	400MHz VideoCore IV
RAM	1 GB SDRAM
Storage	Micro-SD
Ethernet	10/100
Wireless	802.11n/ Bluetooth 4.0
Video Output	HDMI/ Composite
Audio Output	HDMI / Headphone
GPIO	40

7.2. Raspberry Pi 3

Raspberry Pi 3 Is a tiny credit card size computer. Just add a keyboard, mouse, display, power supply, micro SD card with installed Linux Distribution and you'll have a fully-fledged computer that can run applications from word processors and spreadsheets to games.

7.3. Display & Audio 1280x800 IPS Screen

This screen has excellent resolution (1280x800) and IPS display so it is bright, crisp and looks good from any angle. This Screen even has HDMI audio support and can drive two 4-ohm speakers directly.

Table 4. IPS Screen Specs

Dimensions of Screen	105mm x 160mm x 3mm / 4.1" x 6.3" x 0.1"
Weight of Screen	91g
Power of Screen	9 VDC
Display Ratio	16:10
Resolution	1200 x 800
Visible Area	150mm x 95mm 16:10
Display Dimensions	162mm x 104mm x 4mm (6.4" x 4.1" x 0.2")
Brightness	400cd/m2
Contrast	800:1
Display	HSD070PWW1
Weight	290g/10.2oz
HDCP	None

8. Image Processing

For implementation of the system. A simulation of the eye tracking system is done using PC. The software used is OpenCV. For face and eye detection using Viola Jones and Haar Cascade Classifier algorithms were applied. The idea is to detect eyes are open or closed. When the eyes were closed, the system attempted to alert by showing a display message and sound. After the proof of concept, Raspberry Pi board as a main Processor for the system replaces the PC.

The basic idea of detecting the state of the driver revolves around the detection of the eyes. The recent advancement in the image-processing field, there are now multiple real-time methods that could enable the detection and tracking of multiple objects. Bag-of-Words models, Histogram-of-oriented gradients (HOG), Deformable Parts Models, Exemplar models and Viola-Jones are some of the widely used methods for object detection.

Viola-Jones method is one of the most widely used method of object detection. The main feature that makes this method so popular is its ability training is slow but detect fast. Haar basis feature filters are used in this algorithm avoiding, thus avoiding multiplications.

Detection happens inside a detection window. The minimum and maximum window size is selected, and for each size a sliding

step size is selected. The detection window moves across the selected image as follows:

Cascade -connected classifiers contained inside each face recognition filter. The classifier takes care of scanning at a rectangular subset of the detection window and recognize if it looks like a face or not. The next classifier is applied in case the face detected on the off chance that all classifiers give a positive answer, at that point the filter gives a positive answer and the face is perceived. If nothing detected, the next filter in the set of N filters should execute.

Each classifier is composed of Haar feature extractors (weak classifiers). Each Haar feature is the weighted sum of 2-D integrals of small rectangular areas attached to each other. The weights may take values ± 1 . Haar feature extractors are scaled with respect to the detection window size.

Viola-Jones algorithm consists of four stages, Haar features, integral image, AdaBoost, and cascade. To begin with, Haar features detection can be summarized in three steps. First, convert the colored, RGB, images to grey-scale images. Then, integral images are generated. The integral image is a summation of the pixel values of the original image. The summation of the pixels are calculated by choosing a coordinate (x,y) and summing all the values to the top-left of the point including the point itself. Second, different levels of integral images are used to obtain the Haar wavelets. Haar-like features are a combination of two or more rectangles that have different contrast; they can be two-rectangle, three-rectangle, or four-rectangle, and so on. The third step is the AdaBoost, which is an algorithm that looks at relevant features and is used to detect objects, in this case the face and eye. When the features are extracted, a stage comparator will sum all the features and compares them with the threshold. After these steps are done, there is the cascade step; and in this step the stages are cascaded to eliminate any non-face candidates to improve the computation time an accuracy of the detection.

8.1. Algorithm Implementation

The Algorithm is applied sequential. The first step is to detect the face. If no faces detected, this means that the driver is looking to the right or left or down. A counter start counting, and an alert message is sent after 1 second. If faces detected, then the algorithm will extract the eyes from the face. If the driver's eyes opened don't count any values. In case the driver closed his eyes then a counter start counting based on the required timing. After that time an alert message is sent to driver's. Figure 5 shows the application of the algorithm as a simulation of a detected eye.

9. Results

Real-time Eye tracking came up as an added-solution to the wearable biosensors bracelet. This solution helps in an accurate real-time monitoring of the driver's eyes while driving. This system aims to assist the driver not only in case of drowsiness and in case of Fatigue; it can assist in case of driver's distraction due to changing radio stations, chatting with passengers or texting while driving.

The system implemented as a cockpit for testing in-door to validate the system. The system is installed and tested in a real-vehicle to proof the correctness of the proposed system.

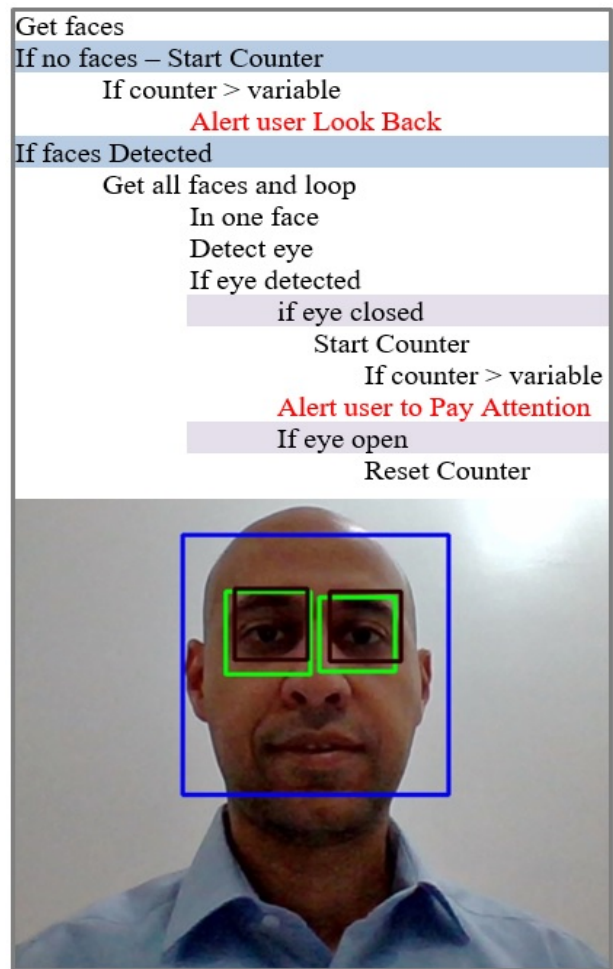


Figure 5. Eye Detection using Viola-Jones Algorithm

9.1. Indoor-System Test

The High-Level design of the simulator shown below (see **Error! Reference source not found.**), where the driver is seated in front of the steering wheel. The steering wheel gives the subject the feeling of controlling a vehicle. The driver is facing a camera and the display screen. In addition, there are pedals that control the rover when the driving mode is manual. One pedal is used to move the car forward and the other one moves it backwards.



Figure 6. Simulator Cockpit Design

The system tested in-door on 20 persons to record a small data-set for analysis and validation of the system. The system will be coupled to the prototype explained earlier, where the driver sat down in front of the steering wheel and imitated distraction or sleepiness as shown in figure. 6. The subject is controlling the robotics rover as he/she is driving a vehicle on roads. In case the user is distracted for 1.5 seconds (pre-set value), the system will respond by a message on the screen Look back! Moreover, the system reacts by switching the rover mode to autonomous in which the rover will run obstacle avoidance algorithm as a simulation of cars and obstacles on roads. In case, the subject closed his/her eyes for 1 second (pre-set value), the system will react by switching the rover mode to autonomous mode.

The recorded data of the subjects for in-door testing is shown in Table 5 and the graph is shown below (see **Error! Reference source not found.**). The results of average correct detection rate are 82 %. This result is very promising to test the system in real vehicle.

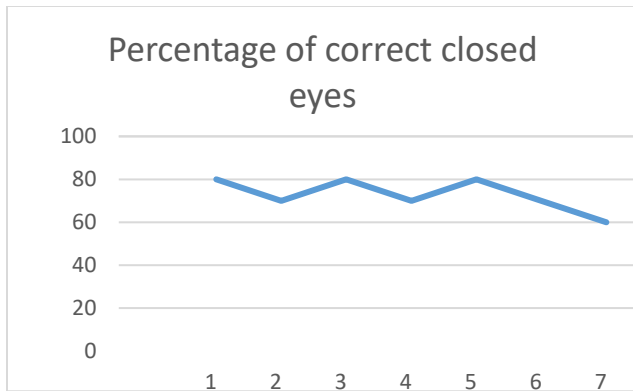


Figure 7. Correct Detection Percentage

Table 5. In-door Real-time Eye tracking Recorded Data

Index	Age	Gender	Samples	Correct Detections	False Detections	Percentage
1	20	M	10	8	2	80
2	23	M	10	7	3	70
3	26	M	10	8	2	80
4	21	M	10	9	1	90
5	36	M	10	10	0	100
6	25	M	10	7	3	70
7	45	M	10	7	3	70
8	36	M	10	8	2	80
9	24	M	10	9	1	90
10	62	M	10	10	0	100
11	45	F	10	8	2	80
12	52	F	10	7	3	70
13	42	F	10	9	1	90
14	35	F	10	10	0	100
15	26	F	10	7	3	70
16	29	F	10	8	2	80
17	27	F	10	7	3	70
18	28	F	10	8	2	80
19	39	F	10	7	3	70
20	34	F	10	10	0	100

			Average Correctness	82
--	--	--	---------------------	----

9.2. Out-door System Test

The system is installed in two different car models for real driving conditions testing as shown in below (see **Error! Reference source not found.**). Once, the system detects eye-closing for 1 second (pre-set value) or distraction detection for 1.5 second (pre-set value) an audible alert sound among with alert message sent to the user for driver’s assistance on road.



Figure 8. Installation of the System in real-vehicles

The system tested on seven subjects. The testing scenario is to drive the car at speed around 40-50 Km/hr. The driver will act based on the assistant instructions. The assistant task is to direct the driver to close eye or distract the eyes by looking right/left and record the data of correct detection to validate the algorithm and system on roads. The recorded data is shown in **Error! Reference source not found.** and the results of average correct detection rate of closed eyes is 72.8 % with an average correct detection rate of distracted eyes of 67.14 % as shown below (see **Error! Reference source not found.**).

As noticed, the results of correct detection reduced compared to the in-door test results. The reasons of this decrease are that the car on the streets is exposed to different lighting conditions that needs a pre-setting. Also, the driver movements in the real-driving conditions increases which needs higher frame rate of camera.

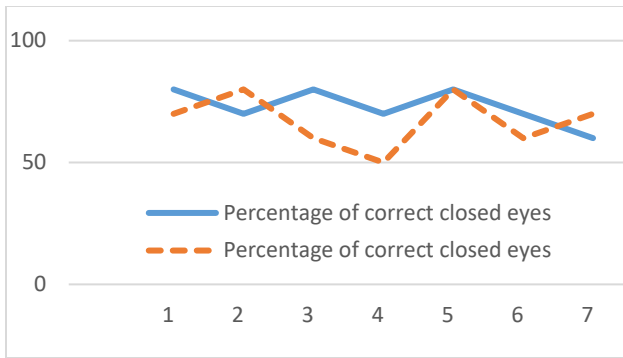
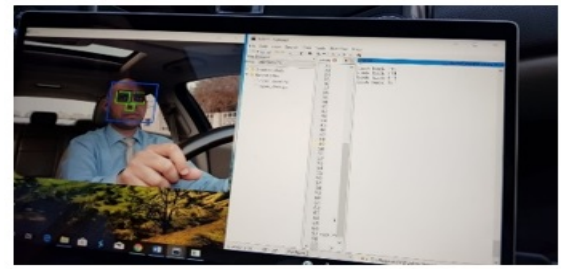


Figure 9 Correct Detection of eye closed and distracted eyes in out-door testing environment

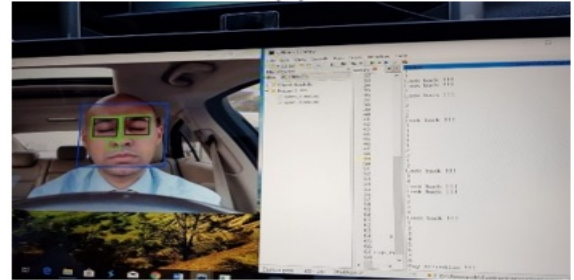
Table 6. Out-door Real-time Eye tracking Recorded Data

Index	Age	Gender	Samples	Correct Detections of closed eyes	Correct Detections of Distracted eyes	Percentage of correct closed eyes	Percentage of correct distracted eyes
1	30	M	10	8	7	80	70
2	25	M	10	7	8	70	80
3	21	F	10	8	6	80	60
4	22	F	10	7	5	70	50
5	24	F	10	8	8	80	80
6	20	F	10	7	6	70	60
7	22	F	10	6	7	60	70
				Average Correctness		72.85 7143	67.142 857

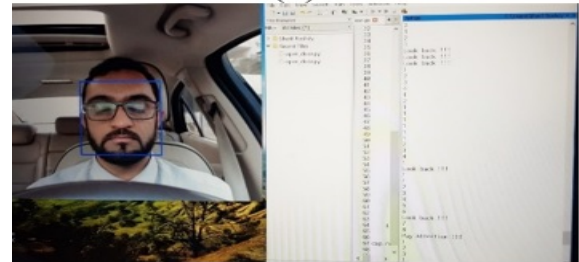
The system testing out-door is shown below (see **Error! Reference source not found.**). The system is tested with the PC to record the system test in out-door environments. The system is validated, and the results were promising to improve the algorithm for more scenarios.



(a)



(b)



(c)

Figure 10. (a) Distracted driver, (b) closed eyes driver, (c) closed eyes driver

10. Conclusion

The complete system is implemented with a developed simulator cockpit for in-door system validation and data collection purposes. The measured average correct detection of the system is 82 %. This system is tested in subjects in average of 33 years old and showed no issues regarding the age. It is valid for people wearing eyeglasses without any problem. In out-door environment, the results of average correct detection rate of closed eyes are 72.8 % with an average correct detection rate of distracted eyes of 67.14 %.

11. Future Works

- The system can't detect during night due to lack of camera version
- The response time of the system when the driver closes his/her, eyes should be a function of vehicle speed.
- The system should be adapted to the looking of mirrors scenario.
- Sensor fusion of the camera with all the wearable biosensors.
- Use Night-vision camera to detect the eyes at night.
- Use better specifications camera will increase the efficiency of the system

References

- [1] <https://www.nhtsa.gov/>, 4/2017.

- [2] "Kuwait Times," 15 10 2017. [Online]. Available: <http://news.kuwaittimes.net/website/>.
- [3] Sherif Said ; Samer AlKork ; Taha Beyrouthy ; M Fayek Abdrabbo, "Wearable bio-sensors bracelet for driver's health emergency detection," in Biosmart ,IEEE, Paris, France, 2017.
- [4] Singh Himani parmar ; Mehul Jajal ; Yadav priyanka Brijbhan, "Drowsy Driver Warning System Using Image Processing," Nternational Journal Of Engineering Development And Research , 2017.
- [5] "http://www.who.int," 2016. [Online].
- [6] Llerena, L.E.; Aronow, K.V.; Macleod, J.; Bard, M.; Salzman, S.; Greene, W.; Haider, A.; Schupper, A. J. , " An evidence-based review: Distracted driver.," *Trauma Acute Care Surg.*, vol. 78, p. 147–152, 2015.
- [7] Bennakhi, A.; Safar, M. , "Ambient Technology in Vehicles: The Benefits and Risks," *Procedia Comput. Sci.*, vol. 83, pp. 1065-1063, 2016.
- [8] Liu, T.; Yang, Y.; Huang, G.B.; Lin, Z, "Detection of Drivers' Distraction Using Semi-Supervised Extreme Learning Machine," in In Proceedings of ELM-2014; Springer, Berlin, Germany, 2015.
- [9] Simons-Morton, B.G.; Guo, F.; Klauer, S.G.; Ehsani, J.P.; Pradhan, A.K, "Keep your eyes on the road: Young driver crash risk increases according to duration of distraction," *J. Adolesc. Health*, vol. 54, p. S61–S67, 2014.
- [10] Recarte, M.A.; Nunes, L.M., " Mental workload while driving: Effects on visual search, discrimination, and decision making," *J. Exp. Psychol. Appl.* , vol. 9, pp. 119-137, 2003.
- [11] Klauer, S.G.; Guo, F.; Simons-Morton, B.G.; Ouimet, M.C.; Lee, S.E.; Dingus, T.A, "Distracted driving and risk of road crashes among novice and experienced drivers," *N. Engl. J. Med.*, vol. 370, p. 54–59, 2014.
- [12] Bergmark, R.W.; Gliklich, E.; Guo, R.; Gliklich, R.E. , "Texting while driving: The development and validation of the distracted driving survey and risk score among young adults.," *Inj. Epidemiol.*, 2016.
- [13] Xia Liu, "Real-time eye detection and tracking for driver observation under various light conditions," *Intelligent Vehicle Symposim*, 2002.
- [14] R. Grace, "Drowsy driver monitor and warning system," in *Proc. Int. Driving Symp. Human Factors in Driver Assessment, Training and Vehicle Design*, 2001.
- [15] Z. G. Yuan, "A real-time eye detection system based on the active IR illumination".
- [16] Eriksson, and N.P. Papanikotopoulos M., "Eye-tracking for detection of driver fatigue," *Proc. Int. Conf. Intelligent Transportation Systems*, pp. 314-318, 1997.
- [17] S. Singh, and N.P. Papanikolopoulo S., "Monitoring driver fatigue using facial analysis techniques," in *Proc. Int. Conf. Intelligent Transportation Systems, Tokyo*, 1999.
- [18] Hyungkeun Jee, Kyunghee Lee, and Sungbum Pan, "Eye and Face Detection using SVM," in *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*, Melbourne, Australia, 2005.
- [19] Jay D. Fuletra ; Viral Parmar, "Intelligent Alarm System for Dozing Driver using Hough transformation," *IJEDR*, vol. 2, no. 2, pp. 2797-2800, 2014.
- [20] Anirban Dasgupta, Anjith George, S. L. Happy, and Aurobinda Routray, "A Vision-Based System for Monitoring the Loss of Attention in Automotive Drivers," *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS*, vol. 14, no. 4, 2013.
- [21] Hari Singh ; Jaswinder Singh, "Human Eye Tracking and Related Issues: A Review," *International Journal of Scientific and Research Publications*, vol. 2, no. 9, 2012.
- [22] Mitharwal Surendra Singh L. ; Ajgar Bhavana G. ; Shinde Pooja S. ; Maske Ashish M. , "EYE TRACKING BASED DRIVER DROWSINESS MONITORING AND WARNING SYSTEM," *International Journal of Technical Research and Applications*, vol. 3, no. 3, pp. 190-194, 2014.
- [23] R. Onken, "DAISY: an adaptive knowledge-based driver monitoring and warning system," in *Proc. Vehicle Navigation and Information Systems Conf.*, 1994.
- [24] L. Fletcher, N. Apostoloff, L. petersson, and A. Zelinsky, "Vision in and out of Vehicles," *IEEE Trans. Intelligent Transportation Systems*, pp. 12-17, 2003.
- [25] F Sayegh, F Fadhli, F Karam, M BoAbbas, F Mahmeed, JA Korbane, S AlKork, T Beyrouthy "A wearable rehabilitation device for paralysis," 2017 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, 2017, pp.1-4.

Simulation-Optimisation of a Granularity Controlled Consumer Supply Network Using Genetic Algorithms

Zeinab Hajiabolhasani^{*1,2}, Romeo Marian², John Boland¹

¹*School of Information Technologies and Mathematical Sciences, University of South Australia, 5095, Australia*

²*School of Engineering, University of South Australia, 5095, Australia*

ARTICLE INFO

Article history:

Received: 29 August, 2018

Accepted: 05 December, 2018

Online: 20 December, 2018

Keywords:

Consumer Supply Network

Simulation-Optimisation

Granularity

Genetic Algorithms

ABSTRACT

The decision support systems regarding the Supply Chains (SCs) management services can be significantly improved if an effective viable method is utilised. This paper presents a robust simulation optimisation approach (SOA) for the design and analysis of a granularity controlled and complex system known as Consumer Supply Network (CSN) incorporating uncertain demand and capacity. Minimising the total cost of running the network, calculating optimum values of orders and optimum capacity of the inventory associated with each product family are the objectives pursued in this study. A mixed integer non-linear programming (MINLP) model was formulated, mathematically described, simulated and optimised using Genetic Algorithms (GA). Also, the influence of the problem's attributes (e.g. product classes, consumers, various planning horizons), and controllable parameters of the search algorithm (e.g. size of the population, crossover rate, and mutation rate) as well as the mutual interaction of various dependencies on the quality of the solution was scrutinised using Taguchi method along with regression. The robustness of the proposed SOA was demonstrated by a series of representative case studies.

1. Introduction

The main challenges affecting today's Supply Chains (SCs) are globalisation, environmental and technological turbulences and rapid changes in economy capacity. They have provoked companies to recognise that, in order to remain competitive in the global market, they need to gain more from their SCs.

Supply Chains are defined as links (relationships) between every unit (enterprise) in a manufacturing process from raw materials to customers. Traditionally, products were made and flowed to consumers through SCs. However, due to globalisation and complexity of the economy, today's SCs are better characterised as Supply Networks (SNs).

Consumer Supply Networks (CSNs) refer to complex networks consisting of sets of companies working in unison to supply, manufacture, distribute and deliver final products and services to end-users (Figure 1), being controlled by information flow.

CSNs are examples of industrial systems that are naturally large, complex, stochastic, and dynamic. These attributes translate into difficulties in representing the actual behaviour and in

planning, optimising and anticipating performance. Also, the combination of these attributes makes the choice of an appropriate solution methodology difficult at best, if not simply impossible at this point in time [1].

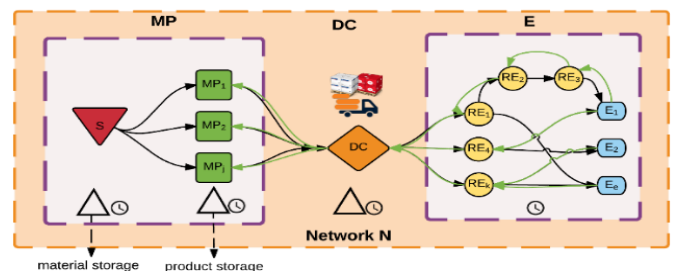


Figure 1 Three echelon Consumer Supply Network

Different methodologies have been utilised to solve this class of complex problem; simulation and optimisation methods are widely used to tackle such problems.

Simulation is a powerful tool for modelling, analysis, and validation of CSNs. However, its major disadvantage is that it will produce a very detailed analysis but strictly for a given

* Zeinab Hajiabolhasani, Email: zeinab.hajiabolhasani@unisa.edu.au

configuration. Simulation cannot change the configuration of the system, and any optimisation would be searching for the best combination of variables for a given system.

A recurrent, key issue when attempting to optimise CSN is the granularity of the model. An appropriate granularity – the size of the smallest indivisible unit (of product, part, flow, time, etc.) of the process – makes the difference between a successful implementation of the optimisation methodology and an algorithm that does not converge or gets consistently stuck in local optima. Additionally, the choice of the granularity of the model has to be easy to translate in practice – a purely theoretical solution that cannot be implemented in real life is of little help.

This paper is an extension of the work initially has been presented in Intellisys Conference [2] in which a unique simulation optimisation approach (SOA) within an integrated methodology was developed. A small-scale Multi-Period, Multi-Product Consumer Supply Network (MPMPCSN) model, using mixed integer non-linear programming (MINLP) was designed. Then, the optimum quantity of orders was determined incorporating GA optimisation algorithm which simultaneously results in the total inventory cost minimisation. This way, the unique advantages of simulation were incorporated with optimisation method and higher quality solutions were achieved. Also, the quality of the solutions that were obtained by the proposed framework was checked by fine-tuning of the search algorithm's parameters combining the simulation model with the Taguchi method. Hence, in this study, a series of computational trials on realistic test problems are designed and analysed to demonstrate the generalisability of the proposed SOA for problems of similar size at different granularity levels.

The rest of this article is organised as follows: Section 2 is devoted to reviewing modelling methodologies that were used to solve CSNs problems. Section 3 presents the proposed MINLP model. Section 4 provides details about granularity. The optimisation module of the SOA methodology is described in Section 4. The numerical examples are given in Section 6 and discussed in Section 7. Section 8 concludes the paper.

2. Literature Review

A number of potential solution methods for the class of problems of similar size and complexity have been developed in the literature ranging from classical mathematical programming to hybrid and systematic methods [1, 3].

Optimisation methodologies combined with mathematical models are mainly contributed to solutions validation. A stable optimal solution can be obtained by a given objective function subject to several constraints. However, they are unable to provide the gradient of design space over time [4]. The extent of the optimisation problem cannot be expanded beyond a certain limit as the complexity of the problem adversely affects the computational costs which make less efficient and less practical [5]. This concern can be addressed by using, simulation methodologies.

Simulation models can deal with all attributes of CSNs problems which makes them a powerful analytical tool in this area [6]. In particular, CSN simulation provides a model that suitably represents, processes associated with specific business units such as ordering system, manufacturing plant, distribution centres, etc.

in the presence of uncertainty [7, 8]. Simulation modelling methods alongside with mathematical and models based on algorithms almost always come together. The main advantage of simulation approaches is a possibility to explore *what-if* scenarios that provide a deeper understanding of the dependencies in a system. The operations of a real system that are usually very dangerous, expensive, or impractical to implement can be evaluated according to their resilience and robustness subject to various predefined inputs (e.g. time horizon, resources, etc.) and at any desired granular level via simulation modelling. Using computer programming, the performance of a real system subject to controlled and environmental changes can be simulated. Therefore, many input values and their combinations can be explored through simulation models [9]. Also, simulation models offer flexibility in developing and assessment of different scenarios, with reasonably high-speed processing. In addition, an embedded standard reporting system make them unique in modelling, analysing, and validating of complex systems.

As pointed out, independent deployment of optimisation and simulation methodologies has some benefits. However, it also has limitations. The main drawback of simulation models is that they can only work with a set configuration of a solution. On the other hand, finding the optimal solution by independently using the traditional optimisation approaches incurs heavy computational cost. Therefore, the integration of the two methods may lead to a uniquely efficient optimisation.

SOA is a key factor of modern design across industries [3]. SOA is often used in the design, modelling and in analyses of systems. It can provide an optimal setting for set of parameters for a simulation model [10]. But due to high computational requirements, scientists have not given much attention to the use of SOA in CSNs [10-12]. Consequently, SOA turns into a hot research topic for optimisation of CSNs. The optimisation core together with a simulation model in SOA, can search the solution space globally (ergodicity of GA) whereas the simulation module acts as a quality assessment unit.

Following the advances in computational power, increased efforts have been made to leverage simulation for optimisation/simulation-based optimisation of hybrid systems with behaviours that can be discrete or continuous [13]. CSNs are hybrid systems with a high level of complexity.

Inventory control planning problems have been tackled using many metaheuristic algorithms [5, 14-18]. GA was widely used to solve related problems [19]. Through exploring the solution space, GA finds optimal or near optimal solutions. But, like in other evolutionary algorithms (EA), GA cannot carry out self-validation. GA risks to converge to local optima [20]. Hence, a valid question is whether or not the obtained solution is a high-quality candidate.

The parameters of the search algorithm - population size, crossover and mutation and rates, as well as the interaction between these parameters have significant impacts on the quality of solutions. As the entire search population or its fitness function might be highly affected by variation of these parameters. This necessitates implementation of a mechanism that can offer parameters tuning is essential. However, it is very hard to perform perfect tuning due to complexity among the interactions of EA's parameters. Most often, trial and error of EA's parameters is used in OR studies. However, experimentally tuning the parameters is less practical and very expensive [21]. We thus, propose using

statistical methods based on experiments as a more robust approach [22].

In [23], the authors present a multi-echelon SN simulation-based optimisation model for a multi-criteria P-D design. The model offers concurrent optimisation of the network's structure, the set of the control strategies, and the quantitative parameters of the strategy for control. The modelling, simulation and then optimisation of networked entities are performed using a graphical interface designed in C++ programming. In this study, the candidate solutions are evaluated by a discrete-event simulation (DES) module. A multi-objective GA algorithm is developed aiming at finding compromised solutions regarding structural, qualitative and quantitative variables. The toolbox developed in the research considers a real Production-Distribution model which makes it a unique decision support system. However, there is no evidence shown with regards to parameters tuning of the GA algorithm.

In [24], the authors describe a two-phase Mixed Integer Linear Programming model addressing planning and scheduling systems of a build-to-order SN system. They use GA to optimise the aggregate costs of both subsystems. Three different scenarios were developed, in which distinct recombination rates for genes was used to improve the quality of solutions.

In [25], the researcher model a P-D network over a tactical planning horizon with uncertain demands and capacity. The proposed algorithm incorporates a simulation and an optimisation module; each calculates the total costs of the network for P-D. The problem is mathematically formulated by a MILP, and the fitness function (total cost) is evaluated via the simulation core. Then the solution resulting from the optimisation module is compared with the obtained output from the simulation module recursively. This procedure iterates until there is a set difference between two solutions. This study reports on data obtained from the implementation of the proposed SOA on a SN problem of a reduced scale. Although the simulation and the optimisation modules are both included in the proposed approach, there is no interaction or connection between them. The application of the simulation module is used to produce initial values for the parameters of the mathematical model. Also, the capacity to generalise the model for similar or larger problems was not addressed. Moreover, no evidence was shown around approaching a solution with better quality if different configurations were chosen for the optimisation parameters.

In [15], the authors developed a modified Particle Swarm Optimisation model (MPSO) for a location-allocation Supply Network problem. They formulated a two-echelon Distribution Network (DN) considering multi-product and multi-period inventory, subject to uncertainty of seasonal demands. The determination of the orders quantity and the vendors' location are pursued as the main objectives in this paper. They use Taguchi to tune the parameters of the MPSO. They considered parameter tuning in their model and they performed a sensitivity analysis for similar problems with different granularity levels.

In a similar study, In [26], the researcher developed a PSO algorithm attempting to find the maximum profit for a channel of a two-echelon SN for a single product. Sales quantity and production rate were used as decision variables of their model. Using a combination of GA, PSO, and simulated annealing (SA), they conduct a detailed sensitivity analysis. However, the

improvement of the proposed heuristic is computed by using another heuristic. This seems very inefficient.

In [27], the authors proposed a simulation optimisation approach to reduce the number of delayed customer orders while costs are kept under control for an integrated production-distribution supply chain. The hybrid modelling combined linear programming and discrete event simulation. This research is a great potential of using SOA approach; however, no effort was made considering the tuning of the control factors of the GA algorithm.

In [28], the researchers developed an agent-based simulation optimisation model through which an online auction policy within the context of the agricultural supply chain was optimised. Three different scenarios namely, oversupply, balance and insufficient supply with different demand and supply quantities were presented to obtain the optimal lot-size and to determine the optimum online auction policy to control inventory. The investigation towards improving the solution quality derived from the proposed methodology was not provided.

An important observation concerning SOA studies is that, in almost all studies, the tuning of the model's variables (e.g. lead time, production rate, etc.) was only attempted in the optimisation module for small problems. Good examples are included in [20] and [22]. On the other hand, evidence in this regard seems to be missing in some studies [23, 29]. Furthermore, very few ([15, 24]) indicated efforts for tuning the optimisation parameters - selection methodology, mutation, and crossover in GA or swarm's cognitive and social components in PSO. They reported that this had been done by trial and error - a typical approach used in the majority of OR studies [21]. The simulation model is run several times, then the better solution is selected. Due to the complexity of the interaction of parameters of the search algorithm as well as the high computational cost, it is unclear how many iterations would be sufficient for a given size problem. Besides, as the scale of the problem increases, the complexity of interactions increases exponentially. Therefore, the difficulties corresponding to this class of SNP problems will further escalate if a more detailed model is simulated. So, it is necessary to study in more depth the variation of the solution quality.

This paper presents an integrated simulation-optimisation approach to solve a class of CSN problem using GA. The objective is to minimise the total cost while an optimum/near optimum inventory level associated to each product family is obtained. An important feature of the under-investigated problem is that both demand and the inventory capacity are uncertain. The randomness of the uncertain parameters is captured by the simulation model. The optimal quantities are searched by GA. Also, a fine-tuning mechanism for the optimisation algorithm's controllable parameters is applied using Taguchi experimental design and ANOVA to improve the quality of the solution. In Section III, the mathematical model, parameters and notations of the proposed problem are summarised.

3. Mathematical Model

This section presents a mathematical model for a multi-product multi-period consumer supply network. The mathematical model presented here consider a planning period of T (indexed by t), a set of product family P (indexed by i) and a set of retailers R (indexed by j) with the limited budget and inventory restrictions.

The parameters in the model are the following:

- D_{ijt} Demand for product family i by retailer j in period t
- D_{ijT} Demand for product family i by retailer j at the end of period T
- I_{0i} Initial inventory level for product family i
- O_{minijt} Minimum quantity of product family i manufactured for retailer j in period t
- O_{maxijt} Maximum quantity of product family i manufactured for retailer j in period t
- V_{max} Maximum capacity of the inventory at DC
- V_t Total capacity of inventory at DC in period t
- a_{ijt} Cost for the ordering of product family i
- b_{ijt} Cost for purchasing one unit of product family i at time t
- c_{ijt} Storage cost for one unit of product family i in period t
- d_{ijt} Handling cost at DC for one unit of product family i in period t
- e_{ijt} Cost for backordering one unit of product family i in period t
- f_{ijt} Cost for transporting one unit of product family i in period t
- $\mathcal{A}_T O$ Total cost of ordering at the end of period T
- $\mathcal{B}_T I$ Total cost of storage in inventory at the end period T
- $\mathcal{C}_T I$ Total cost of handling in inventory at the end of period T
- $\mathcal{D}_T D$ Total cost of purchasing at the end of period T
- $\mathcal{E}_T O$ Total cost of order shortage at the end period T
- $\mathcal{F}_T O$ Total cost of transportation at the end of period T
- \mathcal{C}_T The total network costs at the end of period T
- σ_1 The backorder intensity rate for product family i at the end of period T
- σ_2 The capacity severity rate for product family i at the end of period T

The objective function (1) comprises the minimisation of the total CSN costs, consisting of ordering costs, purchasing costs, transportation costs from manufacturing plants (MP) to retailers (RE), inventory holding and handling costs at the distribution centre (DC), and backordering costs subject to a set of constraints present in (2-4). Constraint (1) represents the quantity of order of a product family i in a period t bounded by the upper and the lower limits. Note, the maximum quantity of an order for product family i from retailer j cannot exceed maximum n folds of the maximum quantity of the demand for the entire planning period T . Constraint (2) is the capacity of the inventory denoted by V_T . The order quantity is a positive integer that is normalised between 0 and 1 by (4) denoted by \hat{O} . Table 1 and Table 2 shows a numerical representation of O_{ijt} , \hat{O}_{ijt} and D_{ijt} for $i = 3, j = 5$ and $t = 2$.

$$\min \sum_{t=1}^T \sum_{j=1}^R \sum_{i=1}^P C_{ijt}(O_{ijt}, I_{ijt}) \tag{1}$$

$$C_T(O_{ijt}, I_{ijt}) = \mathcal{A}_T(O_{ijt}) + \mathcal{B}_T(I_{ijt}) + \mathcal{C}_T(I_{ijt}) + \mathcal{D}_T(D_{ijt}) + \mathcal{E}_T(I_{ijt}) + \mathcal{F}_T(O_{ijt}) ; \forall i, j \geq 0$$

$$\begin{cases} \mathcal{A}_T(O_{ijt}) = a_{ijt} \cdot O_{ijt} \\ \mathcal{B}_T(I_{ijt}) = b_{ijt} \cdot I_{ijt} \\ \mathcal{C}_T(I_{ijt}) = c_{ijt} \cdot I_{ijt} \end{cases} \quad \begin{cases} \mathcal{D}_T(D_{ijt}) = d_{ijt} \cdot D_{ijt} \\ \mathcal{E}_T(I_{ijt}) = e_{ijt} \cdot I_{ijt} \\ \mathcal{F}_T(O_{ijt}) = f_{ijt} \cdot O_{ijt} \end{cases}$$

$$C_T(O_{ijt}, I_{ijt}) = \text{minimise} \sum_{t=1}^T \sum_{j=1}^R \sum_{i=1}^P a_{ijt} \cdot O_{ijt} + b_{ijt} \cdot I_{ijt} + c_{ijt} \cdot I_{ijt} + d_{ijt} \cdot D_{ijt} + e_{ijt} \cdot I_{ijt} + f_{ijt} \cdot O_{ijt}$$

subject to:

$$O_{min} \leq O_{ijt} \leq O_{max} \tag{2}$$

$$O_{min}, O_{max} = [0 \quad n * \max(D_{ijT})] ; \quad n > 1 \tag{3}$$

$$V_{max} \leq V_T$$

$$O_{ijt} = \min([O_{min} + (O_{max} - O_{min} + 1) \times \hat{O}], O_{max}) \tag{4}$$

$$0 \leq \hat{O} \leq 1$$

Table 1. Numerical representation of \hat{O}_{ijt}, O_{ijt}

\hat{O}_{ijt}	\hat{O}_{11t}	\hat{O}_{12t}	\hat{O}_{13t}	...	\hat{O}_{1jt}
	\hat{O}_{21t}	\hat{O}_{22t}	\hat{O}_{23t}	...	\hat{O}_{2jt}
	\vdots	\vdots	\vdots	\vdots	\vdots
	\hat{O}_{i1t}	\hat{O}_{i2t}	\hat{O}_{i3t}	...	\hat{O}_{ijt}
\hat{O}_{ij1}	0.771	0.134	0.681	0.414	0.820
	0.699	0.568	0.332	0.247	0.962
	0.697	0.425	0.106	0.929	0.581
\hat{O}_{ij2}	0.338	0.040	0.182	0.887	0.991
	0.670	0.306	0.771	0.135	0.092
	0.017	0.394	0.973	0.116	0.447
O_{ijt}	O_{11t}	O_{12t}	O_{13t}	...	O_{1jt}
	O_{21t}	O_{22t}	O_{23t}	...	O_{2jt}
	\vdots	\vdots	\vdots	\vdots	\vdots
	O_{i1t}	O_{i2t}	O_{i3t}	...	O_{ijt}
O_{ij1}	161	240	36	111	57
	176	172	52	145	231
	103	96	97	58	56
O_{ij2}	259	248	53	237	288
	212	158	87	124	68
	110	99	196	164	132

Table 2. numerical representation of D_{ijt}

D_{ijt}	D_{11t}	D_{12t}	D_{13t}	...	D_{1jt}
	D_{21t}	D_{22t}	D_{23t}	...	D_{2jt}
	\vdots	\vdots	\vdots	\vdots	\vdots
	D_{i1t}	D_{i2t}	D_{i3t}	...	D_{ijt}
D_{ij1}	259	248	53	237	288
	212	158	87	124	68
	110	99	196	164	132
D_{ij2}	11	26	17	35	38
	72	93	80	42	6
	69	61	87	39	85

Note: D_{11t} presents the quantity of product family 1 to be manufactured for consumer 1 in time interval $t = 1$ is 259 unit.

The I_{ijt} and O_{ijt} are related to the decisions regarding the inventory level and the quantity of orders that are calculated by (5). O_{ijt} is the main decision variable, since I_{ijt} is obtained recursively from O_{ijt} . The demand quantity, D_{ijt} , is unknown but bounded. It can be expressed by probabilistic distribution functions such as normal or uniform distribution functions. In this model, a uniform distribution is used to model D_{ijt} using (6), where D_{min}, D_{max} are the lower and upper bounds, respectively.

Also, each product family has a set volume (v_i) so the total volume of the order i.e. the total volume occupied by the inventory, V_{max} , is calculated by (7)

$$O_{ijt} = I_{ijt} - I_{ijt-1} + D_{ijt} \tag{5}$$

$$D_{ijt} \sim U(D_{min}, D_{max}) \tag{6}$$

$$V_{max} = \sum_{i=1}^P \sum_{j=1}^R \sum_{t=1}^H v_i \times I_{ijt} \tag{7}$$

$$v_i \sim U(0,1)$$

If a solution breaks any constraint (c_i) it is infeasible and therefore the associated evaluation should be penalised in proportion to how violently they break the constraints. In this problem α_1 and α_2 are defined and assigned to the fitness function via (8). The problem size and substantially the changes in the planning period result in changes of α_1, α_2 .

$$C_T(O_{ijt}, I_{ijt}) = \left\{ \left(\sum_{t=1}^T \sum_j^R \sum_{i=1}^P \mathcal{A}_T \cdot O_{ijt} + \mathcal{B}_T \cdot I_{ijt} + \mathcal{C}_T \cdot I_{ijt} + \mathcal{D}_T \cdot D_{ijt} + \mathcal{E}_T \cdot I_{ijt} + \mathcal{F}_T \cdot O_{ijt} \right) + \alpha \sigma_1 \right\} \times (1 + \beta \sigma_2) \tag{8}$$

$$\sigma_1 = \frac{\sum_t^T \sum_j^R \sum_i^P I_{ijt}}{TRP}$$

$$\sigma_2 = \frac{I_{ijk} < 0}{\left(V_{max} - \frac{\sum_t^T V_t}{T} \right)}$$

Also, the average backlogged orders, and the average volume occupied by the inventory are denoted by σ_1 and σ_2 , respectively. In associate with the planning policy in-use, the values of σ_1, σ_2 may vary. For example, if the customer satisfaction rate is %100, which means shortages are not allowed and $\sigma_1 = 0$. Conversely, if a company unable to deliver their promises on time then σ_1 can be set according to the safety stock level. Note, in both cases, the inventory capacity cannot be exceeded, thus $\sigma_2 = 0$. So, a solution candidate is regarded feasible if both conditions are satisfied.

4. Granularity

In systems engineering literature, granularity translates into the level of detail one can decide to consider in a model or decision-making process where the same functionality is expressed with different ‘sized’ designs [30]. In SN, the size of the problem determines the granularity level of the problem which has a significant influence on the computation time and the algorithm’s efficiency. Measures such as the number of product families, the number of facilities, planning periods, etc. are some important factors which affect the granularity level [31]. In this study, in order to verify the robustness of the proposed methodology, three case studies with different granularity levels are considered for the design of experiments represented by a tree structure with two levels L_1 and L_2 (Figure 2). The leaves at L_1 denoted by $[P_S, P_M, P_L]$, correspond to an individual scenario with a distinct problem size, known as *Small*, *Medium* and *Large-scale* problems. L_1 is developed based on the problem size categories proposed by Mousavi, Bahreininejad, Musa and Yusof [15], shown in Table 3. The roots at L_2 are the number of experiments considered for each category. This is determined according to the number of parameters and the levels of variation of a specific parameter which will be developed using Taguchi method (see Section 6).

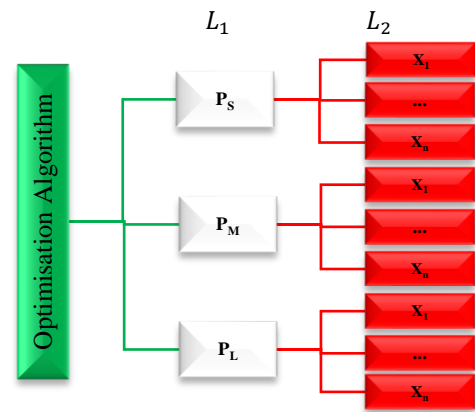


Figure 2 Hierarchical structure proposed for implementation phase

Table 3. Sizes of the proposed instances [15]

Problem Size	Product Family (P)	Manufacturing Plants (MP)	Retailer (RE)	Periods (T)
<i>Small</i>	[1-5]	[1-5]	[5-10]	[1-3]
<i>Medium</i>	[6-10]	[1-10]	[11-20]	[1-5]
<i>Large</i>	[11-15]	[11-15]	[20-30]	[6-10]

Note: a problem with $P = 7, MP = 6, RE = 11$ and $T = 2$ is counted as a Medium-scale problem.

5. Solution Approach

To solve the MPMPCSN problem discussed in this paper, GA optimisation method is used. GA are based on principles of natural selection and genetics to evolve better solutions through multiple consecutive generations. *Selection*, *Crossover* and *Mutation* are implementations in GA of similar phenomena occurring in the natural world. [23]. Based on the quality of solutions, they have a probability to be selected and evolve in new generations and converge towards optimality. Finally, the solutions are tested against termination criteria (evolving procedure). A good search space and genetic operators must maintain an equilibrium between exploration and exploitation and this is key in reaching optimality [32-34]

5.1. Generation and Initialisation

The first step in implementing the GA is to generate a random population of solutions (chromosomes). Chromosomes are resizable according to problem’s attributes and vary based on the problem type, level of complexity, number and type of variables, granularity, etc. Each chromosome consists of several atomic structures - *genes* representing the characteristics of the solution (e.g. number of suppliers, position of manufacturing plants, types of products considered, etc.) [35]. Real coding has been used for this type of problem (Figure 3).

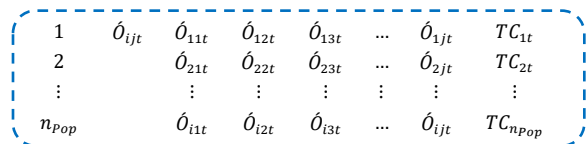


Figure 3 Chromosome representation

The performance of the GA is affected by two opposing factors; population size and computation time. The larger the population size; the longer takes the computation time. The population size should be large enough to incorporate sufficient variation in one generation from which the children in the next

generations are produced. GA is designed to evolve over a number of generations. Hence, having a large population has a serious impact on the computation time. A carefully selected population size that offers sufficient variety but does not permit a fast-enough evolution is needed.

5.2. Genetic Operators: Selection, Crossover, and Mutation

Genetic operators may affect the optimal fitness value for the designed algorithm. The GA operators presented in this paper are selection, crossover and mutation. Roulette Wheel, Tournament and Ranked are the most popular selection mechanisms that are used in this study [33, 36].

In the following step, the offspring population is created by applying single point crossover and mutation. So, new offsprings are produced by combining the characteristics of two parents that can be better than both parents if they take the best characteristics from each of the parents. This mechanism should be performed with a certain probability. Throughout this study, P_c and P_m are referring to crossover and mutation probabilities respectively. Two individuals are produced per randomly selected parents followed by mutating gens of offspring population with specified probability. The mutation is implemented to preserve the variety of the solution pool and prevent GA getting stuck in local optima by exploring the entire search space and maintaining diversity in the population [37]. It is likely that some randomly lost genetic information recovered through mutation. P_m should be set carefully too as such the diversity in the population is preserved but does not negatively affect the overall, fitness of the current population by removing good solutions. Mutation can finely tune the balance between exploration and exploitation. Typically, the mutation rate is small (<2-5%).

5.3. Simulation

After initialising the first population, each chromosome is evaluated for fitness. Fitness function is a metric used to measure the quality of the represented solution. The fitness value of a chromosome is the most important factor in GA evaluation that is always problem dependent [38]. The fitness function defined for MPMPCSN is the minimum cost of running the network. So the lower the fitness value, the higher is the survival chance of a chromosome.

5.4. Stopping Conditions

The optimal/near optimal solution is achieved through an iterative procedure until the stopping condition is satisfied. Choosing the termination criteria depends on the complexity of the problem structure as well as the size of the solution pool [39]. Often, the maximum number of generation is adopted which is the case in this study.

The traditional GA has several shortcomings. As a result of premature convergence, the search parameters (selection, crossover, mutation) may not be very useful towards the end of a search procedure [40]. Also, obtaining an absolute global optimum is not guaranteed, however providing good solutions within a reasonable time is generally expected [41, 42]. Also, GA may not be effective if the starting point in search space was at a great distance from optimal solutions [43]. This deficiency limits the use of GA in real-time applications. However, it can be overcome if GA is hybridised with other local search methods where a closed-form expression of the objective function can be appropriately

performed [42]. Simulation tools are unique methods that are tightly integrated with mathematical and algorithmic based models. Overall, to improve GA performance and obtain accurate solutions, the population size, selection mechanism, crossover and mutation rates and the computational time are required to be turned. Further validation and evaluation of the proposed model and the solution approach is discussed in the following section.

6. Computational Experiments

This section provides experimental results obtained from applying the proposed SOA methodology on practical tests associated to MPMPCSN problems with different granularity levels.

A manufacturing CSN with a central distribution centre is considered in which orders received from consumers are being processed. The demand quantities for P_S , P_M and P_L were randomly generated first and remained unchanged throughout the rest of the optimisation algorithm (see. Appendix A, Table 22-Table 24), because the variation of D_{ijt} causes changes of other parameters. Also, associated purchase cost per unit of product family P_i and the corresponding volume v_i for P_S , P_M and P_L are given in Appendix A (Table 21). All other related costs of running the network consist of ordering cost, backordering cost, holding cost, handling cost and transportation cost are computed via (9)-(14). In addition, the fixed parameters of the model are presented in Table 4.

$$a_{ijt} = 0.1 \times d_{ijt} \tag{9}$$

$$b_{ijt} = 0.05 \times d_{ijt} \tag{10}$$

$$c_{ijt} = 0.05 \times d_{ijt} \tag{11}$$

$$1 \leq d_{ijt} \leq 100 \tag{12}$$

$$e_{ijt} = 0.05 \times d_{ijt} \tag{13}$$

$$f_{ijt} = 0.05 \times d_{ijt} \tag{14}$$

Table 4. fixed parameters of the model

Parameters	P	R	T	V_T
Small-Scale	5	2	2	1000
Medium-Scale	6	11	5	10000
Large-Scale	10	25	8	100000

Note: P, R, and T are referred to the Product family, Retailer and Planning period respectively.

7. Results and Discussion

As discussed above, the performance of the GA optimisation algorithm is mostly influenced by its controllable parameters. These parameters are selection method (P_S), crossover and mutation rate (P_c, P_m), population size (n_{pop}) and the maximum number of iteration ($MaxIt$). Thus, though utilising Taguchi Orthogonal Array Design along with Regression Analysis and Optimisation Solver the optimal parameter set was determined. More details are given in the following sections.

7.1. Process of Experiment Design

The main two components of the Taguchi method are the number of parameters and their variation levels. In order to analyse the results obtained from ANOVA (analysis of variance) and S/N ratio (signal to noise), it is required to create a set of tables of

numbers known as *orthogonal arrays*. These tables are then used first to reduce the number of experiments, next to determine the most critical parameters with high impact on the outcomes. In this study, we consider the GA controllable parameters as significant factors in 3 levels (Table 7). The Taguchi Orthogonal Array Design ($L27 - 3^5$) shown in Table 6 is proposed and created by Minitab.

Table 5. The GA parameters' level

Granularity Level Parameters	Small-scale	Medium-Scale	Large-Scale
	Level 1	Level 2	Level 3
P_s	RW	T	R
P_c	0.9	0.85	0.8
P_m	0.1	0.05	0.025
n_{Pop}	[30 60 120]	[100 150 200]	[100 200 300]
MaxIt	[200 100 50]	[500 400 300]	[3500 3000 2000]

RW, T and R referred to Roulette Wheel, Tournament and Ranked Selection method respectively

Table 6. The layout of the orthogonal array for 5 factors in 3 levels

No.	P_s	P_c	P_m	n_{Pop}	MaxIt
S1	1	1	1	1	1
S2	1	1	1	1	2
S3	1	1	1	1	3
S4	1	2	2	2	1
S5	1	2	2	2	2
S6	1	2	2	2	3
S7	1	3	3	3	1
S8	1	3	3	3	2
S9	1	3	3	3	3
S10	2	1	2	3	1
S11	2	1	2	3	2
S12	2	1	2	3	3
S13	2	2	3	1	1
S14	2	2	3	1	2
S15	2	2	3	1	3
S16	2	3	1	2	1
S17	2	3	1	2	2
S18	2	3	1	2	3
S19	3	1	3	2	1
S20	3	1	3	2	2
S21	3	1	3	2	3
S22	3	2	1	3	1
S23	3	2	1	3	2
S24	3	2	1	3	3
S25	3	3	2	1	1
S26	3	3	2	1	2
S27	3	3	2	1	3

7.2. Signal-to-Noise (S/N) Ratio Method

S/N ratios evaluate the size of the apparent effect (signal) against the size of random fluctuations (noise) witnessed in the data. The higher this indicator, the better the compromise is which can be calculated in different ways according to the optimisation problem (minimisation/maximisation) [44]. In this study, S/N ratio values are calculated to determine the best combination of GA control factors. The proposed optimisation algorithm was run four times for each parameter set to obtain more refined solutions. The numerical results for the Small, Medium and Large-scale problem are reported in Table 7, Table 8 and Table 9, respectively.

This problem is aimed to minimise the response value (y). Therefore, to minimise the mean-square deviation (MSD) from the target value 0 and maximise the S/N ratio, MSD has to be

calculated using (15). The signal to noise (S/N) ratio, in this case, is defined by (16), where n is the sample size.

Table 7. Taguchi experimental design and design data of GA for small-scale problem

Trial	Function Evaluation (TC)				μ	σ	P_s	P_c	P_m	n_{Pop}	MaxIt
	Run 1	Run 2	Run 3	Run 4							
1	52545.31	52838.97	52824.62	52798.99	52751.97	138.76	RW	0.9	0.1	30	200
2	57736.13	57984.64	54800.67	56440.22	56740.41	1459.71	RW	0.9	0.1	30	100
3	55082.79	54767.13	55334.41	55983.30	55291.91	516.06	RW	0.9	0.1	30	50
4	57348.28	56895.83	58086.99	58118.95	57612.51	595.84	RW	0.85	0.05	60	200
5	59594.91	58612.27	61314.77	61253.54	60193.87	1321.56	RW	0.85	0.05	60	100
6	60380.16	62646.26	60710.87	59366.54	60775.96	1371.79	RW	0.85	0.05	60	50
7	55536.78	54608.65	55060.74	54506.04	54928.05	471.97	RW	0.8	0.025	120	200
8	55135.85	54540.19	54946.94	56517.07	55285.01	858.15	RW	0.8	0.025	120	100
9	57518.01	59179.99	56537.55	57925.29	57790.21	1094.37	RW	0.8	0.025	120	50
10	52410.97	52718.79	52428.90	52416.20	52493.72	150.24	T	0.9	0.05	120	200
11	53368.53	52881.84	53767.00	52857.57	53218.73	434.73	T	0.9	0.05	120	100
12	58698.26	55432.41	56344.90	57940.46	57104.01	1484.56	T	0.9	0.05	120	50
13	54263.36	56283.66	55064.54	55837.51	55362.27	889.02	T	0.85	0.025	30	200
14	56139.17	56388.68	57656.13	56204.80	56597.20	713.81	T	0.85	0.025	30	100
15	62448.49	94741.69	60631.15	98432.34	79063.42	20304.09	T	0.85	0.025	30	50
16	52413.82	52417.87	52439.46	52418.27	52422.36	11.58	T	0.8	0.1	60	200
17	53546.80	54432.52	53665.56	52804.39	53612.32	666.48	T	0.8	0.1	60	100
18	62686.18	56408.68	56602.45	56552.31	58062.41	3083.61	T	0.8	0.1	60	50
19	54034.56	53650.51	53214.02	53760.76	53664.96	341.24	R	0.9	0.025	60	200
20	56947.05	58519.69	57332.37	56946.45	57436.39	744.73	R	0.9	0.025	60	100
21	62368.65	58889.81	64213.45	64114.90	62396.70	2486.75	R	0.9	0.025	60	50
22	52472.93	52454.69	52466.57	52462.89	52464.27	7.61	R	0.85	0.1	120	200
23	54151.02	54381.73	54913.21	54443.82	54472.45	319.71	R	0.85	0.1	120	100
24	59054.53	58677.67	59390.45	59848.09	59242.69	497.66	R	0.85	0.1	120	50
25	54123.74	53139.69	53600.31	53588.71	53613.11	402.34	R	0.8	0.05	30	200
26	62582.39	57133.15	57636.73	58226.32	58894.65	2498.75	R	0.8	0.05	30	100
27	76782.74	63219.40	67855.77	65419.86	68319.44	5951.48	R	0.8	0.05	30	50

Note: (μ : mean, σ : Standard deviation)

Table 8. Taguchi experimental design and design data of GA for medium-scale problem

Trial	Function Evaluation (TC)				μ	σ	P_s	P_c	P_m	n_{Pop}	MaxIt
	Run 1	Run 2	Run 3	Run 4							
1	3055303	3053526	3046047	3050184	3051265.00	4074.87	RW	0.9	0.1	200	500
2	3149794	3154852	3180213	3164676	3162383.75	13396.09	RW	0.9	0.1	200	400
3	3372901	3350114	3335613	3323874	3345625.50	21114.59	RW	0.9	0.1	200	300
4	3200575	3185842	3197118	3191536	3193767.75	6464.29	RW	0.85	0.05	150	500
5	3355893	3308538	3369709	3382514	3354163.50	32301.14	RW	0.85	0.05	150	400
6	3499418	3511169	3529597	3529401	3517396.25	14775.75	RW	0.85	0.05	150	300
7	3432256	3440475	3410509	3433997	3429309.25	13022.82	RW	0.8	0.025	100	500
8	3575145	3520148	3586398	3537586	3554819.25	31141.79	RW	0.8	0.025	100	400
9	4555883	4146796	3846552	4203898	4188282.25	290903.67	RW	0.8	0.025	100	300
10	3051447	3066724	3034552	3045986	3049677.25	13368.15	T	0.9	0.05	100	500
11	3156857	3217344	3129544	3179152	3170724.25	37114.87	T	0.9	0.05	100	400
12	3281164	3310920	3406627	3340245	3334739.00	53652.67	T	0.9	0.05	100	300
13	3077422	3072374	3047223	3078703	3068930.50	14727.31	T	0.85	0.025	200	500
14	3182456	3188477	3166677	3221685	3189823.75	23144.54	T	0.85	0.025	200	400
15	3436084	3441521	3417875	3435688	3432792.00	10294.60	T	0.85	0.025	200	300
16	2991777	2972519	2986549	2982617	2983365.50	8146.47	T	0.8	0.1	150	500
17	3057430	3030744	3064818	3033992	3046746.00	16926.05	T	0.8	0.1	150	400
18	3172227	3184862	3188181	3173263	3179633.25	8079.53	T	0.8	0.1	150	300
19	3360788	3373308	3403272	3440016	3394346.00	35280.59	R	0.9	0.025	150	500
20	3503662	3492818	3501245	3457735	3488865.00	21267.49	R	0.9	0.025	150	400
21	6083231	6707308	6357912	5970323	6279693.50	328268.14	R	0.9	0.025	150	300
22	3099402	3117656	3130297	3111689	3114761.00	12846.33	R	0.85	0.1	100	500
23	3243754	3249067	3272814	3255208	3255210.75	12634.31	R	0.85	0.1	100	400
24	3410574	3462829	3421737	3409948	3426272.00	24965.85	R	0.85	0.1	100	300
25	3232477	3281042	3288839	3245727	3262021.25	27198.93	R	0.8	0.05	200	500
26	3372780	3354390	3375793	3360978	3365985.25	10031.33	R	0.8	0.05	200	400
27	3542951	3526380	3567064	3540808	3544300.75	16865.55	R	0.8	0.05	200	300

Note: (μ : mean, σ : Standard deviation)

Table 9. Taguchi experimental design and design data of GA for large-scale problem

Trial	Function Evaluation (TC)				μ	σ	P_s	P_c	P_m	n_{pop}	MaxIt
	Run 1	Run 2	Run 3	Run 4							
1	6197853	6182641	6205040	6171968	6189375	14895.51	RW	0.9	0.1	100	3500
2	6171968	6197853	6182641	6205040	6189375	14895.51	RW	0.9	0.1	100	3000
3	6026883	6036349	6064389	6092117	6054934	29462.9	RW	0.9	0.1	100	2000
4	6171329	6156044	6162801	6148266	6159610	9813.874	RW	0.85	0.05	200	3500
5	6189910	6160183	6168566	6183770	6175607	13646.61	RW	0.85	0.05	200	3000
6	6276588	6197853	6256788	6232034	6240816	33949.63	RW	0.85	0.05	200	2000
7	5609430	5583614	5604952	5587145	5596285	12806.33	RW	0.8	0.25	300	3500
8	6219773	6220941	6220798	6296291	6239450	37896.94	RW	0.8	0.025	300	3000
9	6393839	6421235	6397965	6500502	6428385	49567.4	RW	0.8	0.025	300	2000
10	5765313	5783485	5797242	5786545	5783146	13270.95	T	0.9	0.05	300	3500
11	6145198	6115210	6141100	6146131	6136910	14630.8	T	0.9	0.05	300	3000
12	6181667	6166755	6174604	6150186	6168303	13526.68	T	0.9	0.05	300	2000
13	5766122	5797580	5768570	5819409	5787920	25393.04	T	0.85	0.025	100	3500
14	6330538	6295429	6350012	6405480	6345365	46003.23	T	0.85	0.025	100	3000
15	6421425	6446814	6429805	6425072	6430779	11227.04	T	0.85	0.025	100	2000
16	6124129	6234488	6150018	6149727	6164591	48152.91	T	0.8	0.1	200	3500
17	6132648	6141044	6166400	6151393	6147871	14537.94	T	0.8	0.1	200	3000
18	5803931	5783648	5803967	5805327	5799218	10400.52	T	0.8	0.1	200	2000
19	5930494	5953702	5898563	5878441	5915300	33387.83	R	0.9	0.025	200	3500
20	6231820	6227294	6232032	6276543	6241922	23183.89	R	0.9	0.025	200	3000
21	6401559	6416416	6425043	6414352	6414342	9699.475	R	0.9	0.025	200	2000
22	6103190	6123392	6079358	6102566	6102126	17999.54	R	0.85	0.1	300	3500
23	5729010	5873701	5867909	5715137	5796439	86089.04	R	0.85	0.1	300	3000
24	6103190	6123392	6079358	6122019	6106989	20598.34	R	0.85	0.1	300	2000
25	6271088	6235294	6240251	6212358	6239748	24169.66	R	0.8	0.05	100	3500
26	6219361	6297088	6249893	6228781	6248781	34642.66	R	0.8	0.05	100	3000
27	6402903	6433575	6407674	6422437	6416647	14017.7	R	0.8	0.05	100	2000

Note: (μ : mean, σ : Standard deviation)

$$MSD = \frac{1}{n} \sum_{i=1}^n y_i^2 \tag{15}$$

$$\frac{S}{N} = -10 \log(MSD) \tag{16}$$

The example of the calculation of S/N ratio for the control parameter P_s is shown below (column 1 of Table 10) and the results correspond to each case study are summarised in Table 10, Table 11 and Table 12. The difference between the levels of factors in the Table 10- Table 12 determines which parameter has more effect on the quality characteristics (the total cost of the network).

$$Level 1 = \frac{(-94.44 - 95.08 - 94.85 - 95.21 - 95.59 - 95.68 - 94.80 - 94.85 - 94.24)}{9} = -95.08$$

$$Level 2 = \frac{(-94.40 - 94.52 - 95.13 - 94.86 - 95.05 - 98.16 - 94.39 - 94.58 - 95.28)}{9} = -95.16$$

$$Level 3 = \frac{(-94.60 - 95.18 - 95.91 - 94.40 - 94.the 72 - 95.45 - 94.59 - 95.41 - 96.72)}{9} = -95.22$$

$$Difference = |highest\ value| - |lowest\ value| \\ = |-95.22| - |-95.08| = 0.14$$

As it can be seen from Table 10, the control factor $MaxIt$, by far is the most important factor that impacts on S/N ratio (1.19), n_{pop} , P_m , P_c and P_s are also significant factors. Table 11 shows $MaxIt$, P_s and P_m are approximately double of P_c and n_{pop} . Also, in Table 12 while control factor P_c has a negligible effect in influencing the S/N ratio in P_L problem, the contribution of all other four parameters (P_s , P_m , n_{pop} and $MaxIt$) to the S/N is more than 10%.

The S/N ratios computed for the data set P_s , P_c , P_m , n_{pop} and $MaxIt$ (Table 10-Table 12) are essential for sketching the S/N

ratio response diagrams for P_s , P_m and P_L problems (0). So, a higher S/N ratio is related to a data set with the minimum variation which is considered as the best data set.

Table 10. The response table of S/N ratio of P_s Problem

	Selection (P_s)	Crossover Rate (P_c)	Mutation Rate (P_m)	Population Size (n_{pop})	Generation ($MaxIt$)
Level 1	-95.08	-94.90	-94.80	-95.46	-94.63
Level 2	-95.16	-95.46	-95.25	-95.16	-95.00
Level 3	-95.22	-95.10	-95.41	-94.84	-95.83
Difference	0.14	0.56	0.61	0.63	1.19

Table 11. The response table of S/N ratio P_m Problem

	Selection (P_s)	Crossover Rate (P_c)	Mutation Rate (P_m)	Population Size (n_{pop})	Generation ($MaxIt$)
Level 1	-130.7	-130.9	-130	-130.3	-130
Level 2	-130	-130.3	-130.4	-130.9	-130.3
Level 3	-131.1	-130.6	-131.3	-130.6	-131.4
Difference	1.1	0.5	1.3	0.6	1.4

Table 12. The response table of S/N ratio P_L Problem

	Selection (P_s)	Crossover Rate (P_c)	Mutation Rate (P_m)	Population Size (n_{pop})	Generation ($MaxIt$)
Level 1	-135.8	-135.6	-135.6	-135.9	-135.5
Level 2	-135.7	-135.7	-135.8	-135.8	-135.8
Level 3	-135.8	-135.8	-135.7	-135.6	-135.9
Difference	0.1	0.2	0.2	0.3	0.4

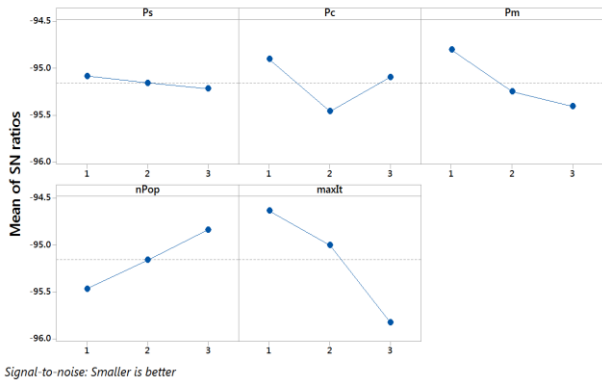
Therefore, the best values associated with P_s , P_c , P_m , n_{pop} and $MaxIt$ corresponding to P_s , P_m and P_L problems are as follows: for P_s , level 1 (Roulette Wheel selection), level 1 (90% crossover), level 1 (10% mutation), level 2 (120 chromosomes) and level 1 (200 iterations), respectively; for P_m level 2 (Tournament selection), level 2 (85% crossover), level 1 (10% mutation), level 1 (200 chromosomes) and level 1 (500 iterations), respectively; For P_L level 2 (Tournament selection), level 1 (90% crossover), level 1 (10% mutation), level 3 (300 chromosomes) and level 1 (3500 iterations), respectively. This can be observed from S/N ratio response diagrams too (Figure 4). The rows show difference values in Table 10-Table 12 determine the contribution level of each parameter in obtaining lower cost. So, the total cost of running the network, for example for P_m problem, is mostly affected by the number of generation, mutation rate, the selection method, population size and crossover rates of the GA algorithm. To determine the significant level of these parameters, ANOVA method is utilised for which the data given in Table 7- Table 9 are going to be used again. Results obtained from ANOVA are summarised in Table 13-Table 15.

7.3. ANOVA Method

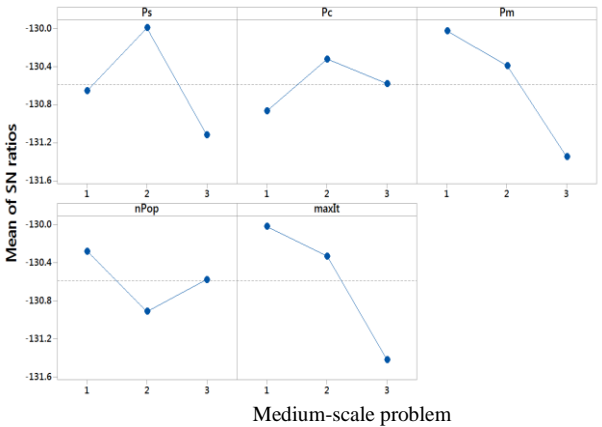
From ANOVA, the percentage contribution ratio (PCR) of each parameter can be calculated. PCR indicates the significance of all main factors and their interactions on the output. The calculation is performed by comparing the mean square (MS) against an estimate of the experimental errors at specific confidence levels. The total sum of squared deviations (SS_T) from the total mean S/N ratio is calculated via (17).

$$SS_T = \sum_{i=1}^n (\eta_i - n_m)^2 \tag{17}$$

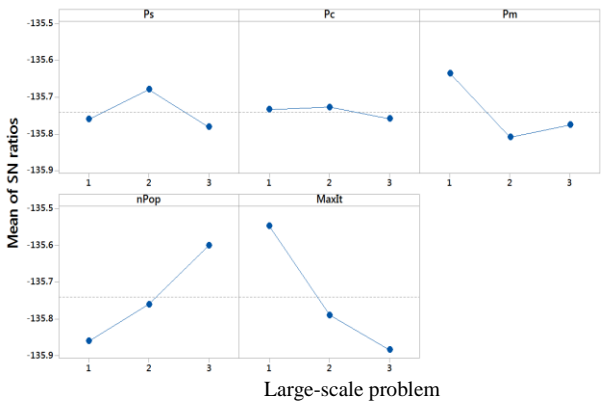
where n is the number of experiments in the orthogonal array and η_i is the mean S/N ratio for the i^{th} experiment.



Small-scale problem



Medium-scale problem



Large-scale problem

Figure 4 The main effect diagram for S/N Ratio response diagram for GA parameters ($P_s, P_c, P_m, n_{Pop}, MaxIt$)

The ANOVA tables for S/N ratios corresponding to the data in Table 10-Table 12 are summarised in Table 13- Table 15. The terms SS_T and MS_T are corresponding to the total sum of squared and the total mean square, respectively. Also, the F-ratios and P-values provided in “F” and “P” columns are calculated via (18) and (19), respectively. F-ratio indicates which parameter ($P_s, P_c, P_m, maxIt$) have a significant effect on the quality characteristic (TC) and P-value determines the significant percentage of the parameters on the quality characteristic (TC).

$$F = \frac{SS_T}{MS_T} \quad (18)$$

$$P = \frac{SS_T}{SS_T} \quad (19)$$

Table 13. Results obtained from ANOVA for Small-scale problem

Source	DF	SS_T	MS_T	F	P
P_s	2	19727169	9863585	0.38	0.685
P_c	2	2.76E+08	1.38E+08	5.32	0.006
P_m	2	3.33E+08	1.66E+08	6.41	0.002
n_{Pop}	2	3.49E+08	1.75E+08	6.72	0.002
$MaxIt$	2	1.24E+09	6.22E+08	23.96	0
Error	97	2.52E+09	25971697		

Table 14. Results obtained from ANOVA for medium-scale problem

Source	DF	SS_T	MS_T	F	P
P_s	2	4.86E+12	2.43E+12	14.27	0
P_c	2	1.69E+12	8.44E+11	4.96	0.009
P_m	2	7.30E+12	3.65E+12	21.45	0
n_{Pop}	2	2.07E+12	1.03E+12	6.08	0.003
$MaxIt$	2	8.19E+12	4.10E+12	24.08	0
Error	97	1.65E+13	1.70E+11		

Table 15. Results obtained from ANOVA for large-scale problem

Source	DF	SS_T	MS_T	F	P
P_s	2	1.02E+11	5.1E+10	1.52	0.223
P_c	2	1.2E+10	5.98E+09	0.18	0.837
P_m	2	3.09E+11	1.55E+11	4.61	0.012
n_{Pop}	2	5.93E+11	2.96E+11	8.85	0
$MaxIt$	2	1.06E+12	5.30E+11	15.82	0
Error	97	3.25E+12	3.35E+10		

Note: SS and V stand for the sum of squared and the variance respectively.

It can be observed from Table 13 that the difference between the mean values of the level of the control factor P_s (selection method) is insignificant ($0.68 > \alpha = 0.05$). Therefore, any selection strategy can be chosen for implementation of the proposed SOA for small-scale problem. However, the difference between the mean values of crossover rates (P_c), mutation rate (P_m) and the number of iteration ($MaxIt$) is significant ($0.006, 0.002$ and $0.002 < \alpha = 0.05$). Thus, the best control factor setting for maximising the S/N ratio is P_c at level 1, P_m at level 1, n_{Pop} at level 2 and $MaxIt$ at level 1. In the Medium-scale problem, all of the control factors are highly contributing to the performance of the SOA (Table 14). According to Table 15, only P_m, n_{Pop} and $MaxIt$ are significantly influenced on the performance of the SOA in Large-scale problem, while there is no restriction in choosing the selection strategy and the crossover rate.

7.4. Confirmation test

The final step of the verification phase is to perform the confirmation test with the optimal level of the GA parameters drawn based on the Taguchi’s design approach for each case study (Table 16).

Table 16. The best combination of the GA parameters

	P_s	P_c	P_m	$nPop$	$MaxIt$
Small-Scale	R	0.9	0.1	120	200
Medium-Scale	T	0.85	0.1	200	500
Large-Scale	R	0.9	0.05	100	3500

The results obtained from the proposed methodology and GA solver associated with P_s, P_m and P_L problems along with the average of the best and the worst results are summarised in Table 17. The quality measurement of the solution is determined according to the value of standard deviation (σ). Therefore, the solution candidate with the maximum σ is considered as the worst solution and the one with the minimum value is regarded as the

best solution. Hence, the experiments No. 15 and No. 22 are the worst and the best scenario for the Small-scale problem, respectively.

Table 17. The total optimised cost

Problem Size	Small-scale	Medium-scale	Large-scale
Optimal Scenario	49966.28(\$)	2921429.2(\$)	5971604 (\$)
Best Scenario	52464.27(\$)	3051265(\$)	6102126 (\$)
Worst Scenario	79063.41(\$)	6279694(\$)	6239450 (\$)

As can be seen from Figure 5, the proposed algorithm shows better performance compared with the best and the worst solutions acquired from GA solver ($5\% \cong \$ 2498$). A similar improvement was also experienced in Medium-scale and the Large-scale problem with $4\% \cong \$ 129835.5$ and $2\% \cong \$ 130522$, respectively.

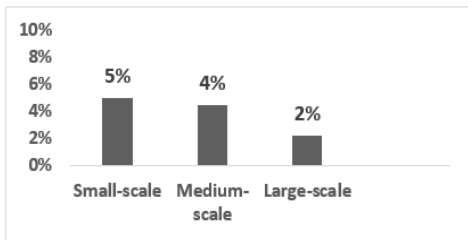
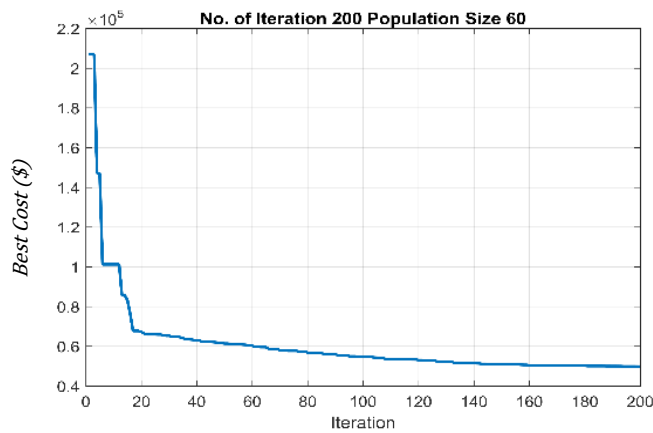
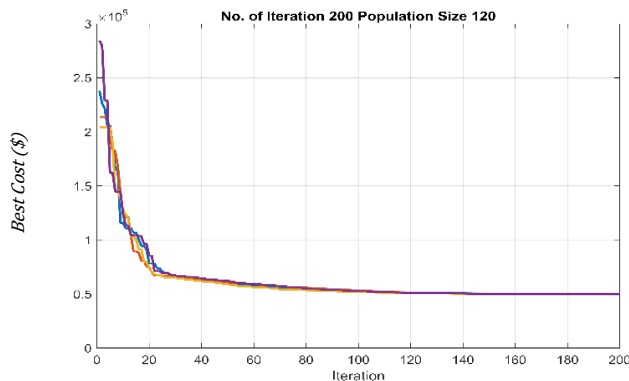


Figure 5 Improvement rates obtained from the tuning procedure

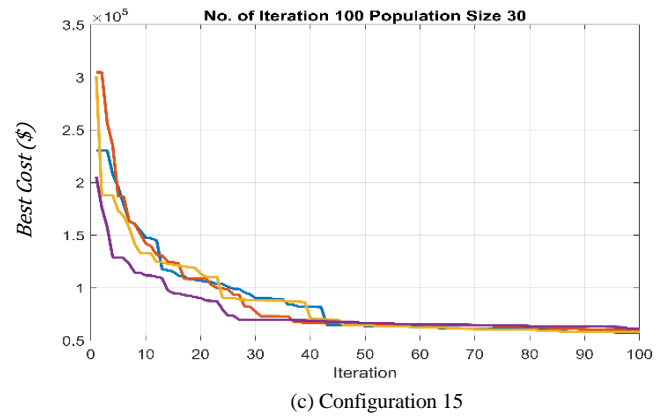
Also, the results obtained from the proposed SOA algorithm, and the GA solver associated with P_S , P_M , and P_L case studies are depicted in Figure 6-Figure 8.



(a) Proposed SOA

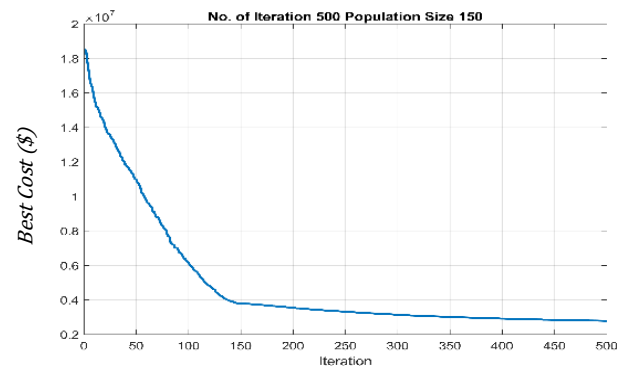


(b) Configuration 22

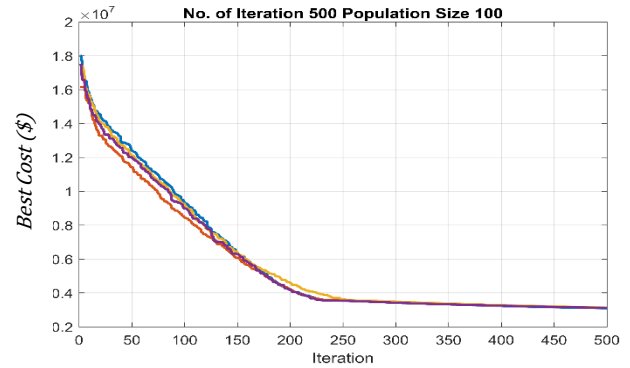


(c) Configuration 15

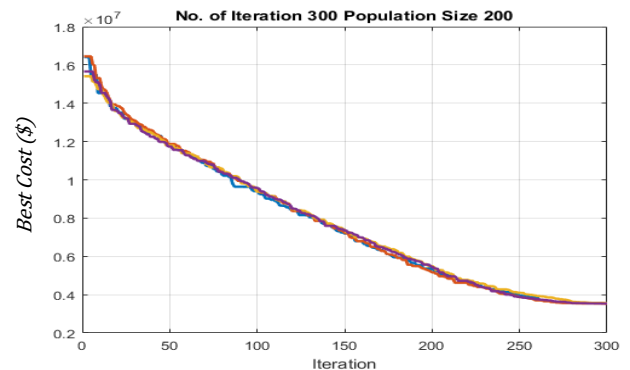
Figure 6 Results obtained from (a) the proposed SOA methodology, (b) the GA optimiser (S22) and (c) the GA optimiser (S15) for P_S



(a) Proposed SOA

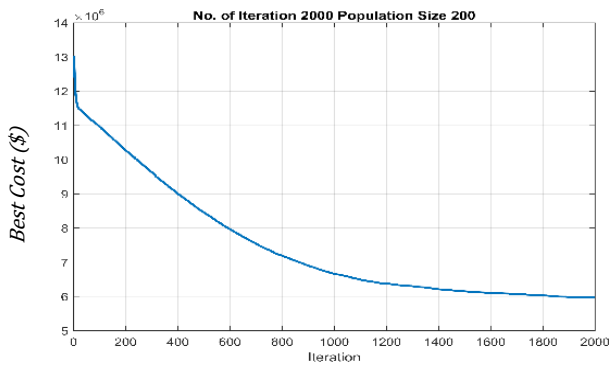


(b) Configuration 21

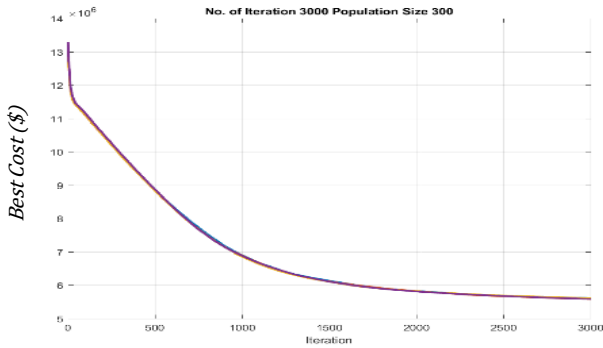


(c) Configuration 1

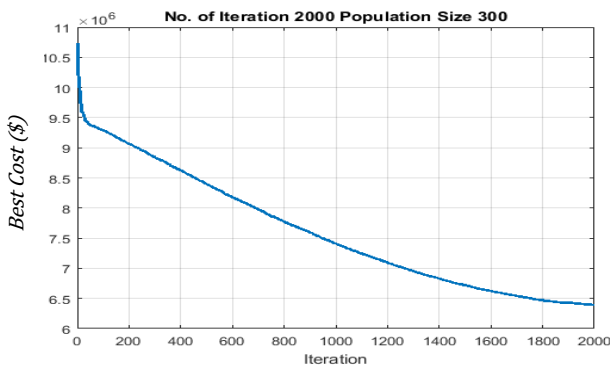
Figure 7 Total cost achieved from implementing (a) the proposed SOA methodology, (b) the GA optimiser (S21) and (c) the GA optimiser (S1) for P_M



(a) Proposed SOA



(b) Configuration 23



(c) Configuration 9

Figure 8 Total cost achieved from implementing (a) the proposed SOA, (b) the GA optimiser (S23) and (c) the GA optimiser (S9) for P_L

Table 18-Table 20 present the optimum quantities associated with each product family to be manufactured for consumers over the given planning horizon.

Table 18. The Optimum Solution for Small-scale problem

	P_1	P_2	P_3	P_4	P_5
T1	11	1	54	4	1
T2	10	11	1	5	80
Total	21	12	55	9	81

Table 19. The Optimum Solution for Medium-scale problem

	P_1	P_2	P_3	P_4	P_5	P_6
T1	136	314	362	220	450	276
T2	391	396	292	575	403	197
T3	369	658	557	574	464	349
T4	499	656	831	433	404	509
T5	577	622	727	681	1013	1086
Total	1972	2646	2769	2483	2734	2417

Table 20. The Optimum Solution for large-scale problem

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
T1	802	706	543	477	471	488	1026	768	670	590
T2	579	480	740	915	561	771	994	820	775	822
T3	710	811	917	608	877	703	952	791	946	1077
T4	1354	1128	630	1161	1058	1222	1090	1099	1427	1187
T5	1507	1771	1624	1429	1524	1229	1145	1537	1254	1554
T6	1685	1935	1762	1952	2055	1802	1848	1903	1397	1698
T7	1997	2192	2097	2037	2118	2435	1883	1918	2276	2854
T8	1904	2411	2159	2765	2271	2542	2309	2604	2437	1998
Total	10252	10371	10455	10606	10575	9608	9815	10685	10187	9990

8. Conclusion and outlook to future

In this paper, an advanced decision-making system for a class of CSN problems was proposed. A novel SOA algorithm incorporating GA as its optimisation module was designed for MPMPCSN problem. The robustness and effectiveness of the proposed methodology was verified through performing twenty-seven computational trials on three practical test problems at different granularity levels (small-scale, medium-scale, large-scale). In addition, a tuning mechanism was recommended to improve the quality of the obtained solutions that was affected by controllable parameters of the optimisation module. To this end, two statistical techniques known as ANOVA and Taguchi methods were utilised. The optimum levels associated to the controllable parameters of GA were determined as following: for P_S , level 1 (Roulette Wheel selection), level 1 (90% crossover), level 1 (10% mutation), level 2 (120 chromosomes) and level 1 (200 iterations), respectively; for P_M level 2 (Tournament selection), level 2 (85% crossover), level 1 (10% mutation), level 1 (200 chromosomes) and level 1 (500 iterations), respectively; For P_L level 2 (Tournament selection), level 1 (90% crossover), level 1 (10% mutation), level 3 (300 chromosomes) and level 1 (3500 iterations), respectively. The proposed SOA was resulted in 5%, 4% and 2% improvement in total cost of CSN associated to P_S , P_M and P_L problems respectively, in contrast to only using GA solver. Also, it was observed that the computational cost and time were reduced significantly.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors are grateful to the Australian Mathematical Society (*Aust MS*) for providing the Lift-up fellowship which financially supported this work.

References

- [1] Z. Hajiabolhasani, R. Marian, and J. Boland, "Consumer Supply Network Planning: Literature Review And Analysis," *Journal of Multidisciplinary Engineering Science Studies*, vol. 3, no. 3, pp. 1519-1538, 2017.
- [2] Z. H. Abolhasani, R. Marian, and J. Boland, "Simulation-Optimisation of Multi-Product, Multi-Period Consumer Supply Network using Genetic Algorithms," in *Intelligent Systems Conference (INTELLISYS)*, London, UK, 2017, pp. 34-44: IEEE, 2017.
- [3] X.-S. Yang, S. Koziel, and L. Leifsson, "Computational Optimization, Modelling and Simulation: Past, Present and Future," in *ICCS 2014. 14th International Conference on Computational Science*, 2014, vol. 29, pp. 754-758: Procedia Computer Science.
- [4] T. Hennies, T. Reggelin, J. Tolujew, and P.-A. Piccut, "Mesoscopic supply chain simulation," *Journal of Computational Science*, vol. 5, no. 3, pp. 463-470, 2014.
- [5] R. A. Alive, B. Fazlohhahi, B. G. Guirimov, and R. R. Aliev, "Fuzzy-genetic approach to aggregate production-distribution planning in supply chain management," *Information Sciences*, vol. 177, pp. 4241-4255, 2007.

- [6] N. Mustafee, K. Katsaliaki, and S. J. E. Taylor, "A review of literature in distributed supply chain simulation," presented at the Simulation Conference (WSC), 2014 Winter, 7-10 Dec. 2014, 2014.
- [7] T. M. Pinho, J. P. Coelho, A. P. Moreira, and J. Boaventura-Cunha, "Modelling a biomass supply chain through discrete-event simulation**This work was supported by the FCT - Fundação para a Ciência e Tecnologia through the PhD Studentship SFRH/BD/98032/2013, program POPH - Programa Operacional Potencial Humano and FSE - Fundo Social Europeu," *IFAC-PapersOnLine*, vol. 49, no. 2, pp. 84-89, 2016/01/01/ 2016.
- [8] F. Campuzano and J. Mula, *Supply chain simulation (A system dynamics approach for improving performance)*. Springer, 2011.
- [9] G. Dellino, J. P. C. Kleijnen, and C. Meloni, "Robust optimization in simulation: Taguchi and Response Surface Methodology," *Int. J. Production Economics*, vol. 125, pp. 52-59, 2010.
- [10] A. Huerta-Barrientos, M. Elizondo-Cortés, and I. F. d. I. Mota, "Analysis of scientific collaboration patterns in the co-authorship network of Simulation—Optimization of supply chains," *Simulation Modelling Practice and Theory*, vol. 46, pp. 135-148, 2014.
- [11] X. Wan, J. F. Pekny, and G. V. Reklaitis, "Simulation-based optimization with surrogate models—Application to supply chain management," *Computers and Chemical Engineering*, vol. 29, pp. 1317–1328, 2005.
- [12] J. Y. Jung, G. Blaua, J. F. Pekny, G. V. Reklaitis, and D. Eversdykb, "A simulation based optimization approach to supply chain management under demand uncertainty," *Computers and Chemical Engineering*, vol. 28, pp. 2087–2106, 2004.
- [13] M. C. Fu, "Optimization via simulation: A review," *Annals of Operations Research* vol. 53, pp. 199-247, 1994.
- [14] J.-H. Kang and Y.-D. Kim, "Inventory control in a two-level supply chain with risk pooling effect," *International Journal of Production Economics*, vol. 135, no. 1, pp. 116-124, 2012.
- [15] S. M. Mousavi, A. Bahreinejad, S. N. Musa, and F. Yusof, "A modified particle swarm optimization for solving the integrated location and inventory control problems in a two-echelon supply chain network," *Intell Manuf*, 2014.
- [16] F. T. S. Chan and A. Prakash, "Inventory management in a lateral collaborative manufacturing supply chain: a simulation study," *International Journal of Production Research*, vol. 50, no. 16, p. 15, 15 August 2012 2012.
- [17] O. Labarthe, B. Espinasse, A. Ferrarini, and B. Montreuil, "Toward a methodological framework for agent-based modeling and simulation of supply chains in a mass customization context," *Simulation Modelling Practice and Theory*, vol. 15, no. 2, pp. 113-136, 2007.
- [18] F. Longo and G. Mirabelli, "An advanced supply chain management tool based on modeling and simulation," *Computers & Industrial Engineering*, vol. 54, no. 3, pp. 570-588, 2008.
- [19] A. Alrabghi and A. Tiwari, "State of the art in simulation-based optimisation for maintenance systems," *Computers & Industrial Engineering*, vol. 82, pp. 167-182, 4// 2015.
- [20] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 309-313, 2015.
- [21] A. E. Eiben, R. Hinterding, and Z. Michalewicz, "Parameter control in evolutionary algorithms," *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*, vol. 3, no. 2, pp. 124-141, 1999.
- [22] J. Sadeghi, S. M. Mousavi, S. T. A. Niaki, and S. Sadeghi, "Optimizing a multi-vendor multi-retailer vendor managed inventory problem: Two tuned meta-heuristic algorithms," *Knowledge-Based Systems*, vol. 50, pp. 159-170, 2013.
- [23] H. Ding, L. Benyoucef, and X. Xie, "Stochastic multi-objective production-distribution network design using simulation-based optimization," *International Journal of Production Research*, vol. 47, no. 2, pp. 479-505, 2009.
- [24] A. D. Yimer and K. Demirli, "A genetic approach to two-phase optimization of dynamic supply chain scheduling," *Computers & Industrial Engineering*, vol. 58, no. 3, pp. 411-422, 4// 2010.
- [25] A. Nikolopoulou and M. G. Ierapetritou, "Hybrid simulation based optimization approach for supply chain management," *Computers & Chemical Engineering*, vol. 47, pp. 183-193, 12/20/ 2012.
- [26] M. Seifbarghy, M. M. Kalani, and M. Hemmati, "A discrete particle swarm optimization algorithm with local search for a production-based two-echelon single-vendor multiple-buyer supply chain," *Journal of Industrial Engineering International*, journal article vol. 12, no. 1, pp. 29-43, 2016.
- [27] E. M. Frazzon, A. Albrecht, M. Pires, E. Israel, M. Kück, and M. Freitag, "Hybrid approach for the integrated scheduling of production and transport processes along supply chains," *International Journal of Production Research*, vol. 56, 2018.
- [28] J. Huang and J. Song, "Optimal inventory control with sequential online auction in agriculture supply chain: an agentbased simulation optimisation approach," *International Journal of Production Research*, vol. 56, no. 6.
- [29] S. K. Shukla, M. K. Tiwari, H.-D. Wan, and R. Shankar, "Optimization of the supply chain network: Simulation, Taguchi, and psychoclonal algorithm embedded approach," *Computers & Industrial Engineering*, vol. 58, no. 1, pp. 29-39, 2// 2010.
- [30] B. Unhelkar, *Practical object oriented design*. Thomson Social Science Press, 2005.
- [31] J. Arthur F. Veinott, "Lectures in Supply-Chain Optimization," S. U. Department of Management Science and Engineering, Ed., ed. Stanford, California, 2005.
- [32] B. Fahimnia, L. Luong, and R. Marian, "Genetic algorithm optimisation of an integrated aggregate production–distribution plan in supply chainsn," *International Journal of Production Research*, vol. 50, no. 1, pp. 81-96.
- [33] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. MA: Addison-Wesley, 1989.
- [34] N. M. Razali and J. Geraghty, "Genetic Algorithm Performance with Different Selection Strategies in Solving TSP," presented at the Proceedings of the World Congress on Engineering 2011 Vol II, London, U.K., 2011.
- [35] B. Fahimnia, "An Integrated Methodology for the Optimisation of Aggregate Production-Distribution Plan in Supply Chains," *Doctor of Philosophy, Mechanical and Manufacturing Engineering, University of South Australia*, 2010.
- [36] R. Marian, L. Luong, and K. Abhary, "Assembly sequence planning and optimisation using genetic algorithms: part I. Automatic generation of feasible assembly sequences," *Applied Soft Computing* vol. 2, no. 3, pp. 223-253.
- [37] J.-F. Cordeau, M. Gendreau, A. Hertz, G. Laporte, and J.-S. Sormany, "New heuristics for the vehicle routing problem," in *Logistic Systems: Desing and Optimization*, A. Langevin and D. Riopel, Eds. United States of America: Springer, 2005, pp. 279-298.
- [38] R. M. MARIAN, "Optimisation of assembly sequences using genetic algorithms," *DOCTOR OF PHILOSOPHY, School of Advanced Manufacturing and Mechanical Engineering, UNIVERSITY OF SOUTH AUSTRALIA*, 2003.
- [39] F. T. Chan, S. Chung, and S. Wadhwa, "A hybrid genetic algorithm for production and distribution," *Omega*, vol. 33, no. 4, pp. 345-355, 2005.
- [40] Z. H. Abolhasani, R. M. Marian, and L. Luong, "Optimization of Multi-Commodities Consumer Supply Chain- Part 1- Modelling," *Journal of Computer Science*, vol. 9, no. 12, p. 16, 2013.
- [41] H. Xing, X. Liu, X. Jin, L. Bai, and Y. Ji, "A multi-granularity evolution based Quantum Genetic Algorithm for QoS multicast routing problem in WDM networks," *Computer Communications*, vol. 32, pp. 386-393, 2009.
- [42] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, D. E. Goldberg and J. R. Koza, Eds. *Evolutionary algorithms for solving multi-objective problems*, 2nd ed. (Genetic and Evolutionary Computation). New York: Springer, 2007.
- [43] Jeong Hee Hong, K.-M. Seo, and T. G. Kim, "Simulation-based optimization for design parameter exploration in hybrid system: a defense system example," *Simulation: Transactions of the Society for Modeling and Simulation International*, vol. 89, no. 3, pp. 362-380, 2013.
- [44] P. Genin, S. Lamouri, and A. Thomas, "Multi-facilities tactical planning robustness with experimental design," *Production Planning & Control*, vol. 19, no. 2, pp. 171-182, 2008/03/01 2008.

Appendix A

Table 21. The volume of product family P ($v_{P=1:i}^{G=\{M,L\}}$) used in Medium and Large -scale problem (non-linear Constraint)

Product Family	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}
Purchase Cost (\$)	72	51	16	74	100					
$v_i^{G=S}$	3	4	2	5	1					
Purchase Cost (\$)	161	138	148	185	162	113				
$v_i^{G=M}$	5	6	1	3	5	3				
Purchase Cost (\$)	103	167	159	197	171	160	118	109	178	104
$v_i^{G=L}$	1	2	3	3	4	5	4	1	6	2

Note: The volume of product family P_1 in the Medium-scale problem is $v_1^M = 5$

Table 22. The DEMAND QUANTITY ASSOCIATED TO product FAMILY i ordered by consumer j at PERIOD t for P_5

		RE_1	RE_2			RE_1	RE_2
T_1	P_1	50	60	T_2	P_1	64	74
	P_2	1	64		P_2	42	96
	P_3	32	74		P_3	18	29
	P_4	7	25		P_4	40	45
	P_5	10	78		P_5	24	69

Table 23. The DEMAND QUANTITY ASSOCIATED TO product FAMILY i ordered by consumer j at PERIOD t for P_M

		RE_1	RE_2	RE_3	RE_4	RE_5	RE_6	RE_7	RE_8	RE_9	RE_{10}	RE_{11}
T_1	P_1	58	4	37	98	76	81	8	4	52	94	76
	P_2	17	6	54	8	100	43	60	95	100	48	100
	P_3	15	81	72	59	19	73	92	77	86	24	97
	P_4	48	46	88	42	79	50	20	56	97	40	54
	P_5	91	39	33	31	20	81	44	19	68	71	97
	P_6	56	79	66	27	100	36	75	50	41	56	12
T_2	P_1	6	44	48	20	18	88	34	96	10	4	36
	P_2	31	55	26	20	97	79	60	55	47	21	79
	P_3	59	72	37	33	41	47	91	55	1	46	44
	P_4	54	2	67	89	85	82	71	32	92	13	44
	P_5	91	81	17	48	62	90	38	8	65	1	5
	P_6	55	15	28	41	38	43	74	19	1	73	5
T_3	P_1	10	49	47	44	33	3	37	86	69	35	81
	P_2	60	23	64	89	81	61	21	5	91	42	7
	P_3	25	23	92	40	100	12	45	70	62	16	96
	P_4	85	54	17	18	99	41	96	98	90	82	50
	P_5	86	77	72	64	13	89	13	29	20	63	76
	P_6	97	35	58	63	24	55	48	14	76	74	75
T_4	P_1	84	90	41	26	17	75	53	69	76	61	55
	P_2	16	59	4	33	19	70	33	24	99	86	21
	P_3	46	59	75	41	10	83	84	46	24	99	22
	P_4	62	86	16	41	33	83	82	39	53	93	33
	P_5	94	4	15	39	77	30	56	54	6	41	10
	P_6	84	89	61	61	24	31	27	100	76	1	75
T_5	P_1	75	90	75	79	23	17	5	69	81	80	30
	P_2	55	36	13	37	36	84	22	97	24	33	41
	P_3	34	55	83	75	29	17	40	44	94	23	87
	P_4	84	35	3	90	93	51	34	95	77	32	62
	P_5	56	63	42	25	6	100	23	1	83	59	100
	P_6	96	80	74	13	60	36	94	62	58	83	21

Table 24. The DEMAND QUANTITY ASSOCITED TO product FAMILY i ordered by consumer j at PERIOD t for P_j

	RE_1	RE_2	RE_3	RE_4	RE_5	RE_6	RE_7	RE_8	RE_9	RE_{10}	RE_{11}	RE_{12}	RE_{13}	RE_{14}	RE_{15}	RE_{16}	RE_{17}	RE_{18}	RE_{19}	RE_{20}	RE_{21}	RE_{22}	RE_{23}	RE_{24}	RE_{25}	
T_1	P_1	32	92	33	68	98	7	10	39	73	48	33	38	69	4	37	27	67	22	26	41	64	59	18	70	61
	P_2	60	43	24	58	65	18	76	82	17	76	66	1	45	97	37	43	5	6	43	30	56	3	43	9	14
	P_3	45	40	48	40	85	63	87	52	65	99	2	5	60	10	20	23	12	64	77	46	47	44	94	20	51
	P_4	41	61	49	73	71	77	41	32	8	25	96	3	21	4	78	62	53	7	32	37	39	70	83	88	52
	P_5	59	18	20	38	12	54	99	59	24	21	17	91	78	49	22	19	94	23	60	24	83	90	51	38	11
	P_6	85	7	23	66	81	46	2	90	89	55	28	5	35	82	27	80	39	36	82	60	94	100	72	14	27
	P_7	29	38	71	27	31	78	77	38	92	69	22	23	52	74	80	66	79	45	44	76	71	4	87	40	7
	P_8	66	38	25	84	25	41	43	3	46	50	33	80	68	60	2	99	63	92	95	33	3	71	2	23	64
	P_9	17	5	53	77	60	41	62	44	67	74	29	91	60	31	76	8	79	34	7	31	58	96	97	88	31
	P_{10}	89	63	63	27	44	9	43	5	34	84	61	31	97	75	66	2	88	13	86	52	1	5	74	6	66
T_2	P_1	23	74	81	32	73	63	90	46	80	94	68	94	95	7	60	94	12	54	48	96	17	23	25	4	66
	P_2	18	30	87	13	26	49	96	86	91	93	66	40	45	28	75	83	17	7	47	6	27	63	61	88	79
	P_3	61	66	66	99	86	15	45	50	24	41	33	37	55	68	41	91	57	29	54	52	76	9	57	33	75
	P_4	85	45	16	40	44	23	16	42	70	16	76	99	59	92	12	7	34	37	58	98	60	71	55	85	26
	P_5	26	46	61	12	18	18	21	27	93	20	79	60	91	60	11	75	7	7	42	64	63	80	96	92	79
	P_6	17	62	74	10	91	38	43	62	69	67	52	34	50	34	32	58	21	35	61	37	34	10	67	4	77
	P_7	27	79	6	82	52	91	84	38	97	1	60	26	87	40	50	70	86	91	89	56	33	94	98	7	47
	P_8	18	41	62	54	35	31	84	44	43	67	6	41	41	98	65	23	76	10	58	85	94	15	33	60	65
	P_9	53	74	6	25	14	2	31	46	35	43	84	20	89	17	99	52	78	57	50	72	17	44	79	13	78
	P_{10}	62	87	81	26	62	3	11	80	16	32	46	21	41	51	76	12	19	84	16	95	78	6	83	76	59
T_3	P_1	38	50	78	66	43	80	19	41	85	53	42	80	86	11	38	30	96	48	94	97	80	41	57	91	43
	P_2	32	73	63	53	59	33	88	6	11	75	10	6	15	16	86	6	62	84	41	2	20	67	59	57	17
	P_3	9	28	34	33	77	74	74	60	48	40	10	76	34	82	12	25	14	68	48	6	45	3	61	65	61
	P_4	65	64	88	47	27	26	99	66	62	93	93	24	71	33	12	95	26	28	43	45	71	25	80	97	64
	P_5	1	89	71	52	35	48	96	96	4	35	46	18	54	4	54	60	2	27	99	14	99	29	70	78	87
	P_6	43	73	37	77	60	2	73	5	6	20	82	89	40	27	22	56	82	98	80	7	42	28	57	47	88
	P_7	5	33	46	57	58	9	68	36	81	82	61	50	18	58	78	32	2	32	39	16	53	9	18	15	96
	P_8	29	42	72	71	32	92	87	6	64	10	32	87	72	81	100	12	19	28	77	4	67	20	55	44	86
	P_9	99	67	73	22	97	89	20	93	29	30	39	23	17	84	47	63	82	97	58	17	96	45	21	67	89
	P_{10}	48	95	41	50	67	41	12	49	72	80	39	50	92	50	49	3	25	99	66	87	62	84	55	10	44
T_4	P_1	81	34	10	12	36	99	33	44	95	86	71	58	42	28	9	64	54	1	52	99	32	19	26	62	90
	P_2	82	61	59	94	46	54	33	75	55	64	23	74	77	100	15	9	96	20	45	91	85	45	27	86	16
	P_3	44	80	25	49	87	96	71	48	49	95	59	10	80	35	61	81	33	57	11	28	80	21	31	49	95
	P_4	48	46	75	32	13	62	70	7	98	25	44	43	59	38	89	45	32	71	76	25	22	75	24	15	26
	P_5	33	16	90	51	12	2	79	88	88	46	14	46	24	11	37	42	5	78	6	89	95	32	74	19	27
	P_6	22	37	45	26	71	32	1	16	80	10	23	54	83	85	91	17	71	83	17	93	10	10	88	4	88
	P_7	17	52	39	79	6	82	18	100	23	92	62	90	43	11	81	88	55	23	38	40	43	33	69	89	50
	P_8	73	28	75	56	53	37	69	92	79	20	54	92	7	87	4	54	74	60	66	78	100	100	82	15	76
	P_9	28	58	71	75	28	27	22	52	59	57	51	74	94	26	28	86	29	18	99	97	70	9	58	51	22
	P_{10}	60	52	2	45	59	18	88	11	35	28	51	50	85	60	88	40	43	25	6	79	37	53	51	64	9
T_5	P_1	47	12	3	29	57	31	59	71	80	37	70	12	87	25	72	70	70	84	47	85	12	49	60	32	17
	P_2	75	16	41	37	25	41	81	98	41	73	38	28	24	52	45	61	51	54	63	98	27	31	46	76	19
	P_3	31	44	98	4	66	2	14	76	16	86	36	47	66	21	77	46	38	10	57	45	26	92	71	58	53
	P_4	64	89	8	71	84	2	47	99	90	65	19	73	28	98	62	59	38	7	43	7	81	9	53	31	29
	P_5	57	86	81	91	15	71	44	91	6	80	38	34	63	12	28	57	94	20	95	44	26	79	82	37	89
	P_6	33	58	9	41	72	94	86	29	21	6	38	47	14	21	89	46	24	33	80	75	86	75	64	40	16
	P_7	5	56	21	11	98	42	29	19	96	33	44	9	45	58	49	83	29	21	46	49	88	58	15	35	20
	P_8	43	88	21	48	83	41	11	50	70	28	8	12	93	97	82	39	19	62	19	39	47	32	38	79	52
	P_9	89	53	51	50	76	77	53	9	29	36	26	20	74	3	49	58	27	2	5	43	56	62	87	50	65
	P_{10}	88	5	34	34	2	51	34	13	6	64	8	68	26	27	24	4	46	36	2	86	87	70	41	58	74
T_6	P_1	29	71	70	31	35	30	44	2	57	2	68	100	87	30	8	19	41	72	53	68	35	17	59	12	19
	P_2	63	84	72	4	94	70	46	15	57	66	53	12	44	92	32	16	70	9	44	85	2	44	97	93	90
	P_3	27	52	34	60	66	53	25	69	42	48	33	87	5	31	48	83	47	47	59	38	84	67	17	73	91
	P_4	5	41	92	97	82	83	93	83	93	83	29	86	36	70	98	35	81	67	11	93	67	59	29	7	82
	P_5	5	46	99	88	53	99	32	12	17	69	100	7	6	92	80	81	90	44	55	31	56	46	57	49	94
	P_6	87	2	89	80	73	77	48	37	78	62	10	77	69	19	14	71	62	10	85	14	25	23	39	60	56
	P_7	85	9	91	96	83	52	97	37	98	1	61	28	99	94	45	99	94	38	77	51	33	6	87	27	78
	P_8	6	71	70	62	85	47	33	42	48	69	40	81	55	3	65	89	12	82	100	81	84	37	93	14	87
	P_9	81	7	56	100	6	85	22	3	86	41	9	84	74	27	2	26	87	55	22	20	96	98	17	88	10
	P_{10}	75	91	24	84	87	22	7	60	93	5	90	46	13	80	51	11	3	9	22	77	76	74	39	94	69
T_7	P_1	25	91	72	2	28	5	42	15	76	59	32	16	45	82	62	5	87	79	6	25	5	36	19	35	98

CNN-based Automatic Coating Inspection System

Lili Liu¹, Estee Tan¹, Zhi Qiang Cai^{*,2}, Xi Jiang Yin¹, Yongda Zhen^{*,1}

¹Department for Technology, Innovation and Enterprise, Singapore Polytechnic, Singapore, 139651

²School of Electrical and Electronic Engineering, Singapore Polytechnic, Singapore, 139651

ARTICLE INFO

Article history:

Received: 30 November, 2018

Accepted: 13 December, 2018

Online: 26 December, 2018

Keywords:

Coating corrosion assessment

Deep transfer learning

Instance aware segmentation

Hidden corrosion assessment

Active thermography

ABSTRACT

The application of protective coatings is the primary method of protecting marine and offshore structures from corrosion. Coating breakdown and corrosion (CBC) assessment is a major aspect of coating failure management. Evaluation methods can result in unnecessary maintenance costs and a higher risk of failure. To achieve a comprehensive collection of data for CBC assessment, an unmanned aerial system (UAS), assisted by the latest technological innovations, will be used to facilitate data collection in inaccessible locations. A convolutional neural network (CNN)-based CBC assessment system is developed to provide objective assessment of the severity of coating failure. This method is more suitable for inspecting large areas by capturing and analyzing pictures/videos of the target area than the surveyor's existing manual inspection solution. In this paper, deep learning-based object detection in the CBC assessment system has been developed to provide an effective CBC assessment for the marine and offshore industries. By using active thermal imaging, it can identify corrosion behind the coating. This will greatly improve the efficiency and reliability of coating inspection.

1 Introduction

Protective coatings are the main route used to prevent corrosion of marine and offshore structures. Checking the protective coating is a key issue for asset management, but traditional visual inspection methods are time-consuming and labor-intensive. The project aims to develop an automatic coating inspection system for corrosion management applications. The developed system will be able to quickly and comprehensively screen and evaluate coating conditions and can be used as a scanning tool to help investigate the identification and classification of coating failures.

In order to achieve a comprehensive collection of data for coating condition assessment, a micro-aerial vehicle (MAV) assisted by the latest technological innovations, will be used to facilitate data collection in inaccessible locations. An image-based coating breakdown and corrosion (CBC) assessment system has been developed to provide objective assessment of the severity of coating failure. After inspection is completed, the

automatic CBC assessment system (A-CAS) is able to generate an inspection report. The inspection method developed and the test data obtained in this paper can assist the inspection and maintenance team to improve efficiency and productivity. It reduces maintenance cost, improves structural integrity and minimizes risk of failure. Compared to conventional inspection, A-CAS can reduce coating failure evaluation time and manpower.

The structure of the paper is as follows. Section 2 lists related works, Section 3, Section 4 and Section 5 present innovative detection solutions for CBC evaluation, introduces image-based CBC evaluation methods, and explains non-destructive hidden corrosion assessment method. Finally, conclusions are drawn in Section 6.

2 Background

This paper is an extension of work originally presented in the 13th IEEE conference on industrial electronics

*Zhi Qiang Cai, School of Electrical and Electronic Engineering, Singapore Polytechnic, Singapore, 139651, 65 - 6772 1542 & cai_zhi_qiang@sp.edu.sg

*Yongda Zhen, Department for Technology, Innovation and Enterprise, Singapore Polytechnic, Singapore, 139651, 65 - 6772 1455 & zhen_yongda@sp.edu.sg

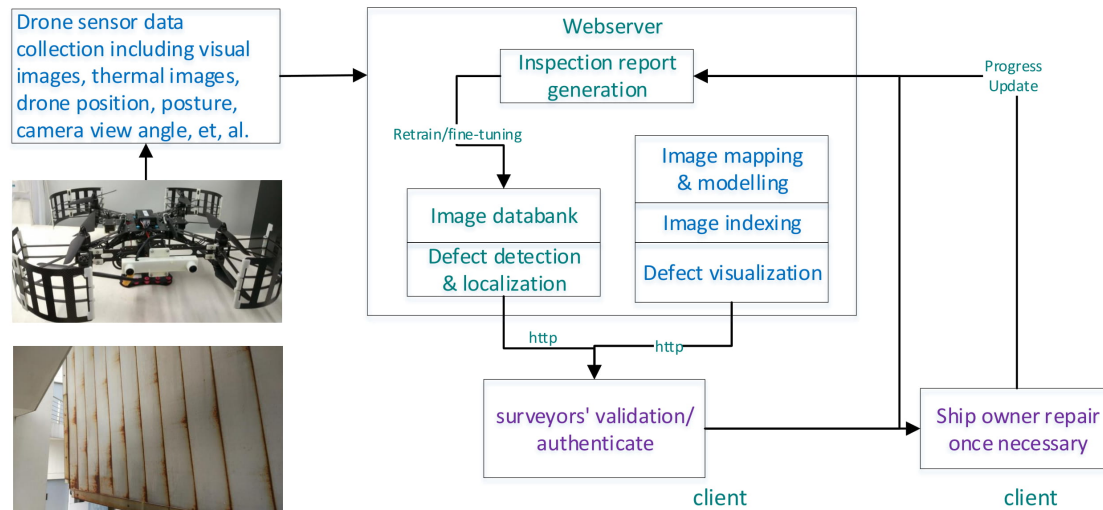


Figure 1: UAS for automated CBC assessment

and applications (ICIEA), 2018 [1]. Referring to automated vision-based coating corrosion detection, to the best of the authors' knowledge, In [2], Jahanshahi and Masri utilize color wavelet-based texture analysis for corrosion detection. Ji et al. [3] apply watershed transform over the gradient of gray images. Siegel et al. [4] choose wavelets for characterizing and detecting corrosion texture in airplanes, Zaidan et al. [5] focus on corrosion texture using standard deviation and entropy as discriminating features. Last but not the least, Ortiz et al. [6] present a solution for coating breakdown and corrosion detection which adopts a semi-autonomous MAV, and an artificial neural network (ANN) to discriminate between pixels suspected/not suspected of corresponding to coating breakdown and corrosion areas through sufficient color and texture descriptors.

2.1 UAS Facilitated Data Collection

Nowadays, visual inspection is widely used in both industry and daily life for security or maintenance checking purposes. Sometimes, visual inspections take place in areas that put the inspector in highly risky situations, hence extra precautions have to be taken, resulting in increasing inspection costs and turnaround time. All these negative factors create difficulties in inspection and would delay the subsequent processes. To meet the needs of inspection, different types of unmanned vehicles were suggested and introduced to replace manual inspection. In coming years, as the technology in aircraft flight control and video processing and transmission systems becomes mature, unmanned aerial system (UAS) would be a solution. It is imperative to use MAVs to replace in-situ human inspection in order to save the cost of preparing the vessel for inspection. Indoor vessel inspection requires the aerial platform to be capable of stable, low speed flight and have anti-collision functions. For research purposes, we customized a MAV with a protective cage and a RGB-D camera for indoor anti-collision coating failure inspection to facilitate our data collection.

In this study, we propose a UAS facilitated CBC assessment system as shown in Figure 1. First of all, a customized autonomous AeroLion drone equipped with necessary inspection tools conducts vessel condition inspection; semi-autonomous navigation technology fulfills the flight capability requirement in indoor environments. The data collected from the drone are then transferred to web-server for post-image analysis including CBC detection and localization in the reconstructed 3D model. Image mapping and image indexing technologies are being developed for traceability of the inspection. The customized Aerolion drone is integrated with a Nvidia Jetson TX2 embedded processor and a ZED RGB-D stereo camera for visual-based simultaneous localization and mapping (vSLAM) mapping, navigation and data collection. Due to the drone's computational power limitation, post-image processing on a separate platform is required for detailed CBC assessment. Figure 1 visualizes an integrated post-processing system. First, a web-server will integrate CBC analysis, image mapping and the CBC visualization model. Then, the HTTP-based client application can be accessed through mobile tablets. Finally, AI-based autonomous CBC inspection results are sent to the surveyor for result verification and report generation. In this paper, we will focus on the automated CBC assessment system.

Instead of image data analysis by certified surveyors, this paper presents one deep learning based visual CBC detection system. The system can automatically recognize and analyze the coating condition by classifying and determining the type of CBC. Corrosion includes but is not limited to CBC on edges, CBC on welds, and CBC on surfaces including hard rust and pitting. Due to the limited size of the available dataset for system development, a transfer learning network was applied to learn the model for CBC detection. An AI-assisted CBC assessment system was successfully developed for the maritime industry. In addition to the visual camera, an infrared/thermal camera will be used for inspection of corrosion behind coating. Ther-

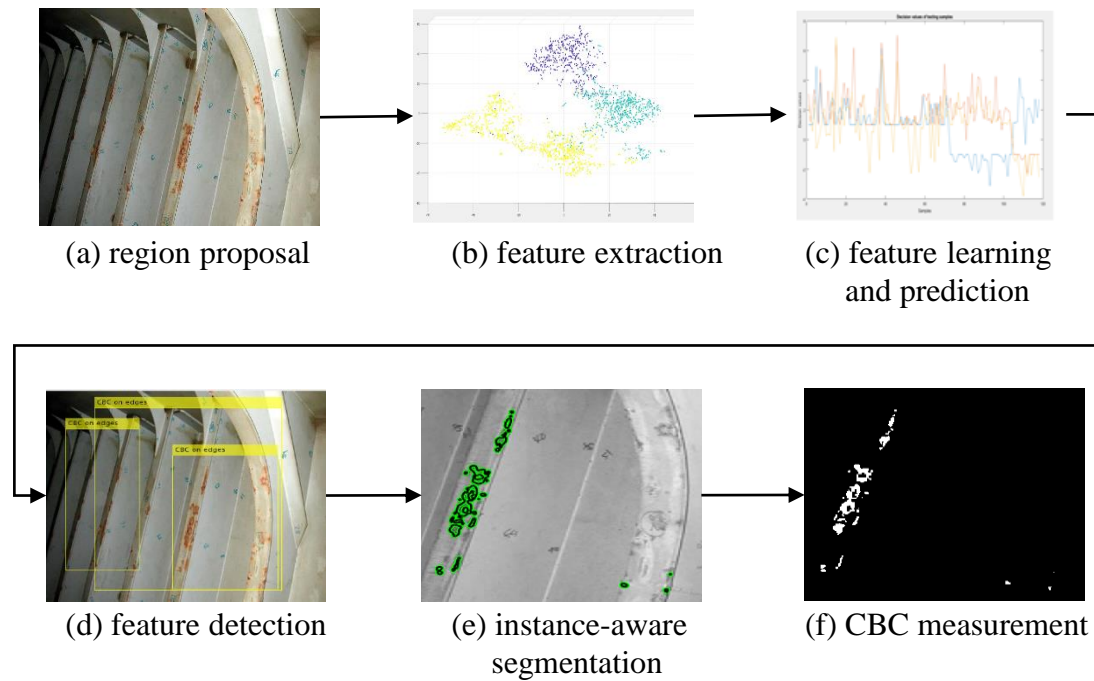


Figure 2: A-CAS system process

mal radiation from areas of the ship's surface with CBC will be different from that without CBC, and this difference can be identified and visualized in some cases. A technology of active infrared thermography to detect the hidden CBC in the vessel structures by combining an external heating source and infrared thermography is developed during the study. The thermal images collected by the drone will be also analyzed by the CBC detection system.

3 Autonomous CBC Assessment System

A vision-based protective coating inspection system was developed for automatic CBC assessment. With this program, even a relatively inexperienced surveyor will be able to make an objective judgment on the severity of CBC. This would improve the efficiency and reliability of coating inspection while reducing the time and manpower cost required. Algorithm development is divided into five phases. They are: region of interest (ROI) proposals, CBC feature extraction, CBC feature learning and prediction, CBC detection and CBC measurements.

In this paper, CBC detection and segmentation are combined for CBC instance aware segmentation. Instance aware segmentation is challenging because it requires proper detection of all CBCs in the image. In practice, due to the limited data set, transfer learning on convolutional activation feature (TLCAF) network [7] is used for CBC feature extraction and feature learning. Models are generated for CBC prediction, segmentation and measurements.

As shown in Figure 2 (b), labeled features are visualized in 3D space and the three categories including

surface CBC, non-coating-failure and edge CBC can be differentiated in hyperplane. Figure 2 shows a typical example of the instance-aware semantic segmentation algorithm for coating failure assessment to speed up the inspection process. The predicted CBC ROIs are reconstructed by background removal. The ROIs are then sent to the next step for active segmentation [8] and CBC measurement. Figure 2 (e, f) demonstrates edge detection and active segmentation, where the background has been masked by active segmentation. The CBC assessment results are more accurate and constant than the ground truth images indicated in [9].

3.1 Feature Extraction and Visualization

Figure 3 lists the five classes of CBC. Different ROIs are manually labeled as different categories. The three categories used in this algorithm are surface CBC, edge CBC and non-coating-failure. Surface CBC can be further categorized into hard rust and pitting. Edge CBC can be divided into CBC on edges and CBC on welds. In order to balance dataset, the three major categories CBC including surface CBC, edge CBC and non-coating-failure are used for prediction for this work.

During the study, TLCAF network [7] is used to learn the region of interest for different types of coating failure including surface CBC and edge CBC. Bounding box is used for region of interest proposal, Faster RCNN [10, 11] framework and vgg19 model [10] are used for convolutional neural network (CNN) feature extraction, and a linear classifier is used for feature classification. Then, the ROIs are reconstructed for different types of coating failure for instance-aware semantic segmentation. Here, active segmentation[8] is used for background removal. The seeds generated

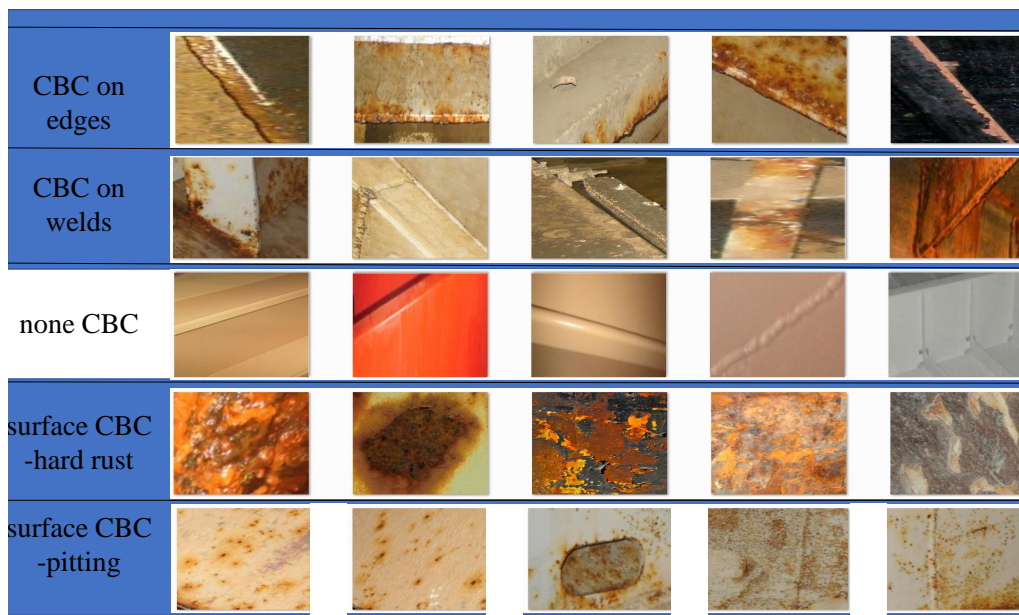


Figure 3: CBC classification

in the previous step are used for auto-guided active selective propagation. Ten iterations of propagation can generate the optimal result which matches with the ground truth which is the value provided by human surveyors' assessment.

2000 randomly selected CBC image features were chosen for the t-distributed stochastic neighbor embedding (t-SNE) [12] high-dimension feature visualization experiment. Figure 4 represents the distribution of three types of CBC features through t-SNE. High-dimension CNN features of the three types of CBC including surface-based CBC (yellow), edge-based CBC (green) and non-coating-failure (blue) can be easily distinguished in hyper-plane for feature prediction.

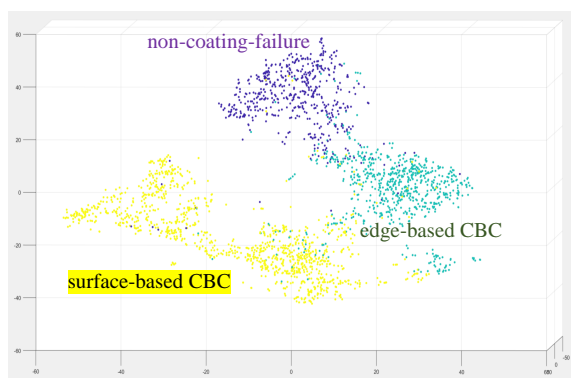


Figure 4: Feature visualization in t-SNE

Figure 5 represents a high-level features learnt from the three categories of extracted CBC features using deep dream. It shows that features for CBC on surfaces are some red dots with sharp gradient descent, while features for CBC on edges/welds are red dots along lines which means the corrosion happens on edges/welds. The non CBC features do not show any red corrosion color or texture. After extracting the R-CNN features, softmax [13] is used for multi-class

feature classification and prediction.

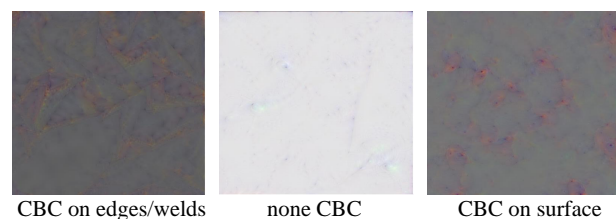


Figure 5: Feature visualization through deep dream

3.2 Feature Learning and Prediction

In the study, 1900 images with CBC were labeled for feature extraction; 12,184 features were extracted and divided into three broad categories, namely edge CBC, non-CBC and surface CBC. A random selection of 2437 features (20% of total features) was used for verification. In the experiment, vgg19 was chosen for network training. Stochastic gradient descent with momentum was used for loss function optimization; the learning rate was set to 1e-6; the maximum number of epochs for training was 20; and a mini-batch with 487 observations was used at each iteration. The max iteration was 9740. Validation frequency was set to 487 iterations duration. Validation patience value, the number of times that the loss on validation set can be larger than or equal to the previously smallest loss before network training stops, was set to five. During the experiment, the training process stopped when it reached the final iteration. The confusion matrix in Figure 7 demonstrates the verification accuracy. The total recognition accuracy rate observed was 89.54%.

The recognition accuracy can be improved by increasing the dataset. Continuous data augmentation will benefit the development of the CBC inspection system, improve CBC generalization and increase accuracy.

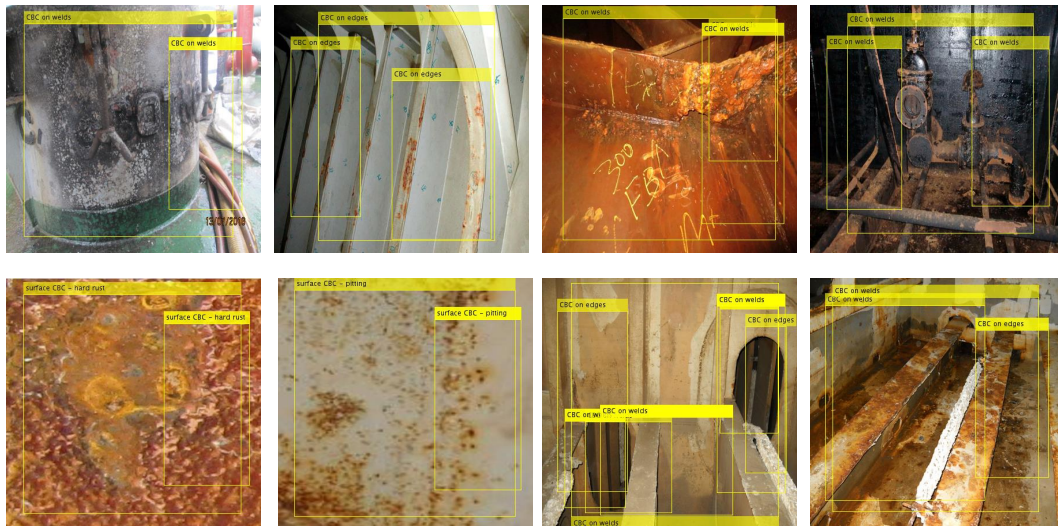


Figure 6: Detected CBC

Accuracy: 89.54%

Output Class	Target Class		
	edges/welds CBC	none CBC	surface CBC
edges/welds CBC	87.0% 703	2.6% 24	10.4% 73
none CBC	2.6% 21	94.8% 876	4.1% 29
surface CBC	10.4% 84	2.6% 24	85.5% 603

Figure 7: Feature prediction result

3.3 Feature Detection

Randomly generated bounding boxes are used to propose ROIs for different categories of CBC prediction. The predicted CBC ROIs are reconstructed with background removal. Instance-aware segmentation was developed for CBC instances' background removal.

The predicted CBC ROIs are reconstructed with background removal. Then, the reconstructed image is sent to the next step for active segmentation [8] and CBC measurement. A database with thousands of photos related to coating failure and corrosion was built up, which is expanding continuously as additional photos are collected from inspections. The system accurately identifies the main types of features, including surface CBC and edge CBC. Figure 6 shows classic examples of ROIs for CBC evaluation to speed up the inspection process. The CBC measurement is more precise and consistent than the ground truth images indicated in [9]. The ground truth images' assessment results are visually estimated by human surveyors and there is bias between different surveyors.

3.4 CBC Measurement

For CBC measurements, hue, saturation and value (HSV) color space is chosen to represent CBC color features. The extracted HSV data are used to identify possible CBC pixels to generate seeds for CBC detec-

tion. The pre-fetched HSV data contain most of the necessary information for pixel-wise CBC measurement. Our system gives users the flexibility to choose how they want to obtain the image that requires analysis. They can choose to either "Upload" or "Capture" an image with the device's camera. The algorithm 1 describes the detailed work flow. The system accurately identified the main types of CBC characteristics, including surface CBC and edge CBC. The study provides a comprehensive A-CAS system for marine and offshore industries. After post-analysis of the RGB image data (448 * 448 bytes), the system was able to promptly generate the inspection report containing field measurement data and the analysis result, as shown in Figure 8. A web-server for data communication will be developed to integrate the local server and client in autonomous on-line process. With the integration of all the techniques developed in this study, the AI-facilitated drone inspection system is able to assist surveyors to increase their efficiency during vessel condition inspection.

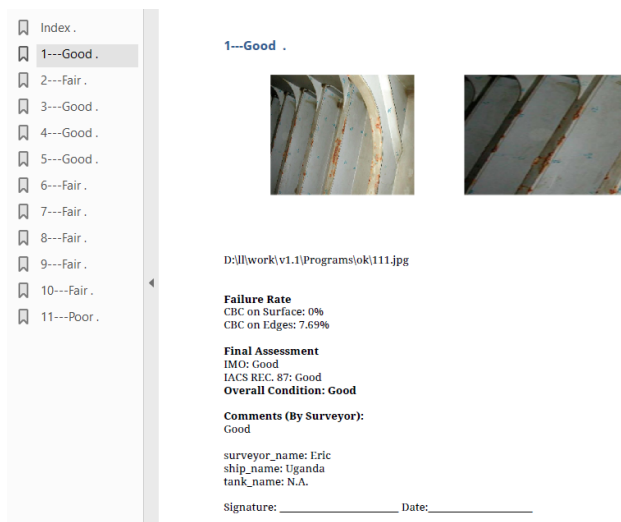


Figure 8: Report generation

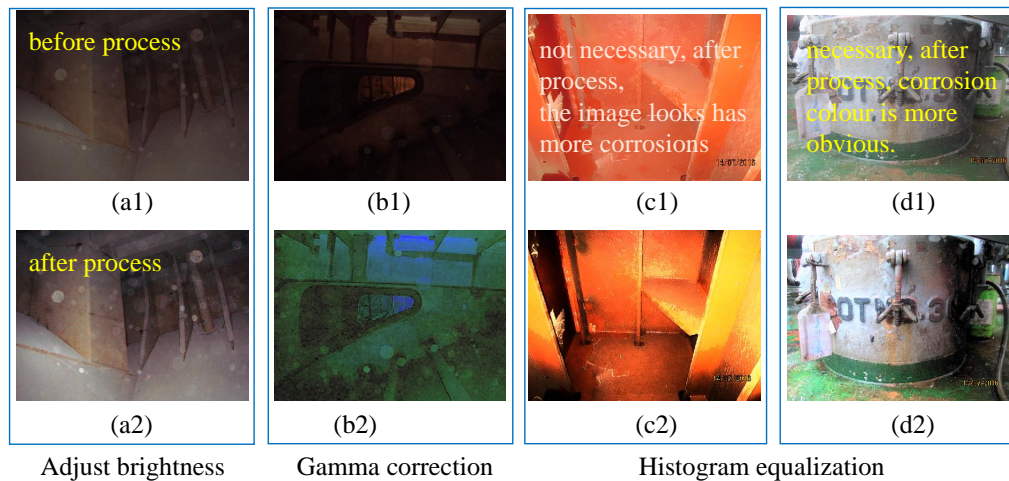


Figure 9: Image Enhancement

Algorithm 1: CBC Assessment System**Input:** The original image**Output:** CBC assessment result

- 1 Input an image and convolve network through different CNN layers;
- 2 Propose relative ROIs;
- 3 ROIs pooling;
- 4 Prediction through fully connected layer with softmax;
- 5 Morphological methods for edge CBC analysis, output processed image;
- 6 If an edge is discovered
 - Edge CBC ROIs reconstruction;
 - Hough transform finds lines;
 - Active selection for propagation;
 - Edge CBC quantitative measurement;
 Endif;
- 7 If surface CBC is detected
 - Surface CBC ROIs reconstruction;
 - Active semantic segmentation;
 - Surface CBC quantitative measurement;
 End-if;
- 8 Adjust the weight for CBC on welds, if any;
- 9 Overall condition grading;
- 10 Generate report;
- 11 The end

The application program can be installed and operated on portable tablets and laptops. The CBC assessment report is automatically generated according to international standards such as IACS Recommendation 87 and IMO recommendations [14] (such as IMO MSC.1/Circ.1330 and IMO MSC.1/Circ.1399). Our program's results are consistent with the ground truth measurements of the surveyors. Our system continues to collect field data to optimize our algorithms and correlate the results with the ground truth results obtained by traditional methods.

4 Active Human Intervention to Improve Accuracy

The section above introduced a fully automated CBC assessment system (A-CAS) for effective coating failure inspection. This method is more suitable for inspecting large areas by capturing and analyzing images of the target area than the surveyor's existing manual inspection solution. However, the machine generated result is not 100 percent correct, and active human intervention and identification can correct and re-adjust these imperfect results. The following section describes how human intervention can fine-tune the result with the help of relevant advanced image processing methods.

4.1 Image Enhancement

Under normal circumstances, lighting will not affect the CBC assessment results. However, non-ideal lighting conditions, such as too high or too low light intensity, can affect the CBC assessment. So a function for adjusting brightness is implemented to facilitate CBC inspection. As shown in Figure 9 (a1) and (a2), we noticed that this function is helpful in certain cases.

Another affiliate function is Gamma correction. By exploiting the non-linear approach of human perception of light and color, Gamma correction works by optimizing the use of bits when encoding images, or transferring the bandwidth of an image. Figure 9 (b1) and (b2) show an example of before and after the Gamma correction process. It shows that this feature is also useful.

The third image enhancement function we implemented is histogram equalization. Histogram equalization was developed to increase contrast. The effect of histogram equalization is demonstrated in the two sets of examples in Figure 9 (c1) and (c2), (d1) and (d2). For CBC, whether histogram equalization helps to improve accuracy is situational dependent. For example, Figure 9 (c1) and (c2) shows that histogram equalization is not necessary, as the dust in Figure 9 (c1) is converted

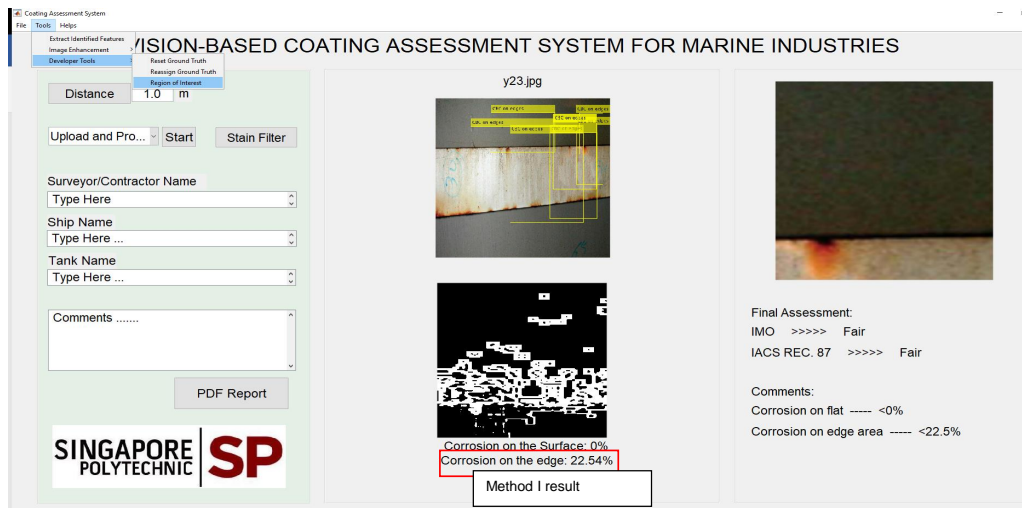


Figure 10: Method I: Auto-report generation

to the CBC color dataset, and the image looks more corroded after the histogram is equalized. For Figure 9 (d1) and (d2), the image before processing, Figure 9 (d1) has low brightness contrast, while Figure 9 (d2) shows that the brightness contrast has increased after processing, and CBC colors in the image better correspond to corrosion colors in the CBC color dataset. Therefore, histogram equalization is necessary for case (d1).

4.2 Ground Truth Adjustment

For special CBC color dataset adjustment, the reassign/reset ground truth function was developed as shown in Figure 11 for result fine-tuning. New special type of HSV values for CBC color can be added to CBC color dataset for CBC measurements.

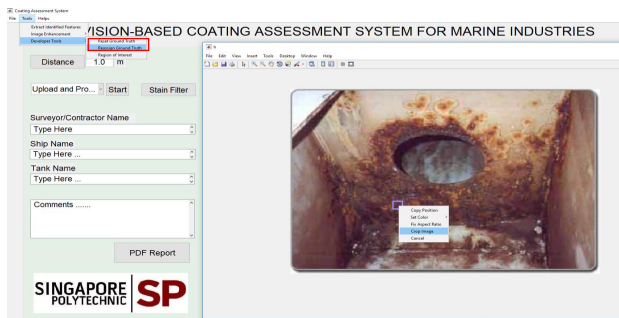


Figure 11: Ground truth adjustment

4.3 Active ROI Selection

There are two ways the system can measure edge CBC. Method I is the system default and the result is auto-generated by the system after input of images through the “Upload and Process”, “Capture and Process” or “Batch Processing” buttons as shown in Figure 12. Figure 10 demonstrates a typical result based on method I. It shows the edge CBC is detected with 22.54%. The ROIs of CBC on edges which are auto-proposed by the CBC detection system is not perfect.

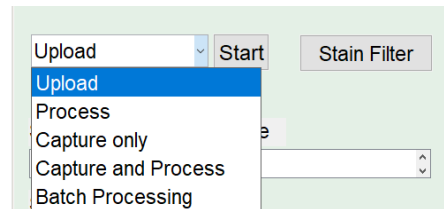


Figure 12: Method I: Auto report generation

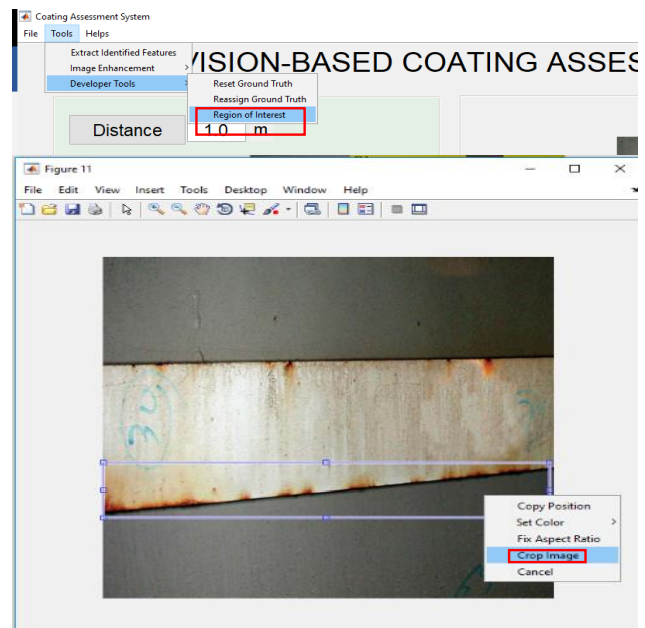


Figure 13: Method II: active ROI selection

Therefore, in order to align with human experience, active ROI selection method, i.e. method II, is developed to let machine interact with human and learn from human. It involves hard assignment of ROI especially for CBC on edges. As shown in Figure 13, the user needs to choose the “Tools” drop-down menu, followed by “Development Tools” and then click “Region of Interest” for ROI selection and edge CBC measure-

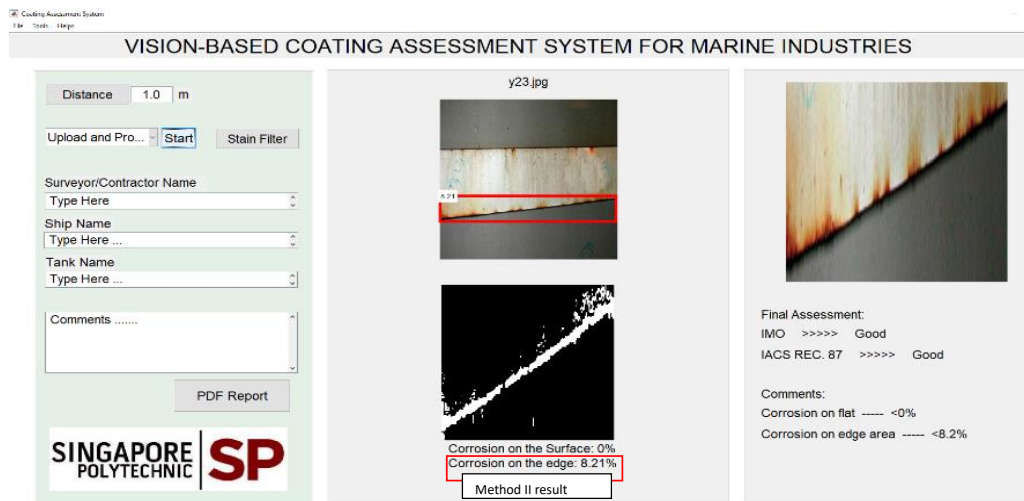


Figure 14: Method II result

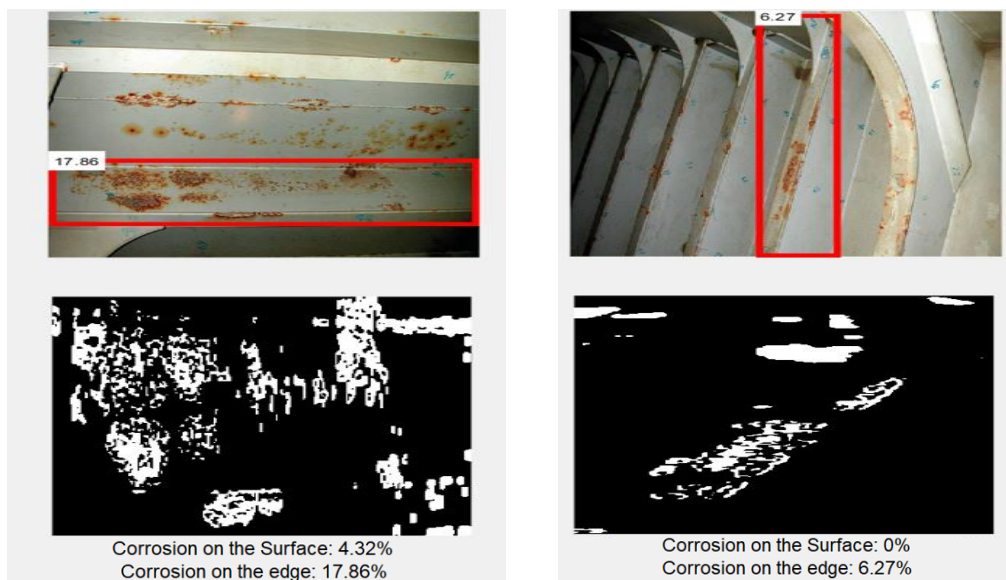


Figure 15: Typical results based on Method II

ment. After that, a new figure pops out to show the original image, where the user can draw a rectangle box on the image to select a ROI, and then right-click to “crop image”. Afterwards, the edge CBC measurement result will be shown in the main GUI in Figure 14. This edge CBC result is 8.21%, which is more accurate than method I. Figure 15 displays two typical results by using method II. The results are more in line with human surveyors’ experience.

5 Non-destructive Hidden Corrosion Assessment

Hidden corrosion behind coated surfaces is a problem that could cause dangerous failure, causing shut down of production processes and high cost to the marine and offshore industry [15, 16]. At present, hidden corrosion can only be visually observed by human

eyesight after cutting and/or coating removal. Compared to existing solutions, active infrared thermography (IRT) is relatively faster and non-destructive. Combined with algorithms for auto CBC detection, condition monitoring with active IRT can maintain higher precision and quality of assessment for consistent defect detection and measurement. While ultrasonic methods are limited to inspection of small areas, active IRT is fit for inspecting large areas, producing thermal images of the target. Active IRT can be used as an auxiliary tool for rough defect area positioning, followed by ultrasonic inspection for precise defect depth measurement. It can also be applied to risk management for cost savings and improving safety in the maritime industry. In the last 10 years, different research institutes have studied active IRT, but this technique has not been widely taken up by industry. The major problem is the long heating duration, making it unfit for on-site assessment. But with advances

in heating technology and sensitivity of thermal cameras, active IRT has potential to “see through” more advanced materials [17, 18, 19].

In this project, a mid-wave thermal imaging camera (FLIR a6702sc) and a halogen heater are chosen for instant heat generation to collect data in near real time for post image processing and analyzing [20]. By using a mid-wave thermal imaging camera FLIR A6703sc in combination with a 1500W halogen heater, the corrosion area could be imaged after only 5-10 seconds of heating. After removing the heat source, a temperature difference develops between the corroded and non-corroded areas, and the thermal contrasts as shown in Figure 16 (a) are captured by the thermal camera. In addition, the video captured is passed through machine learning algorithms for corrosion prediction and assessment. Figure 16 (b) is the processed result of hidden corrosion behind the coating for assessment.

Active infrared thermal imaging developed to detect hidden CBCs can be used to examine areas that are not identifiable by visual inspection. RGB imaged based CBC detection is used for 1st round screening test. Thermal camera is used for 2nd round hidden corrosion assessment once necessary. This will greatly increase the inspection effectiveness and help to predict and reduce the potential risk caused by the hidden corrosion.

6 Summary

The CNN-based automated CBC assessment system developed includes new coating breakdown and corrosion assessment algorithms with higher accuracy to facilitate quantitative data analysis for effective coating failure inspection. Compared to existing manual inspection solutions by surveyors, this method is more suitable for inspecting large areas by means of capturing and analyzing pictures/videos of the target areas.

The system accurately identified the main types of CBC characteristics, including surface CBC and edge CBC. The study provides a comprehensive A-CAS system for marine and offshore industries. A coating condition evaluation report is automatically generated according to international standards such as IACS Recommendation 87 and IMO recommendations. The detection system uses deep TLCAF technology to automate CBC assessment, and an instance-aware semantic segmentation method for CBC measurement and grading was developed. The three main types of CBCs can be distinguished and used for CBC prediction through the last fully connected layer. A model for coating failure prediction is generated, followed by CBC segmentation, CBC quantitative measurement and grading. The developed A-CAS system is going to improve maritime coating inspection works faster and smarter. With the assistance of IRT, the presence of corrosion behind protective coatings will be identifiable. Additionally, the non-destructive approach eliminates the need to remove the coating system.

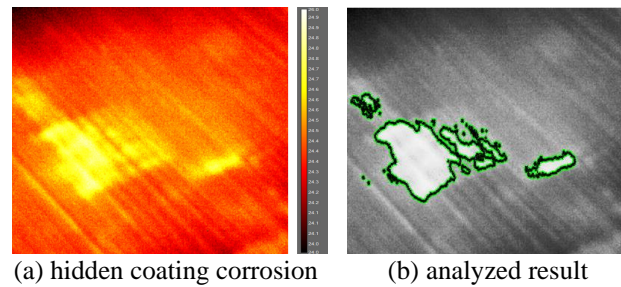


Figure 16: Hidden corrosion

The developed A-CAS system will make maritime coating inspection safer and more efficient. Shipping companies will be able to save on docking costs as well as eliminate losses incurred from goods left on the dock. The process of maritime coating inspection is set to have a makeover in the near future.

The days of cumbersome CBC inspections are gone forever as surveyors’ jobs become highly automated. This automated corrosion assessment system can also be implemented in other industries that also face the problem of inefficient coating inspection. Such industries include the aviation, railway and building industries.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment The author would like to thank Dr. Hai Gu of the American Bureau of Shipping (ABS) for his technical advice and data support. In addition, we are very grateful to Mr. Kelvin Lim Chee Quan, Mr. Ma Yiheng, Mr. Donvis Nguyen, Mr. Yeo Eng Hoe Jason and Mr. Duncan Goh Yitang for their contributions to the study. This work was funded by the Singapore Maritime Institute (grant no. SMI-2015-OF-05) and conducted at SP.

References

- [1] L. Liu, E. Tan, Y. Zhen, X. J. Yin, Z. Q. Cai, Ai-facilitated coating corrosion assessment system for productivity enhancement, in: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2018, pp. 606–610.
- [2] M. Jahanshahi, S. Masri, Effect of color space, color channels, and sub-image block size on the performance of wavelet-based texture analysis algorithms: An application to corrosion detection on steel structures, in: Computing in Civil Engineering (2013), 2013, pp. 685–692.
- [3] G. Ji, Y. Zhu, Y. Zhang, The corroded defect rating system of coating material based on computer vision, in: Transactions on Edutainment VIII, Springer, 2012, pp. 210–220.
- [4] M. Siegel, P. Gunatilake, G. Podnar, Robotic assistants for aircraft inspectors, *Industrial Robot: An International Journal* 25 (6) (1998) 389–400.
- [5] B. Zaidan, A. Zaidan, H. O. Alanazi, R. Alnaqeib, Towards corrosion detection system, *International Journal of Computer Science Issues* 7 (3) (2010) 33–36.
- [6] A. Ortiz, F. Bonnin-Pascual, E. Garcia-Fidalgo, et al., Vision-based corrosion detection assisted by a micro-aerial vehicle in a vessel inspection application, *Sensors* 16 (12) (2016) 2118.

- [7] L. Liu, R.-J. Yan, V. Maruvanchery, E. Kayacan, I.-M. Chen, L. K. Tiong, Transfer learning on convolutional activation feature as applied to a building quality assessment robot, *International Journal of Advanced Robotic Systems* 14 (3) (2017) 1729881417712620.
- [8] S. Dutt Jain, K. Grauman, Active image segmentation propagation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2864–2873.
- [9] A. American Buireau of Shipping, Guidance notes on the inspection maintenance and application of marine coating system third edition, Book, 2007.
- [10] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: *Advances in neural information processing systems*, 2015, pp. 91–99.
- [11] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks* 61 (2015) 85–117.
- [12] L. V. D. Maaten, G. Hinton, Visualizing data using t-sne, *Journal of Machine Learning Research* 9 (Nov) (2008) 2579–2605.
- [13] G. Hinton, R. Salakhutdinov, Replicated softmax: an undirected topic model, in: *Advances in neural information processing systems*, 2009, pp. 1607–1614.
- [14] M. Resolution, 215 (82), “performance standard for protective coatings for dedicated seawater ballast tanks in all types of ships and double-side skin spaces of bulk carriers”, IMO, London, UK.
- [15] P. Traverso, E. Canepa, A review of studies on corrosion of metals and alloys in deep-sea environment, *Ocean Engineering* 87 (2014) 10–15.
- [16] Z. Wang, Y. Cong, T. Zhang, et al., Effect of hydrostatic pressure on the pitting corrosion behavior of 316l stainless steel, *Int J Electrochem Sci* 9 (2014) 778–798.
- [17] Y. Hung, Y. S. Chen, S. Ng, L. Liu, Y. Huang, B. Luk, R. Ip, C. Wu, P. Chung, Review and comparison of shearography and active thermography for nondestructive evaluation, *Materials Science and Engineering: R: Reports* 64 (5) (2009) 73–112.
- [18] J. R. Brown, H. Hamilton, Quantitative infrared thermography inspection for frp applied to concrete using single pixel analysis, *Construction and Building Materials* 38 (2013) 1292–1302.
- [19] L. Liu, I.-M. Chen, E. Kayacan, L. K. Tiong, V. Maruvanchery, Automated construction quality assessment: A review, in: *Mechatronics and its Applications (ISMA)*, 2015 10th International Symposium on, IEEE, 2015, pp. 1–6.
- [20] C. Duberstein, D. Virden, S. Matzner, J. Myers, V. Cullinan, A. Maxwell, Automated thermal image processing for detection and classification of birds and bats, *Offshore Wind Technology Assessment*.

Multi-Objective Path Optimization of a Satellite for Multiple Active Space Debris Removal Based on a Method for the Travelling Serviceman Problem

Masahiro Kanazaki^{*1}, Yusuke Yamada¹, Masaki Nakamiya²

¹Division of Aerospace Engineering, Graduate School of System Design, Tokyo Metropolitan University, 190-0015, Japan

²Department of Aerospace Engineering, Teikyo University, 320-8551, Japan

ARTICLE INFO

Article history:

Received: 14 August, 2018

Accepted: 12 December, 2018

Online: 26 December, 2018

Keywords:

Multiple Space Debris Removal

Trajectory Optimization

Travelling Serviceman Problem

Evolutionary Algorithm

Multi-objective Optimization

ABSTRACT

Space debris removal is currently a critical issue for space development. It has been reported that five pieces of debris should be removed each year to avoid further increase in the amount of debris in orbit. One approach for the removal of multiple pieces of debris is to launch multiple satellites that can each remove one target debris from orbit. The benefit of this approach is that the target debris can be removed without orbit transition, and thus, the satellite can be developed considering simple satellite mechanics. However, to realize this concept, multiple satellites need to be launched. Another approach is to use one satellite to remove multiple pieces of space debris. This approach can reduce the launch costs and achieve efficient removal of space debris. However, the satellite must change its orbit after the removal of each debris piece, and a technique for optimizing the orbit transition is required. In this study, the latter strategy and developed a satellite trajectory optimization method for efficient space debris removal were focused on. The similarity between the problem of multiple space debris removal and the travelling serviceman problem (TSP) were considered, and the TSP solution involving an evolutionary algorithm (EA) was applied. To improve the efficiency of multiple debris removal, the total radar cross-section (RCS), which indicates the amount of space debris, and the total thrust of the satellite was minimized. The TSP solution method was extended to multiple objectives by coupling it with a satellite trajectory simulation. To evaluate the developed method, a set of 100 pieces of space debris was selected from a database. The results indicated a trade-off between the total RCS and total thrust.

1 Introduction

Parts of rockets, spent satellites, and tools left behind during space operations by astronauts continue to remain in orbit as space debris ("debris") [1, 2, 3, 4]. Such debris pose a threat to satellites in operation, including the International Space Station as the debris components may collide with these satellites and cause critical damage. Furthermore, when the amount of debris exceeds a certain value, the involved parts may collide with each other, and further increase the amount of debris. This phenomenon is referred to as the "Kessler syndrome"[5], and a theoretical model has

been developed in this respect. The self-crushing phenomenon called "break up" [6] due to the explosion of residual fuel also increases the amount of debris. Thus, increasing debris is an urgent problem that needs to be solved in future space development, and active debris removal techniques are being studied[7, 8, 9]. To this end, several methods using removal satellites have been studied. For example, the use of extending conductive tethers that connect the removal satellite and a piece of debris, and then eliminate the debris by passing electric current in the Earth magnetic field and providing thrust to the debris by using the generated

*Corresponding Author: Masahiro Kanazaki, 6-6 Hino, Tokyo, Japan, kana@tmu.ac.jp

Lorentz force [7], has been widely investigated. Another possible approach is to attach thrusters to the debris pieces [9] to change their trajectory. Such methods require the satellite to be powered to rendezvous with the debris; therefore, efficient mission scenarios are required.

From the viewpoint of cost-effectiveness, it is desirable to eliminate multiple debris with a removal satellite. In addition to the technology of the removal satellites, the optimization of the order in which debris is collected and/or captured and disposed should be considered. Although the optimization of the meeting order can be achieved by considering general combination optimization problems such as the travelling salesman problem (TSP), the debris moves non-cooperatively towards the dumping satellite in this multiple debris removal problem. Therefore, it is also necessary to investigate the applicability of the TSP solution. In an actual dumping mission, there are multiple objectives such as maximizing the total size of debris to be discarded while minimizing the total amount of energy required for the removal satellite to change its trajectory. In particular, it is predicted that the sum of the sizes of debris represented by the radar cross-section (RCS) and the sum of the trajectory-changing energies associated with each debris are contradictory.

Therefore, in this research, a general methodology for the TSP expanded as a multi-objective problem was applied to the path optimization problem of satellites for active debris removal to eliminate a plurality of debris. The data of 100 observed debris was taken as an example; solve the multi-objective problem by changing the number of target debris to 2, 3, 4, and 5; and determine the trend of the optimum path.

2 Related Study

2.1 Researches on Active Space Debris Removal

In 2017, there were 15,000 pieces of space debris with a size of 10 cm or more. The number of debris pieces is expected to exceed 28,000 in 2116 and more than 65,000 in 2210. However, assuming that five pieces of debris can be removed annually from 2020, the number of debris in 2116 can be reduced to approximately 20,000; if 20 pieces can be removed, this number can be reduced to approximately 16,000, according to the prediction reported in [10]. Although this suggests that multiple debris removal techniques are required, it is expensive to launch and operate multiple removal satellites. Thus, an active debris removal technique using only one satellite operating with path optimization is required.

Conceptual research on multiple debris removal considering cost reduction was carried out at the NASA Goddard Space Flight Center [11]. Conductive tether satellites were adopted in this exercise, and the effectiveness of removal of the assumed 50 pieces of debris was discussed. The optimum trajectory was

calculated, and the weight of the removed debris was evaluated by an estimation formula. The examination was conducted using the TSP solution method. In this study, several case studies focusing on the number of removed debris were examined; however, the required thrust for the removal satellite was not considered in the optimization process and only a single objective was considered.

2.2 Solutions for TSP and TSP-like Problems

The TSP [12], which is a combinatorial optimization problem, is known as a non-deterministic polynomial (NP) problem. To solve the TSP, heuristic methods have been shown to be effective. The problem involves obtaining the shortest travelling distance for a salesman, when the salesman is expected to visit all distributed cities, as shown in Fig. 1. Here, the distance of the path between cities i is d_i . The TSP is one of the most popular problems and an example of evolutionary calculation.

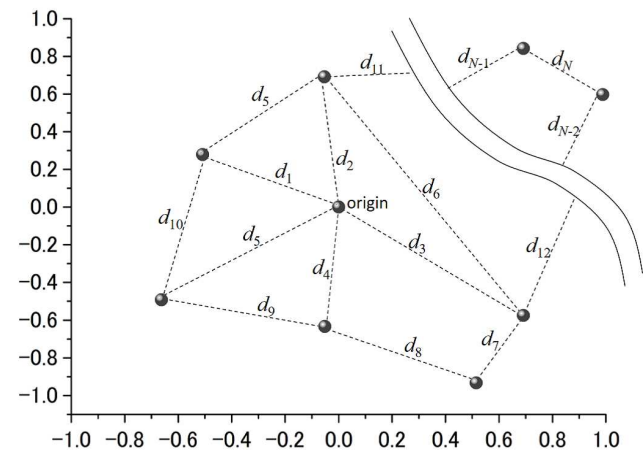


Figure 1: Example of travelling points and routes for the TSP.

A problem similar to the TSP is the "Kyoto Tourism Problem." [13]. When visiting and sightseeing in the city of Kyoto, satisfaction and travel expenses are the factors considered by travelers. The TSP solution method that introduces the concept of the Pareto optimum for the multi-objective problem was applied. In this method, a reciprocal relationship exists between the objectives, and a plan must be proposed based on the travel budget of each traveler.

Another application of the TSP is to reduce a driver's burden and shorten the route for pick up and transfer in consideration of customers and road conditions, which change dynamically [14]. While simulating road conditions and the movement of the customers simultaneously, researchers solved the problem using the evolutionary method. It can be said that this problem is similar to the path optimization problem for active debris removal satellites. However, in reality, it should be noted that debris are uncontrollable

and non-cooperative objects, while customers exhibit cooperative behaviour through communication and self-judgement.

3 Trajectory Optimization for a Multiple Space Debris Removal Satellite

In this research, a plurality of debris is effectively removed by "maximizing the sum total of the RCS of the debris to be removed (RCS_{tot}) and minimizing the sum of the acceleration to move to the i th debris" (ΔV_{tot} which is the amount of ΔV_i for transition).

The debris travels around different orbits (including altitude); thus, it is necessary to increase the speed to meet the orbit in orbit. It is also necessary to minimize the increase in speed to decrease fuel consumption. This problem can be written as follows.

$$\begin{cases} \text{Maximize} & RCS_{tot} \\ \text{Minimize} & \Delta V_{tot} \end{cases} \quad (1)$$

Assuming that N_{debris} pieces of the debris are removed, RCS_{tot} can be expressed as follows:

$$RCS_{debris} = \sum_{i=1}^{N_{tot}} RCS_i. \quad (2)$$

ΔV_{tot} can be written as the summation of the velocity increment ΔV_i for each debris i :

$$\Delta V_{tot} = \sum_{i=1}^{N_{tot}} \Delta V_i. \quad (3)$$

ΔV_i is obtained by solving Lambert's problem [15], which is described in the following subsection.

3.1 Lambert's Problem for Trajectory Evaluation

In this research, the Lambert problem [15] in orbital dynamics was solved to evaluate the path of the removal satellite. This can be rephrased as a problem in which the spacecraft achieves the necessary speed in the orbit to travel from a certain point P to another point Q. The Lambert equation can be expressed as

$$\Delta t = \sqrt{\frac{a^3}{\mu}} [2k\pi + (E - e\sin E) - E_0 - e\sin E_0]. \quad (4)$$

where Δt is the duration of time, a is the semi-major axis, μ is a gravitational parameter, k is the number of revolutions, E is an eccentric anomaly, and e is the orbital eccentricity. ΔV_i is obtained from the target debris position after Δt , the initial position of the debris removal satellite, and Δt is obtained by the expression 4. Figure 2 shows an example of the solution of

the Lambert problem translating the trajectory of the removal satellite with respect to the debris.

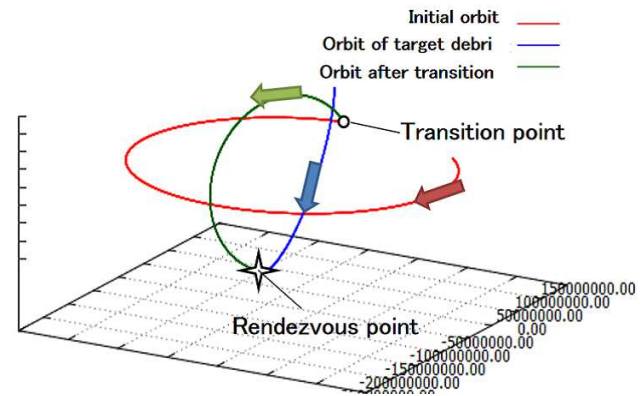


Figure 2: Example of trajectory calculation when the removal satellite heads to the rendezvous point with the debris with one orbit change. Red denotes the original trajectory of the dumped satellite, green denotes the trajectory after the transition, and blue denotes the trajectory of the debris.

4 Combinatorial Optimization and Expansion to Multi-Objective Problem

In this research, the TSP solution methods using a genetic algorithm was applied to the path optimization problem of space debris removal satellites. The following typical algorithm was firstly developed to solve TSP.

- As shown in Fig. 1, it was assumed that all the cities are represented by coordinate points on the plane and N intercity routes are defined.
- A salesman does not visit the same city more than once.

The total travelling distance d_{tot} in Fig. 1 can be written as

$$d_{tot} = \sum_{i=1}^N d_i. \quad (5)$$

4.1 Genetic Operators

4.1.1 Representation of the Combination of the Traveling Path

Path representation (PR) is applied. For example, if the number of visited points is nine, then the order in which the points are visited can be represented as follows::

$$| a | c | b | g | h | i | d | f | e |. \quad (6)$$

4.1.2 Selection

A roulette selection is applied to select individuals for the crossover and mutation operations. In the roulette selection, the selection probability p_i can be based on fitness f_i from N_{pop} :

$$p_i = \frac{f_i}{\sum_{k=1}^{N_{pop}} f_k}. \quad (7)$$

4.1.3 Crossover

Order crossover (OX) [12] is applied to two selected individuals. In OX, the individuals p_1 and p_2 are selected first:

$$\begin{aligned} p_1 &= |a|b|c|d|e|f|g|h|i| \\ p_2 &= |d|a|b|h|g|f|i|c|e|. \end{aligned} \quad (8)$$

Then, the parts of the gene sequence that are copied to children c_1 and c_2 are selected from p_1 and p_2 , respectively. An example is shown in Eq. 9. Here, fourth–seventh genes are copied in that order to the children, and the rest of the genes —which are temporarily represented by $|*|$ — are undecided gene sequence parts to be determined by a crossover:

$$\begin{aligned} c_1 &= |*|*|*|d|e|f|g|*|*| \\ c_2 &= |*|*|*|h|g|f|i|*|*|. \end{aligned} \quad (9)$$

To determine the genes temporarily represented by $|*|$ of c_1 , the following gene sequence of p_2 that corresponds to the undecided gene sequence part of c_1 is copied. The sequential order of this gene sequence remains in the clockwise direction:

$$p_{2b} = |c|e|d|a|b|h|g|f|i|. \quad (10)$$

Here, the gene sequence part that has already been copied from p_1 to c_1 is removed from p_{2b} :

$$p_{2c} = |e|d|b|h|i|. \quad (11)$$

c_1 can be determined by inserting p_{2c} :

$$c_1 = |b|h|i|d|e|f|g|e|d|. \quad (12)$$

c_2 can be determined in the same way:

$$c_2 = |a|b|c|h|g|f|i|d|e|. \quad (13)$$

4.1.4 Mutation

Mutation is used to maintain population diversity, creating individuals with genetic information that cannot be generated only by a crossover. In this research, an inversion mutation that exchanges the genetic information at a position determined by a random number is used. In the inversion method, when p_1 shown in Eq. 8 is chosen as a mutation target, two genes are arbitrarily selected and the positions are exchanged as follows.

$$c_1 = |a|b|e|d|c|f|g|h|i|. \quad (14)$$

In this example, the third and fifth genes of p_1 are exchanged to create c_1 .

4.2 Investigation of the Developed Method with the Typical TSP

As a verification method, a typical TSP (minimizing the total path distance) was solved with 10 randomly distributed cities and the departure point of a salesman. Eight-hundred generations were executed by setting 100 individuals for each generation. It is assumed that the route between the cities is given by a Euclidean distance and the salesman returns to the starting point.

Figure 4 shows the history of the solution (total route distance) obtained by the method outlined in the previous section. From this figure, the total path distance decreases with each generation and converges at approximately 300 generations. An example of the best individual route obtained for each generation is shown in Fig. 1. Figure 4(a), which is the initial generation, has a waste route. Whereas, in Fig. 1(b)-(d), it can be seen that several paths considered to be efficient are being searched. In particular, Fig. 1(c) and (d) show almost the same evaluation value, but the routes are different. This shows that various solutions can be obtained in the design variable space. In this research as well, designers can select the route according to the mission requirements and constraints using the information on the optimum route for the debris removal satellite.

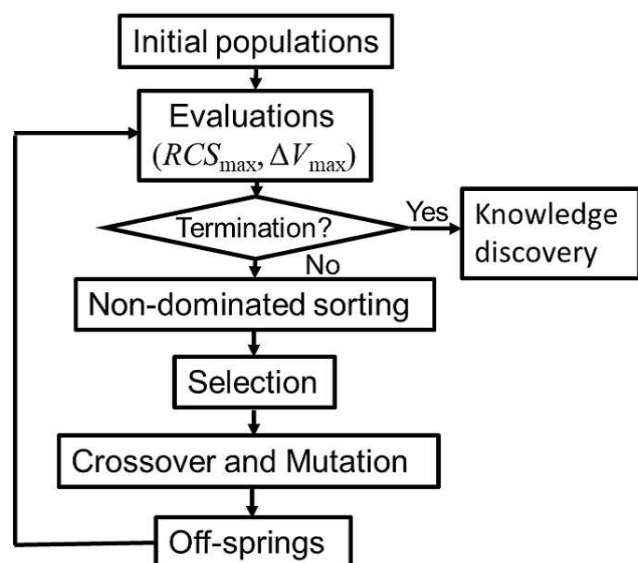


Figure 3: Procedure of multi-objective trajectory optimization of a satellite for multiple active space debris removal.

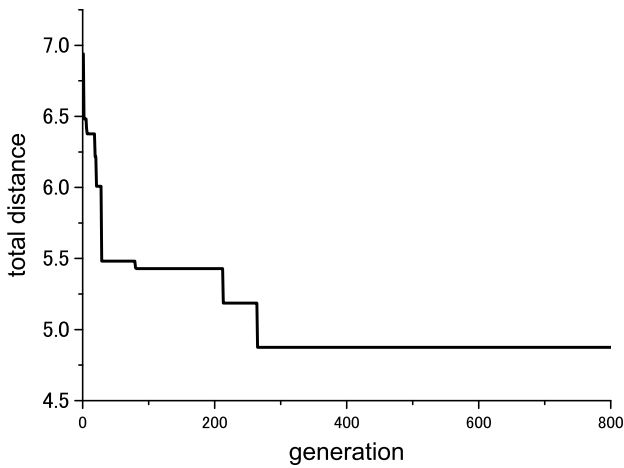


Figure 4: Convergence history of solutions with the developed algorithm

4.3 Non-dominated Sorting for MoPs

The final goal of this research is multi-objective optimization of the orbits of multiple debris removal satellites. For multi-objective optimization, ranking is performed by non-dominated sorting (Fig. 6), which was introduced in the Non-dominated Sorting Genetic Algorithm-II [16]. An elite strategy was adopted to archive good solutions and take over to the next generation without genetic operations such as crossover or mutation.

4.4 Scatter Plot Matrix (SPM)

The solution and design space of the multi-variable design problem obtained by MOEA are observed using SPM [17], which is one of the data mining methods. SPM arranges two-dimensional scatter plots such as a matrix among the objective functions and design variables and facilitates the investigation of the design problem. Each of the rows and columns is assigned attribute values such as design variables, objective functions, and constraint values. The diagonal elements show the same plots. Therefore, it can be said that the SPM shows scatter plots on the upper triangular part of the matrix and the correlation coefficients on the lower triangular part as additional information. modeFrontier™4.2.2 was employed in this study.

5 Results for the Trajectory Optimization of the Multiple Active Space Debris Removal Satellite

In this research, debris data obtained when China destroyed the satellite "Fengyun-1C" in an experiment for the development of an anti-satellite weapon method in 2007 [6] was used. In this experiment, the over 2800 pieces of satellite debris were obtained debris, which is the highest number to date.

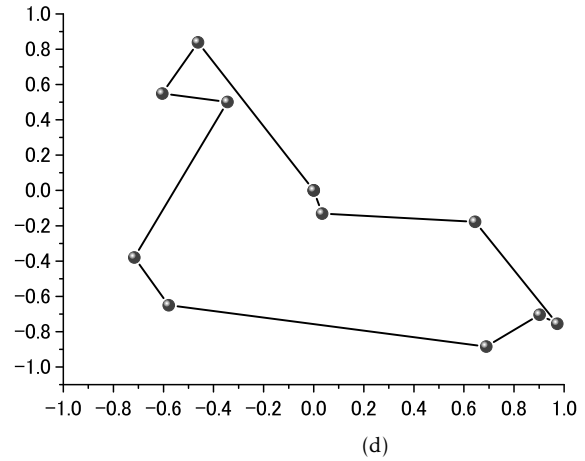
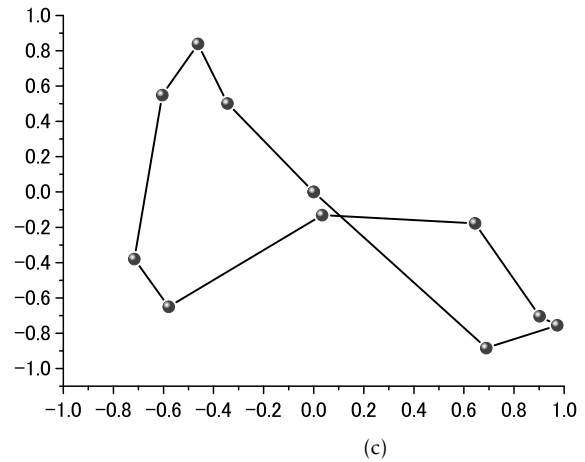
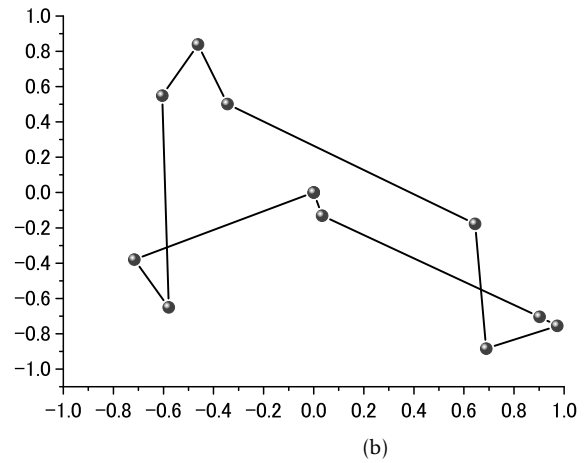
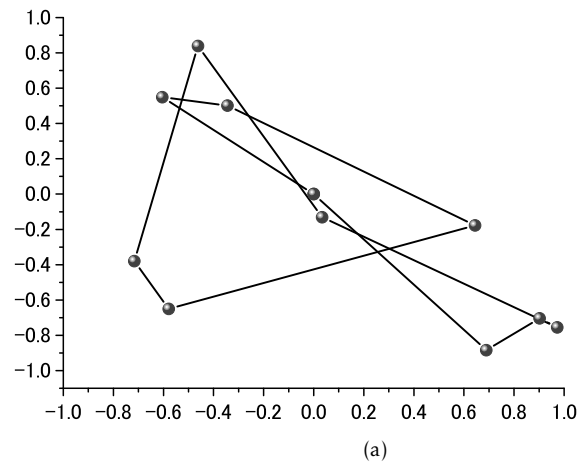


Figure 5: Selected solutions: (a) initial generation, (b) 100th generation, (c)400th generation, and (d) 800th generation

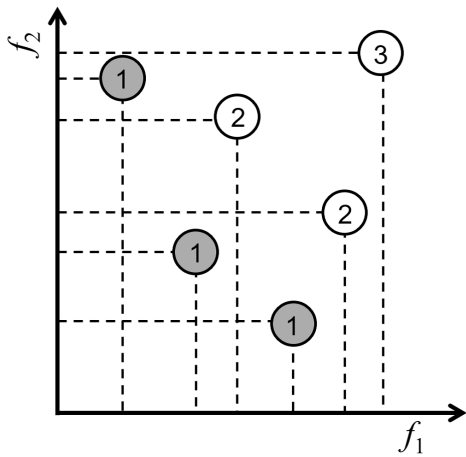


Figure 6: Ranking by the non-dominated sorting algorithm.

The orbital altitude of this debris cloud is located at a high altitude of 800 km; hence, the suspension time is long and is expected to have a negative impact/influence in the long term. Because each piece of debris was tracked from the start, 100 individual debris data which were randomly selected from the catalogue (data table) was used. The frequency distribution of the RCS considered in this research is shown in Fig. 7. RCS_i is based on the observation data published by the North American Aerospace Defence Command.

For optimization, the multi-objective problem shown in the expression 1 in the section 3 was solved. To obtain information on removal efficiency, four cases in which the number of debris to be removed was changed to 2, 3, 4, and 5 were solved. It is assumed that the removal satellite has been in a parking orbit at an altitude of 200 km and departed at 0 o'clock on January 1, 2015. Upon reaching each debris, the satellite conducts a removal operation for 1 h and remains stationary in that orbit until it has to depart for the next debris. Evolutionary calculation was carried out with 100 individuals per generation and 2000 generations were executed.

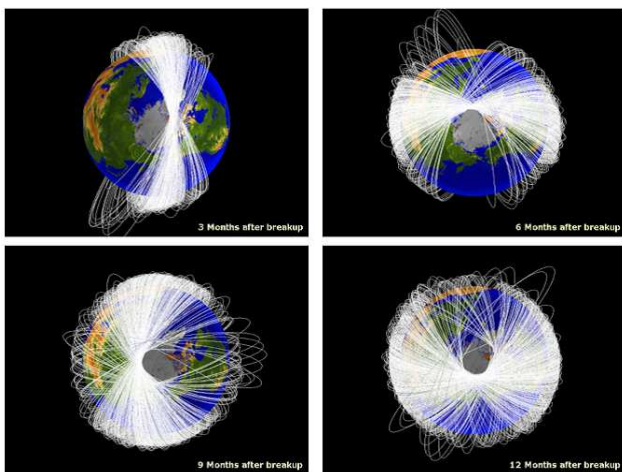


Figure 7: Three months of tracking data for space debris from the satellite destruction test of Fengyun-1C [6].

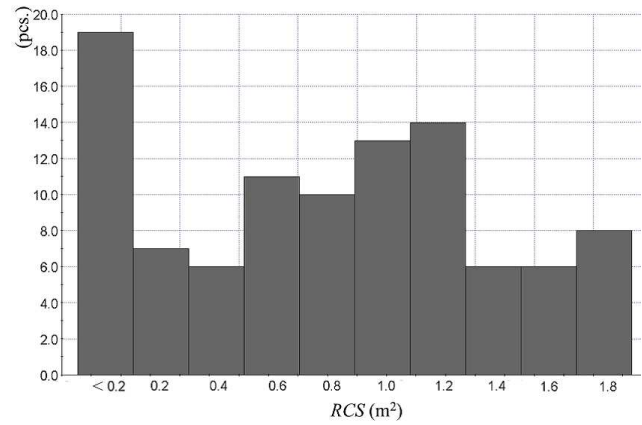


Figure 8: Frequency distribution of the RCS for 100 pieces of debris that were candidates for removal.

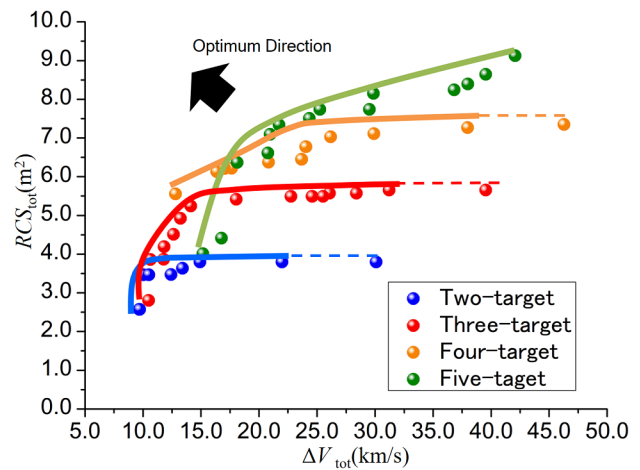


Figure 9: Non-dominated solutions obtained by the developed trajectory optimization method.

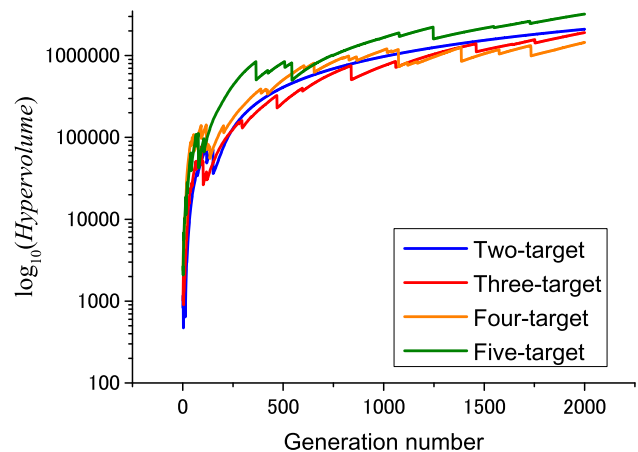


Figure 10: Comparison of hypervolume histories.

5.1 MoP Solutions by Means of MOEA

Figure 9 shows the solution after the 2000th generation. In this figure, the solutions of Rank 1 and 2 were plotted. From Fig. 7, the maximum RCS of one debris is $1.8m^2$; thus, the maximum value of RCS_{tot} in cases of removing two, three, four, and five pieces of debris is $3.6m^2$, $5.4m^2$, $7.2m^2$, and $9.0m^2$, respectively. In each case shown in Fig. 9, the resulting RCS_{tot} shows the maximum possible value. In the figure, the solid line connects Rank 1, and the dotted line connects Rank 2. According to these lines, ΔV_{tot} was distributed over a wide range while RCS_{tot} shows the same maximum values.

The hypervolume convergence history is shown in Fig. 10. In each case, converged solutions were obtained, while it is expected to moderately improve after 2000 generations. This is because the number of solutions constituting the non-dominated solution will increase in each case, while the maximum point of RCS_{tot} has already been determined.

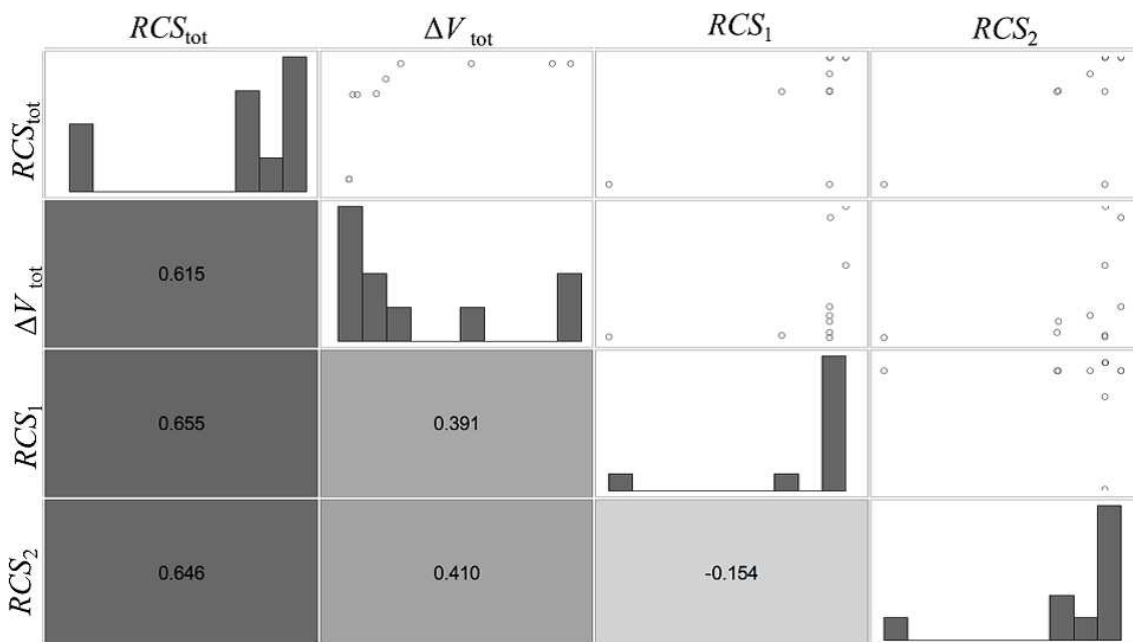
5.2 Visualization of the Design Problem with SPM

Figure 11 shows the SPM for each case. The SPM contains two objective functions and the RCS of the selected debris in the rendezvous order. In Fig. 11(a), both RCS whose debris to be removed show high correlation with both the objective functions. In Fig. 11(b) and (c), the RCS of the last debris to be removed (RCS_2 And RCS_4) is also correlated with both the objective functions. On the other hand, in Fig. 11(d), the RCS (RCS_3) of the third debris to be removed is correlated with each objective function; however, the RCS of the last debris to be removed is not correlated with each objective function.

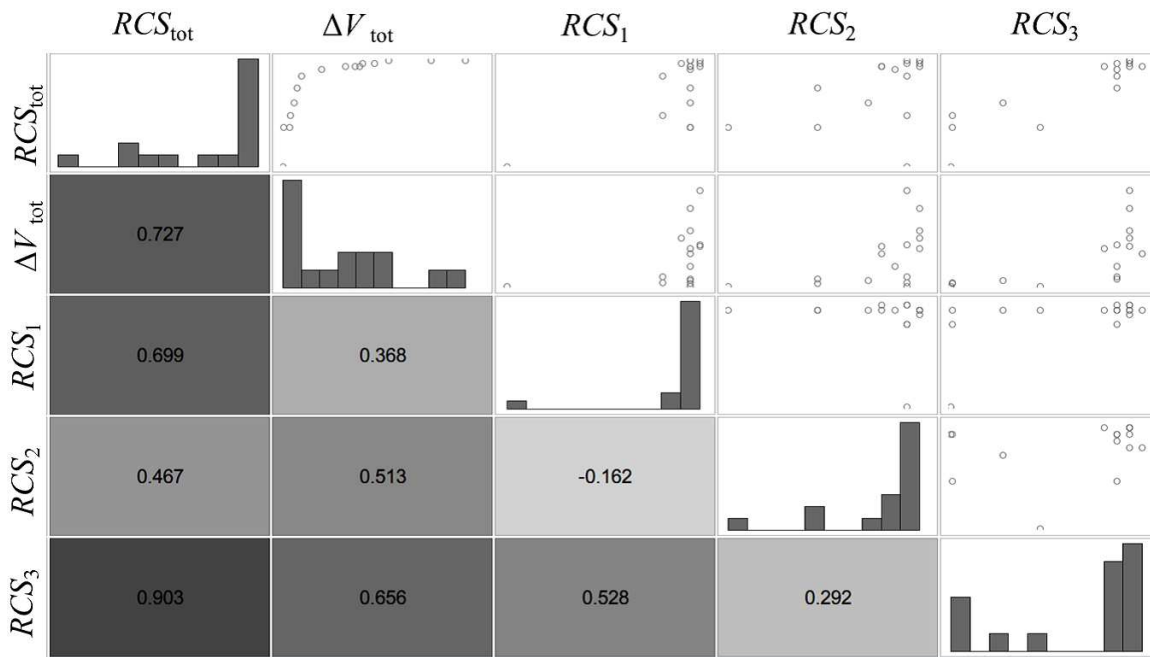
The diagonal elements in Fig. 11 show frequency distributions. Comparing these distributions with those in Fig. 7 shows the deviation. Looking at the frequency distribution of RCS_3 in Fig. 11(d), it was found that relatively large debris (RCS over $1.4m^2$) was removed. Furthermore, as shown in Fig. 11(a) and (b), large debris were selected for RCS_1 and RCS_2 , respectively. However, in Fig. 11(c) and (d), it was found that small debris $0.4m^2$ or less were also removed. This means that large debris should be removed along with small debris at a convenient position from the point of orbit transition to improve the overall efficiency. Figure 11(d) shows that in the case of removing five pieces of debris, there is a strong correlation between RCS_{tot} and RCS_3 . In fact, when RCS_3 is large, RCS_{tot} is also large. On the other hand, when RCS_3 is large, ΔV_{tot} does not necessarily become large. In this way, it is expected that planning multiple debris removal paths will become easier if the order in which to remove larger debris can be decided.

6 Conclusions

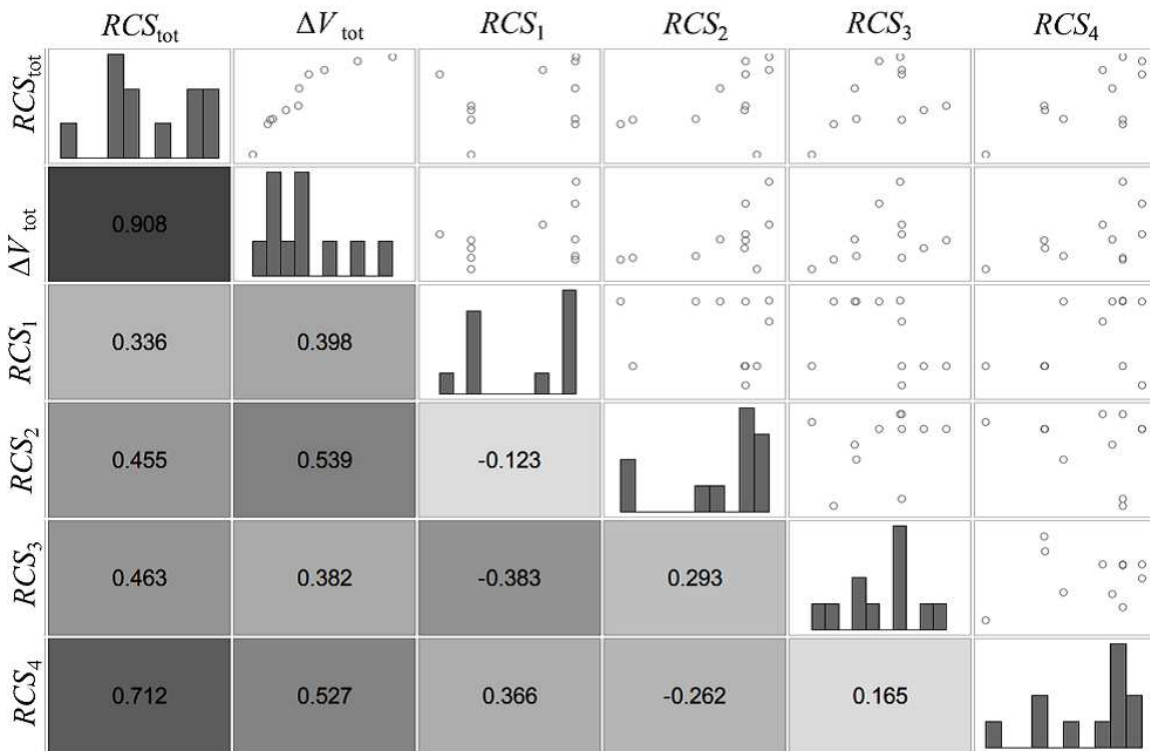
In this research, the path optimization of multiple space debris removal satellites was attempted by applying the TSP solution by using an evolutionary calculation method. The optimization problem considered was the maximization of the sum total of the radar reflection area that represents the size of the debris and minimization of the total velocity increment to rendezvous with the selected debris. Data, including altitude, for hundred pieces of debris were used to investigate our methodology. The Lambert problem was solved for the orbit transition of the removal satellite. In the optimization, order crossover was used as a crossover method. The reverse order method was



(a)



(b)



(c)

used for mutation. Furthermore, as the problem in this study is defined as multi-objective, non-dominated sorting for ranking was employed. In investigating the developed method, multi-objective problems were used by changing the pieces of debris removed by one removal satellite from 2 to 5. According to results, the following information was obtained.

- Because the sum of the maximum radar reflection areas can be calculated from 100 debris candidates in the proposed optimization process, it

is effective to use TSP solution method in this problem.

- There is a trade-off between the sum of the sizes of debris to be removed and the total velocity increment of the removal satellite.
- Even when the sum of the sizes of debris to be removed is the maximum, the total velocity increment was not uniquely determined. Thus, the total velocity increment should be set as the objective function.

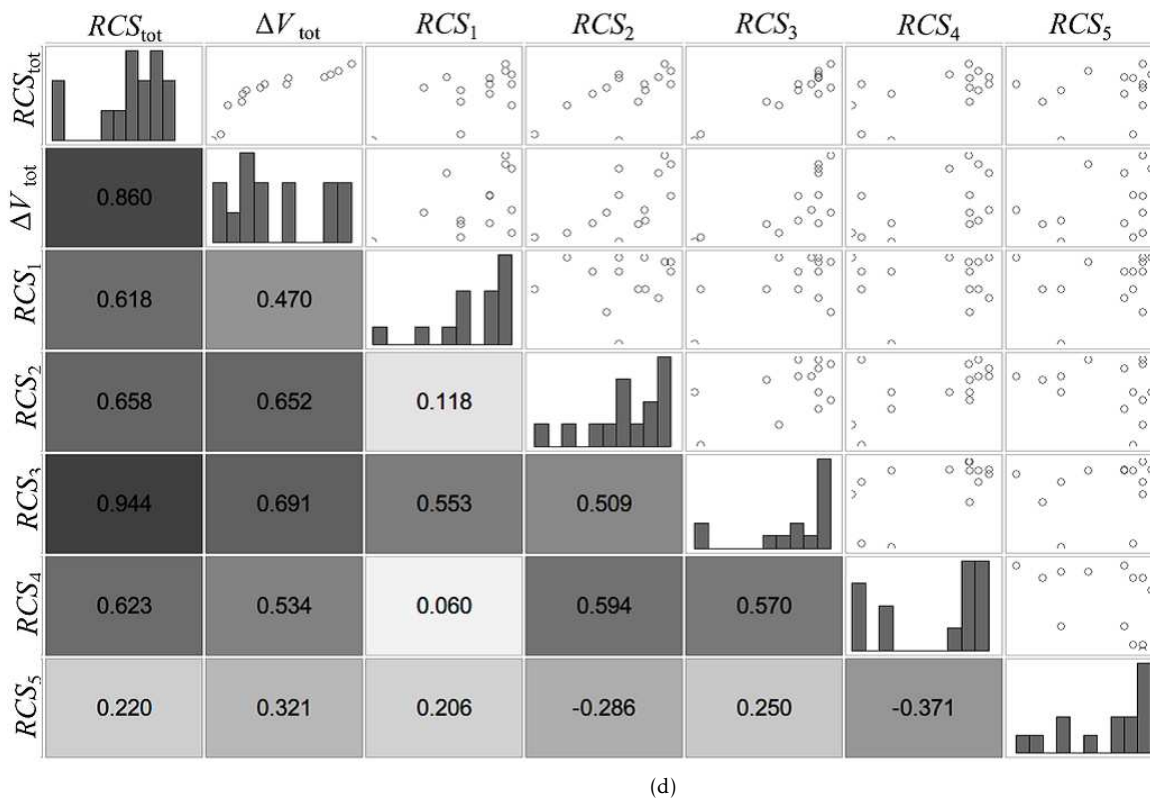


Figure 11: Design problem visualization by SPM: removal of (a) two, (b) three, (c) four, and (d) five pieces of debris.

- By increasing the pieces of removed debris simultaneously, the removal operation can be effective because smaller debris will be removed along with larger debris.
- In the case where five pieces of debris were removed, a positive correlation was observed between the radar reflection area of the third piece of debris and the sum of the radar reflection areas. Such findings can be considered as useful knowledge for mission planning.

References

[1] N. N. Smirnov, *Space Debris: Hazard Evaluation and Debris*, ESI Book Series, Taylor and Francis Group, 2002.

[2] H. Klinkrad, P. Beltrami, S. Hauptmann, C. Martin, H. Sdunnus, H. Stokes, R. Walker, and J. Wilkinson, "The ESA Space Debris Mitigation Handbook 2002," *Advances in Space Research*, Elsevier, **34**(5), 1251–1259, 2004. <https://doi.org/10.1016/j.asr.2003.01.018>

[3] T. Yasaka, R. Hanada, and H. Hirayama, "Geo debris environment: A model to forecast the next 100 years," *Advances in Space Research*, Elsevier, **23**(1), 191–199, 1999. [https://doi.org/10.1016/S0273-1177\(99\)00004-6](https://doi.org/10.1016/S0273-1177(99)00004-6)

[4] D. J. Kessler and B. G. CourPalais, "Collision frequency of artificial satellites: The creation of a debris belt," *Journal of Geophysical Research*, **83**(A6), 2637–2646, 1978. <https://doi.org/10.1029/JA083iA06p02637>

[5] Kessler, D. J., and Cour-Palais, B. G., : Collision frequency of artificial satellites: The creation of a debris belt, *Advances in Space Research*, Elsevier, pp. 2637–2646 (1978)

[6] Johnson, N. L. Stansbery E., Liou, J. C., Horstman, M., Stokely, C. and Whitlock, D., "The characteristics and consequences of the break-up of the Fengyun-1C spacecraft," *Acta Astronautica*, Elsevier, **63**, 149–156, 2013. <https://doi.org/10.1016/j.actaastro.2007.12.044>

[7] V. Aslanov and V. Yudinsev, "Dynamics of large space debris removal using tethered space tug," *Acta Astronautica*, Elsevier, **91**, 149–156, 2013. <https://doi.org/10.1016/j.actaastro.2013.05.020>

[8] Bombardelli, C. and Pelaez, J., : Ion Beam Shepherd for Contactless Space Debris Removal, *Advances in Space Research*, Elsevier, Vol. 34, No. 3, pp. 916–920 (2011)

[9] L. T. DeLucaa, F. Bernellia, F. Maggia, P. Tadinia, C. Pardinib, L. Anselmob, M. Grassic, D. Pavarind, A. Francesconid, F. Branzd, S. Chiesae, N. Viola, C. Bonnalf, V. Trushlyakovg, and I. Belokonov, "Active space debris removal by a hybrid propulsion module," *Acta Astronautica*, Elsevier, **91**, 20–33, 2013. <https://doi.org/10.1016/j.actaastro.2013.04.025>

[10] Payne, T. and Morris, R., : The Space Surveillance Network (SSN) and Orbital Debris, *33rd Annual AAS Guidance and Control Conference*, AAS Paper Number 10-012, (2010)

[11] Barbee, B. W., Alfano, S., Pinon, E., Gold, K. and Gaylor, D., : Design of Spacecraft Missions to Remove Multiple Orbital Debris Objects, *35rd Annual AAS Guidance and Control Conference*, AAS Paper Number 12-017, (2012)

[12] D. B. Fogel, "An evolutionary approach to the travelling salesman problem," *Biological Cybernetics*, Springer, **60**(2), 139–144, 1988. <https://doi.org/10.1007/BF00202901>

[13] K. Kondo, K. Miki, and T. Hiroyasu, "Kyoto touring problem - proposal of a new discrete test problem in multi-objective optimization -," in *IPJS Symposium*, Tokyo, Japan, 2003. <http://jglobal.jst.go.jp/en/public/200902275803303852> (in japanese)

[14] J. Xiao, X. Ma, Z. Zhu, J. Zhou, and Y. Yang, "Multi-objective memetic algorithm for solving pickup and delivery problem with dynamic customer requests and traffic information," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, Vancouver, BC, Canada, 2016. <https://doi.org/10.1109/CEC.2016.7744028>

- [15] D. D. Mueller, R. R. Bate, and J. E. White, *Fundamentals of Astrodynamics*, New Dover Publications, 1971. <https://doi.org/10.1109/4235.996017>
- [16] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197, 2002.
- [17] R. A. Becker, W. S. Cleveland, and A. R. Wilks, "Dynamic graphics for data analysis," *Statistical Science, Institute of Mathematical Statistics*, 2(4), 355–383, 1987. <https://www.jstor.org/stable/2245523>

Perfect Molding Challenges and The Limitations “A Case Study”

Tan Lay Tatt^{*1}, Lim Boon Huat¹, Rosli Muhammad Tarmizi², T. Joseph Sahaya Anand³

¹*Infineon Technologies (Malaysia) Sdn Bhd, Engineering Department, 75350, Malaysia*

²*Infineon Technologies (Malaysia) Sdn Bhd, Engineering Department, 75350, Malaysia*

³*Universiti Teknikal Malaysia Melaka (UTeM), Faculty of Manufacturing Engineering, 76100, Malaysia*

ARTICLE INFO

Article history:

Received: 12 September, 2018

Accepted: 16 December, 2018

Online: 23 December, 2018

Keywords:

Mold Void

Temperature Control

Pressure Control

Transfer Mold

ABSTRACT

Driven by today's market demand, semiconductor is pushing towards the zero-defect direction. The improvement demanded in semiconductor manufacturing is becoming increasingly challenging. In this paper, common molding defects comprise of voids, incomplete fills, and piping holes are studied systematically, focusing on three key areas: 1) Potential mold flow weakness; 2) Molding temperature stability; as well as 3) Defined pressure effects. The in-depth understanding of mold flow in the LF design is achieved via mold flow numerical tool. The numerical model prediction is verified by short shots and end-of-line auto vision data. Advance Process Control (APC) is adopted to measure the stability of key molding parameters like temperature, transfer profile and pressure. The mechanism of transferring the compound in relation to pressure is also analyzed and its effect to molding quality is also assessed. A methodical approach is utilized to understand the process and equipment built-in capabilities from two different equipment manufacturers. The real time transfer profile monitoring is activated for diagnosis of the system issue which leads to the finding of design error of a critical component. The dual temperature controller on one of the systems is analyzed to stabilize temperature for improved compound flow-viscosity control. The process limitations are assessed and transfer profiles are optimized to modify the melt front. By shifting the molding defects to non-critical location, the formation of void at the 500um diameter bonded wire loop peak will be avoided. The verification of potential negative impacts resulted from changes to improve voids, incomplete fills and piping holes are also included in this study. Up-front analysis by adopting numerical tool as a means of understanding the existing design and identifying improvement approach are proven to be useful.

1. Introduction

As today's market requirement on semiconductor is moving towards the direction of zero defect. This demanding requirement has led to improvement not only at manufacturing process control but also at the mold design level without additional cost. The molding process is being chosen due to its high through put per cycle of production. Any imperfection of this process will lead to high quantity of rejections. Ishikawa diagram, Pareto analysis, and Taguchi design of experiments tools are often adopted to address the molding related issues [1,2] but lack of online feedback data to ensure sustainable quality becomes questionable.

^{*}Corresponding Author: Tan Lay Tatt, Infineon Technologies (Malaysia) Sdn Bhd, Contact No: +6062325266 & Email: laytatt.tan@infineon.com

A computer-aided mold design tool with mold shot library built-in to reduce lead time and cost of mold tool design and fabrication has been formulated [3]. This developed methodology does not take into consideration the interaction effects of mold compound, process parameters, and design. In recent years, vacuum assisted molding has been accepted as a potential solution to reduce void formation [4,5]. However, this feature will incur higher running cost. The surge of introduction of Cu wire to replace Au wire due to high Au price [6] has led to new sets of bonding metallurgy and process issues. The effects are pre-mature failures and higher production yield losses [7-11]. Adversely, Cu wire brings along molding advantage of resistant to wire sweep due to its higher modulus. Thicker diameter wire can also provide high resistance to wire sweep during filling of mold cavities [12]. But, it can also

influence the compound flow which can trap the air and cause mold voids.

In this paper, the potential pitfall of molding due to compound flow, temperature stability, pressure control, static DOE results, and interaction effects are being investigated using numerical tool, advance process data control, and in-depth understanding of equipment design.

2. Potential mold flow weakness

The design of package had played the main role that can influence the mold flow pattern [12]. Design of package included die size, wire size, number of wire, wire bonding pattern...etc. Figure 1 showed the DOE matrix for simulation, by using the Autodesk Moldflow.

DOE Matrix	Device #1	Device #2	Device #3
Die Size	Thin	Medium	Thick
Quantity of thin wire *same wire size	1	1	1
Quantity of thick wire *same wire size	3	4	4

Figure 1. Experiment Matrix on different bonding Configuration

Simulation tool with time lapse comparison shows that, the completion of encapsulation process gives faster for bigger die size or higher wire density package as shown in figure 2.

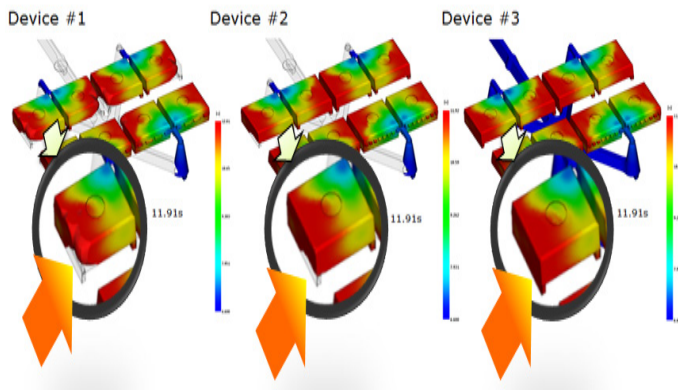


Figure 2. Experiment result on different bonding configuration

This had explained that with constant control on material (compound types), process method and process parameter (cavity temperature, transfer pressure, and transfer profile), package with bigger die size or higher wire density will experience the continuous transfer pressure from plunger after complete fill up the cavity [13]. Higher compactness on package occurs once the compound has been compressed with the set transfer pressure.

3. Temperature Control Sensitivity

Temperature control is one of dominant factors that can greatly affect molding quality [14]. Figure 3 showed the failure mode of mold void detected during production. The magnitude of mold void losses are significantly different by equipment types. “Equipment A” with pin gate concept shows much higher losses

as compared to “Equipment F” which is standard mold concept as shown in Figure 4.

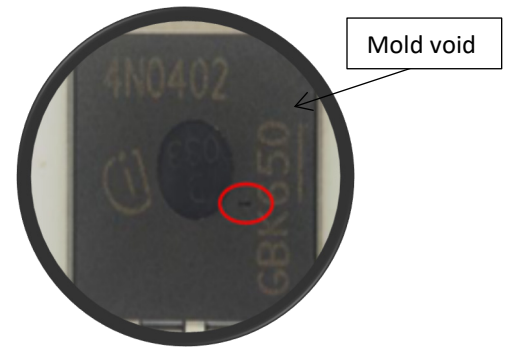


Figure 3. Mold void defect mode

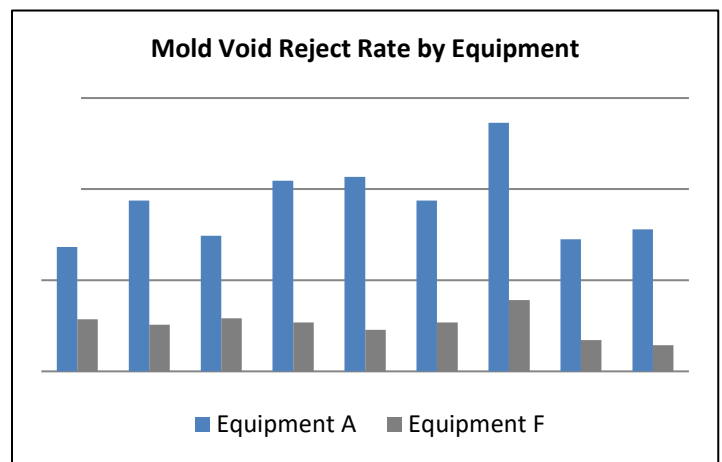


Figure 4. Mold void trend by equipment

Investigation shows that the temperature monitoring trend for “Equipment A” is instable compared to “Equipment F” as shown in Figure 5.

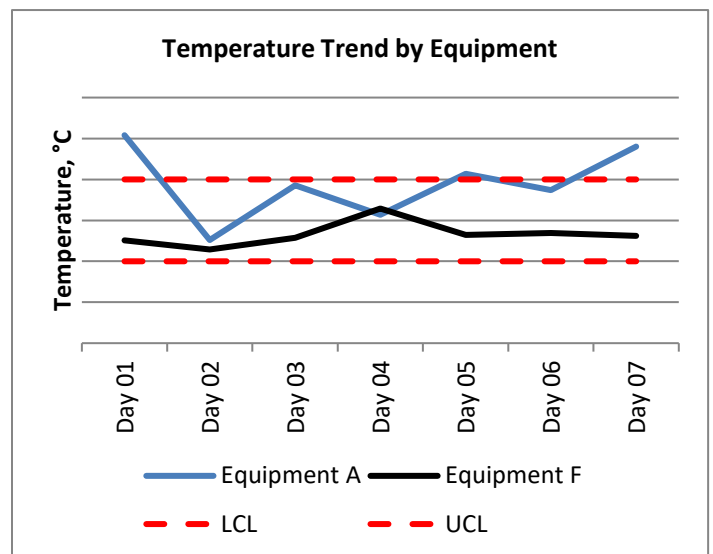


Figure 5. Temperature monitoring Trend

To understand the behaviour of the temperature fluctuations, three areas of focus are defined:

- 1) Mold design;
- 2) Heater system; and
- 3) Robustness of Thermocouple wires.

Mold designs of these two types of equipment are assessed. Figure 6 showed the standard molding concept. The mold runner flow across the rows through the leadframe. The runner which is removed later at the degate station together with the cull. For the standard mold design, mold cavity temperature is stable and sustainable as top cavity and bottom cavity are same temperature setting on solid piece of metal.

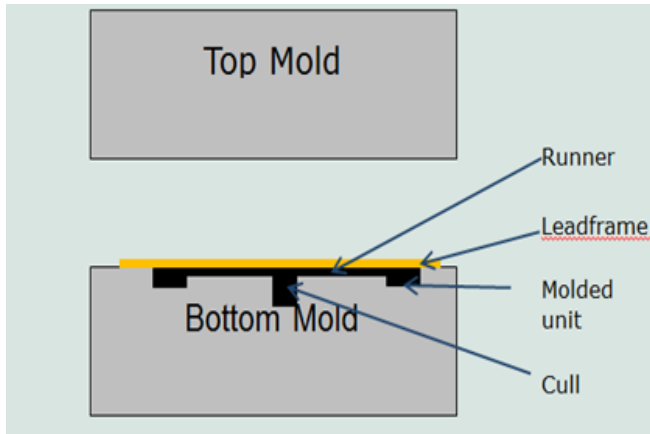


Figure 6. Standard mold design

Figure 7 shows pin gate molding concept. The mold runner flow through a middle plate, middle plate is a moving part whereby it will de-cull the runner and cull during mold open. Due to the middle plate which is moving part and no heater element attached on it, the temperature is always unstable with +/- 5 degree fluctuation caused by heat losses during mold process cycle.

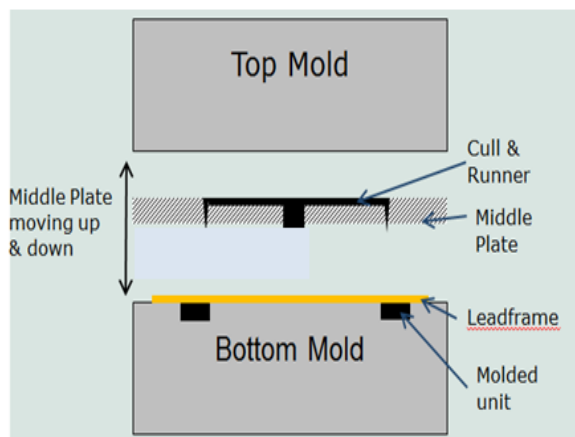


Figure 7. Pin gate mold design

Further investigation found that two individual heater controls are designed into this molding system one on mold housing and the other on mold chase. This leads to complexity of temperature synchronization of two controls: heating and cooling ones [16]. Figure 8 showed the schematic illustration of the heating and control of pin gate molding system. The heater locations and thermocouples and their number can cause the response time to delay for the heaters to turn on or off. All these factors have caused the fluctuation of the temperature of

equipment A to wider ranger. Within a molding system, there are four presses. Figure 9 showed temperature variations within a press and between press-to-press.

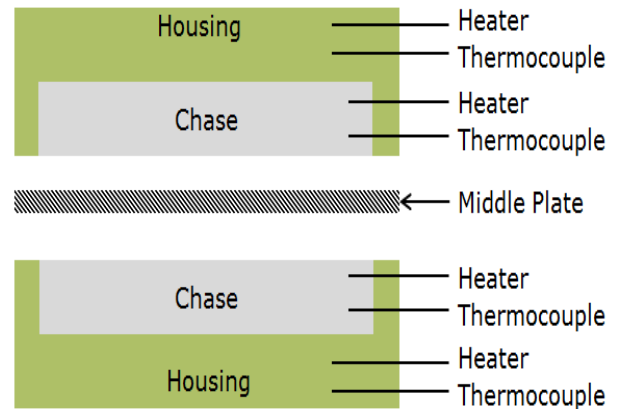


Figure 8. Two individual heaters in one mold press

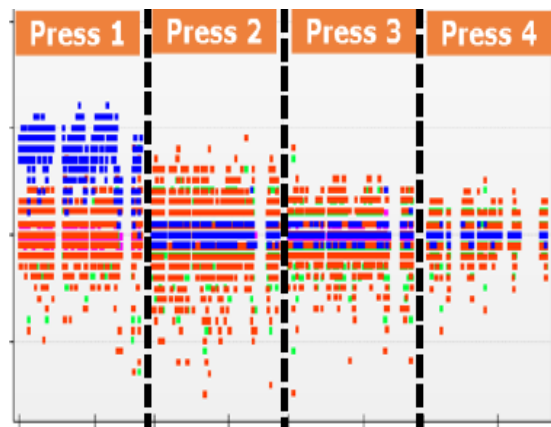


Figure 9. Recorded online temperature variations for four molding presses

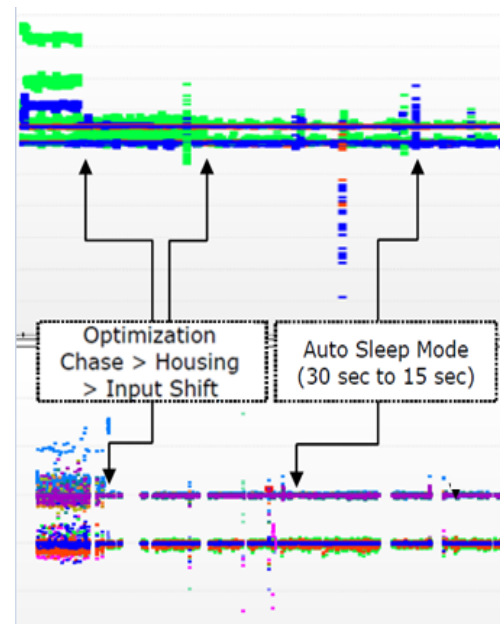


Figure 10. Mold temperature monitoring results after actions taken

After empirical data collection, analysis of the data, and understanding the molding hardware, two key actions are

formulated to overcome the design limitations: a) Single housing heating by disable chase heater to eliminate the temperature control synchronization, and b) Minimizing the waiting time for auto sleep mode to minimize the heat losses at middle plate and cavity. With the two key actions taken, mold temperature has been stabilized as shown in Figure 10.

Figure 10 also showed that sporadically mold temperature can fluctuate on a wide range. Thermocouple wire connector design is the third weakness observed, which can cause unstable temperature. The existing design with thin thermocouple wire attached to TC connector which is easy to spoil if there are any mishandling during install. An enhanced connector design which is more solid and user friendly being introduced into the new molding system. The robust and user friendly connector will not be only replaced by the existing one used in the production floor, but new machine will be introduced with this new connector.

4. Transfer Pressure

In addition to molding temperature, transfer pressure is another key controllable process parameter which can greatly influence the molding quality. In this paper, molding equipment F is used to study the effect of transfer pressure due to its equipment with online real-time pressure monitoring. In this equipment platform, transfer pressure in is monitored under a system called Final Transfer Pressure Control (FTPC). It measures the applied pressure and stops the transfer movement when the required transfer pressure has been reached.

4.1. Process Monitoring Fundamentals

Figure 11 shows a transfer pressure graph in which the transfer position is against the transfer pressure.

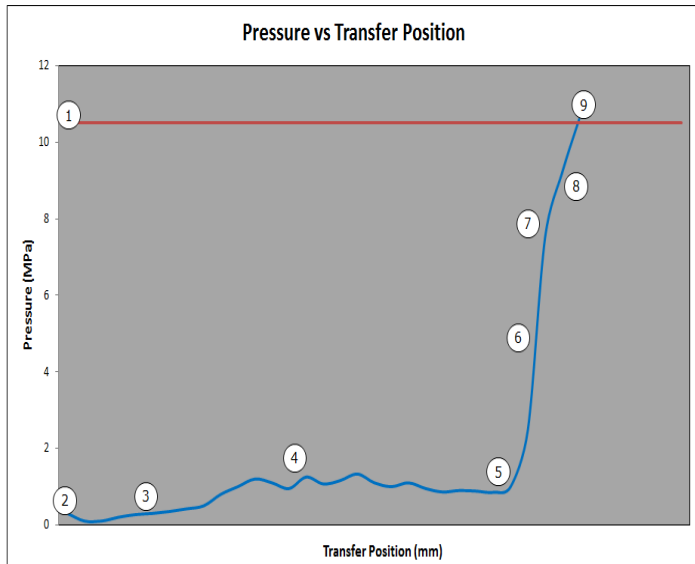


Figure 11. Transfer pressure graph

A description of the event taking place at each number in figure 11 is given below:

1. Red line is transfer pressure set point (i.e. 10.5 MPa).
2. After compound pellets are loaded into the press and they were heat up while the mold closes, the transfer starts moving from its start position which is called pellet load position.

3. There will be a light pressure built up because of the resistance of the plungers in the sleeves and partly melted pellet is pushed up in the sleeve.
4. The compound melts and the transfer moves further up. During the transfer movement, the pressure slightly increases because fluid compound goes through the runners and the narrow gates, which are expected having some resistance.
5. During this transfer process stage, all compound is melted and the pressure built up is caused by the resistance of the compound in the runners, at the gate and in the package.
6. Now, the cavities are completely filled, the air is pushed out through the air vents and the pressure starts to build up in the package. The position from which the line starts rising rapidly equals the cull height.
7. During the fast increase of the pressure, the plunger springs are compressed beyond the spring preload pressure.
8. From this point, the transfer starts compressing the spring.
9. At the final stage, the plunger springs are compressed. The slope of the line depends on the type of spring used in the mold. The transfer move stops when final transfer pressure (required packing pressure) reaches the set point, mold process is completed.

4.2. Transfer Pressure Mechanical System

The transfer pressure mechanical system is dictated by two critical components: 1) Plunger beam and 2) plungers as shown in Figure 12.

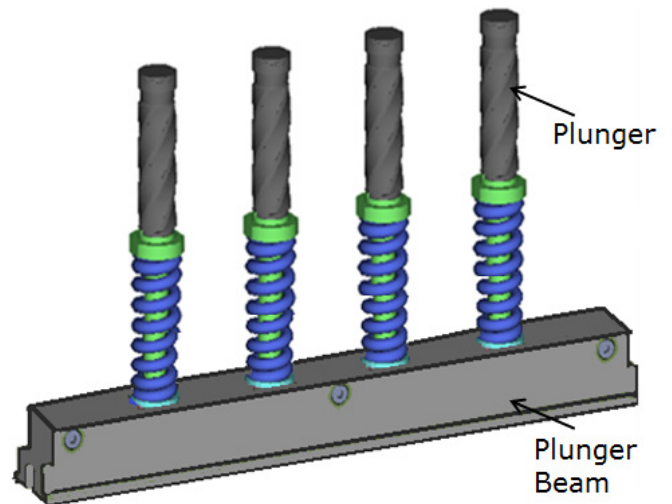


Figure 12. Plunger and plunger beam for molding compound transfer

4.3. Plunger Beam and Plunger

Figure 13 showed the cross section view of an assembled plunger system. The plungers are fitted under a pre-defined tension by means of the compression spring (C). Different spring diameters (recognizable by color type) are used, depending on the required spring force and transfer pressure. The higher part of the plunger (F) is screwed to the lower plunger part (shoulder pin (E) and spring (C)) by means of a long socket screw (D).

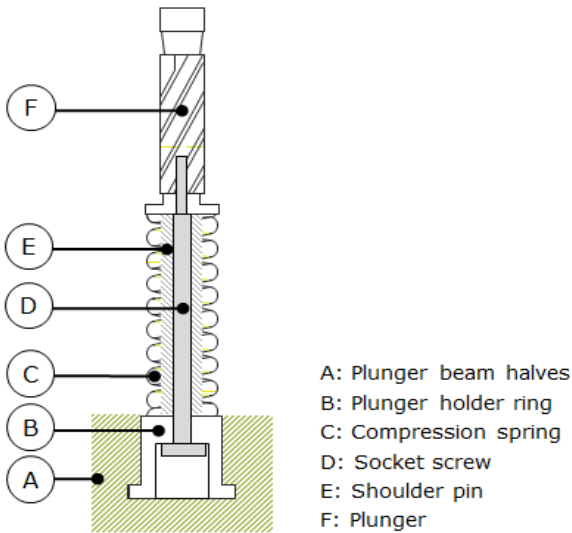


Figure 13. Cross sectional view of plunger beam and plunger

4.4. Transfer movement

The transfer movement consists of four critical pre-set positions. The first position is the starting position which is in compound pellet load position. The second position is the pellet crush position. At this position, the plunger beam waits until the pellet preheat time has finished. The third position is the “Cull Height” (CH) position. Generally, at this point, the plunger cannot move up higher and the cavities are completely filled. But, the transfer movement is not yet applying pressure to epoxy. The end position is the plunger move up to the CH position but the plungers themselves cannot go higher because the cull is blocking this. However, the transfer beam upward movement will cause the cull to compress further by the flexible spring. Figure 14 showed the four critical positions of plunger beam and plunger positions.

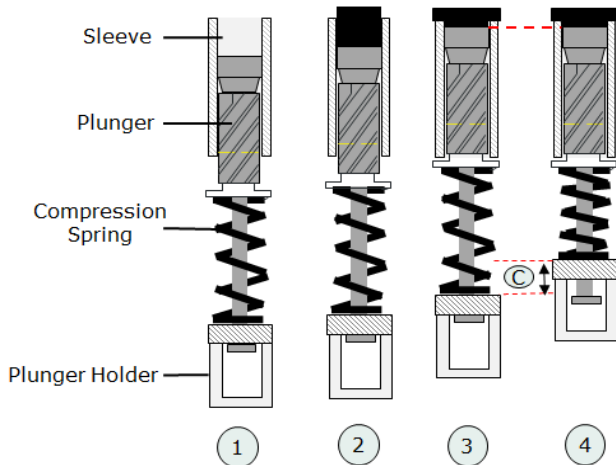


Figure 14. Four critical plunger beam and plunger positions

Cull height setting is one of the crucial parameters that can influence the compactness of the package. However, when the cull height parameter is set lower than the actual cull height, this may deform the product with deform dambar. Therefore, the set end position must be always lower than the cull height. If not, the maximum transfer pressure cannot be attained.

4.5. Transfer Pressure Effects

Transfer end pressure can go higher through the setting of higher plunger position. Different spring type used can deliver different transfer end pressures which purely depend on the cavity designs. When the transfer pressure is too low, this may cause an incomplete fill and air in the package. When the pressure is too high, this may cause die damage or compound leakage. For this paper a case study of plunger beam movement effect on final packing is examined. Figure 15 showed two molding defects due to insufficient transfer pressure.

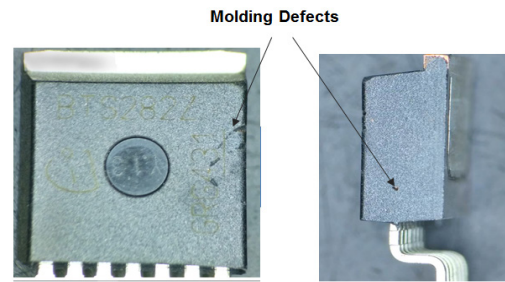


Figure 15. Two modes of molding defect caused by insufficient transfer pressure

The transfer movements and settings in section 4.4 are used as the base for the investigation. The frequent FTPC error triggering at the end filling stage has led to the finding; at the end position the plunger beam upward movement has obstruction. This incomplete final movement had prevented the final compactness of the compound to happen. The root cause of this incomplete plunger beam movement is resulted from fabrication error of plunger holder as shown in Figure 16 below.

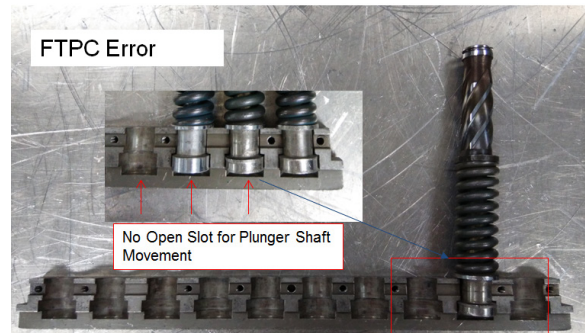


Figure 16. Plunger beam with un-through slot for plunger shaft movement

A new plunger holder with through slot was fabricated and installed into the molding machine as in Figure 17.

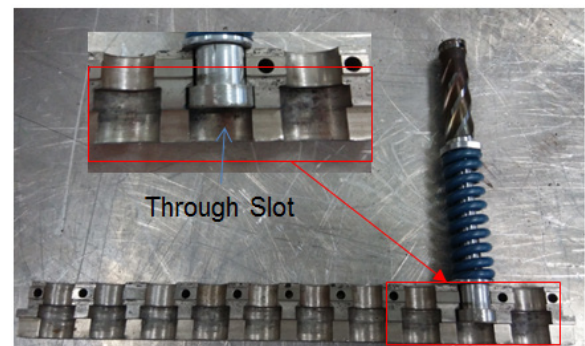


Figure 17. A plunger holder beam with through slot to facilitate free plunger movement

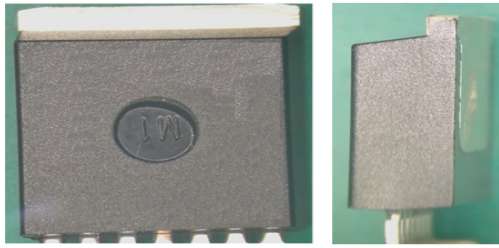


Figure 18. Improvement of molding defects after the corrective action implementation

4.6. Transfer Pressure and Transfer Time Correlation

The effect of transfer time and transfer pressure on short mold and void formation during encapsulation is evaluated in this study. Both short mold and void formation during transfer molding process will affect the long term reliability [17-19] of the package being molded. The tool being used to visually quantify the defect of short mold and void formation is low power scope. Others primary molding process parameters such as top and bottom mold chase temperature, mold compound transfer time are fixed with nominal setting during the evaluation. To minimize the influence of the cleanliness of mold tool on the results, mold die cleaning and conditioning were performed prior to the test. The cleanliness is extended to all the air ventilation area, runner and pot area as well.

The transfer time is defined according to the recommended mold temperature from the melting viscosity curve for compound material A as shown in Figure 19. Output response on short molding and void percentage rate are computed and samples are inspected under low power scope by one dedicated inspector. Table 1 and 2 show the Design of Experiments matrix for transfer pressure and transfer time respectively.

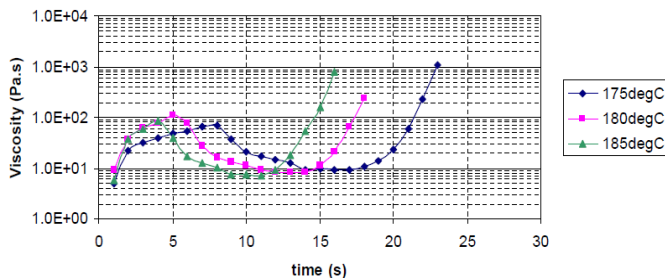


Figure 19. Melting Viscosity chart for compound material A

Table 1. Experiment Matrix of transfer Pressure

DOE Matrix	Mold Temperature	Transfer Time	Transfer Pressure	Remark
1	Fixed	Fixed	Low	Control
2	Fixed	Fixed	High	High Transfer Pressure

Table 2. Experiment Matrix of transfer time

DOE Matrix	Mold Temperature	Transfer Time	Transfer Pressure	Remark
3	Fixed	Medium	Fixed	Control
4	Fixed	Low	Fixed	Shorter transfer time
5	Fixed	High	Fixed	Longer transfer time

High density manufacturing platform is utilized to carry the evaluation. DOE matrix 1 and 2 from Figure 21 show the short molding and void defect inspection results from high density lead frame. The DOE samples are molded with 2 different transfer pressure, low pressure and high pressure. The remaining others primary parameters are all set at nominal values. There is no significant improvement being observed in term of failure rate when transfer pressure increased from low to high with the fixed transfer time [16]. This confirmed that the air trap in the cavity is not due to insufficient compactness during the material transfer to cavity. To gain further understanding, short-shot analysis and mold flow simulation are carried out to further validate the void formation mechanisms and the final air trapped location.

Figure 20 shows the result of shot-shot simulation. It shows the similar behaviour where the air is trapped at end mold flow position and leads to short mold or void formation, which is matching with the current molding defect and location.

DOE matrix 3, 4 and 5 from Figure 21 show the short molding and void inspection result. No significant difference was observed from the matrix 3,4 and 5. According to the viscosity curve, control parameter is the optimum point to achieve the most “liquefy” mold compound flow during the transfer of compound to cavity. The optimum processing parameters, transfer mold is locked at existing parameter for further DOE.

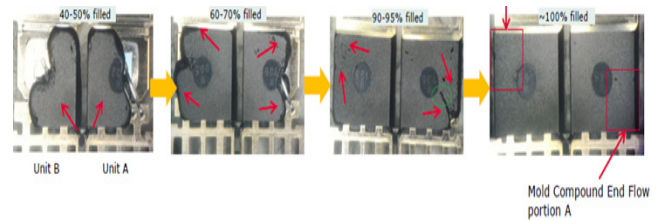


Figure 20. Short-Shot simulation result

DOE Matrix	Mold Temperature	Transfer Time	Transfer Pressure	Result
1	Fixed	Fixed	Low	2F/800
2	Fixed	Fixed	High	2F/800
3	Fixed	Medium	Fixed	2F/800
4	Fixed	Low	Fixed	1F/800
5	Fixed	High	Fixed	2F/800

Figure 21. DOE Matrix result for transfer pressure and transfer time

5. Conclusion

In order to improve the molding processing and reduce the rejection by more than 10%, the following three areas have to be focused: 1) enable only single temperature control instead of dual temperature control to stabilize the temperature. 2) To overcome the challenge of molding process, consideration of internal product structure’s design such as wire size, die size, number of wires, and wire orientation which can cause unwanted mold defects. In addition, up-front analysis by adopting numerical tool to understand the existing design and improvement options are proven to be useful. 3) Fix the fabrication error of the plunger holder to overcome the incomplete plunger beam movement which resulted insufficient transfer pressure. On top of that, online process monitoring such as Advance Process Control (APC) is not

only useful to control the process but also a good tool to diagnose the design weakness.

References

- [1] K.A.Z. Abidin, K.C. Lee, I. Ibrahim and A. Zianudin, "Problem Analysis at A Semiconductor Company: A Case Study on IC Packages", *Journal of Applied Sciences*, 11 (2011) 1937-1944.
- [2] J. Antony, "Taguchi or classical design of experiments: A Perspective from practitioner" *Sensor Rev.*, 26 (2006) 227-230
- [3] M.R. Alam, M.A. Amin, and M.A. Karim, "A Computer-aided Mold Design for Transfer Molding Process in Semiconductor Packaging Industry" *ScienceDirect Procedia Manufacturing* 21 (2018) 733-740
- [4] T.S. Lundstrom and B.R. Gebart. "Influence from process parameters on void formation in resin transfer molding", *Polymer Composite Polymer Composites*, 15 (1994) 25-33
- [5] V.R. Kedari, B.I. Farah, K.-T. Hsiao, "Effects of vacuum pressure, inlet pressure, and mold temperature on the void content, volume fraction of polyester/e-glass fiber composites manufactured with VARTM process", *Journal of Composite Materials*, 45 (2011), 26, 2011 2727-2742
- [6] T.J.S. Anand, K.Y. Chua, Y.S. Leong, W.K. Lim and M.T. Hng, "Microstructural and Mechanical analysis of Cu and Au interconnect on various bond pads" *Current Applied Physics* 13 (2013) 1674 – 1683.
- [7] M.S. Patel, C.P. McCluskey, and M. Pecht, "Effective decapsulation of copper wire-bonded microelectronic devices for reliability assessment, *Microelectronics Reliability* 84 (2018) 197-207.
- [8] M.S. Krishnanramaswami, G. McCluskey, M. Pecht, "Failure mechanisms in encapsulated copper wire-bonded devices", *IEEE 23rd International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2016.
- [9] V.B. Willems, "Early fatigue failures in copper wire bonds inside packages with low CTE green mold compounds", 2012 4th Electronic System Integration Technology Conference, 2012, <http://dx.doi.org/10.1109/estc.2012.6542110>.
- [10] D. Andrews, L. Hill, A. Collins, K.I. Hoo and S. Hunter, "Copper ball bond shear test for two pad aluminum thicknesses, 2014 IEEE 16th Electronics Packaging Technology Conference (EPTC), 2014, <http://dx.doi.org/10.1109/eptc.2014.7028415>.
- [11] S. Manoharan, S. Hunter and P. McCluskey, "Bond pad effects on the shear strength of copper wire bonds, *Electronics Packaging Technology Conference (EPTC)*, 2017 IEEE 19th, IEEE, 2017, December, pp. 1–6.
- [12] L.T. Tan, C.H. Lee, Y.Y. Teo and B.H. Lim, "Perfect Molding Challenges and The Limitations", *IEEE 19th Electronics Packaging Technology Conference (EPTC)*.
- [13] H. Ardebili and M. Pecht, "Encapsulation Technologies for Electronic Applications", William Andrew, Oxford 2009.
- [14] E. Ridengaoqier, R. Fujiki, S. Hatanaka and N. Mishima, "Study on estimation of void ratio of porous concrete using ultrasonic wave velocity", *Journal of Structural and Construction Engineering*, 83 (2018), 943-951.
- [15] G.M.C. Magalhães, G. Lorenzini, M.G. Nardi, S.C. Amico, L.A. Isoldi, L.A.O. Rocha, J.A. Souza and E.D. Dos Santos, "Geometrical evaluation of a resin infusion process by means of constructal design" *International Journal of Heat and Technology*, 34 (2016) S101-S108.
- [16] M. Kubouchi, H. Sembokuya, S. Yamamoto, K. Arai and H. Tsuda, "Decomposition of amine cured epoxy resin by nitric acid for recycling", *Journal of Society in Materials Science, Japan*. 49 (2000) 488–493.
- [17] N.M. Li, D. Das and M. Pecht, "Shelf life evaluation method for electronic and other components using a physics-of-failure (Pof) approach", *Machinery Failure Prevention Technology (MFPT) Conference, MFPT*, 2017.
- [18] N.M. Li and D. Das, "Critical review of US Military environmental stress screening (ESS) handbook", *Accelerated Stress Testing & Reliability Conference (ASTR)*, IEEE 2016, 1-10.
- [19] F.P. McCluskey, N.M. Li and E. Mengotti, "Eliminating infant mortality in metallized film capacitors by defect detection", *Microelectronics Reliability* 54 (2014) 1818–1822.

Omni-directional Dual-Band Patch Antenna for the LMDS and WiGig Wireless Applications

Mourad S. Ibrahim^{1,2,*}

¹Department of Communications and Networks Engineering, College of Engineering, Prince Sultan University, Riyadh, 11586, KSA

²College of Engineering, Modern Sciences and Arts University, 6th October City, Egypt.

ARTICLE INFO

Article history:

Received: 20 September, 2018

Accepted: 16 November, 2018

Online: 23 December, 2018

Keywords:

Dual band

Fifth generation

LMDS

Omnidirectional pattern

WiGig

ABSTRACT

In this paper an omnidirectional dual band monopole antenna at 28 GHz and 60 GHz which is fit for indoor and outdoor wireless applications is developed. The proposed antenna consists of two rectangular patches with a T folded patch. The design, analysis, and optimization processes through this article are executed by the numerical method, Finite Element Method (FEM) and verified with another numerical method, Finite integration Technique (FIT). Good agreement between the results by these two simulators is obtained. The proposed antenna has achieved dual bands with omnidirectional patterns. The first band at 28 GHz is extended from 27.5 GHz to 28.958 GHz with 5.1 % bandwidth and total efficiency of more than 93% along the entire band which serves the LMDS band. The second band at 60 GHz is extended from 45.2 GHz to 84.4 GHz which serves the WiGig band with bandwidth of 60.6% and total efficiency of 85.5% along the entire band. The proposed antenna performance makes it a good candidate for the fifth generation (5G) applications.

1. Introduction

This paper is an extension of work originally presented in ACES [1]. More analysis and parametric studies have been done. The average data per person uses in telecommunication is rapidly increases over the past three decades [2]. This increase is more noticeable mostly with the outgrowth of the wireless communications. The main requirements for wireless communications are the ability to develop a low-cost, light-weight, and low-profile antennas to maintain good performance along a large bandwidth [2].

The shortage in the available global bandwidth has stimulated the reconnaissance of the under-utilized millimeter wave frequency band for the future wireless communications [3].

Multiband antennas have massive applications in millimeter band applications. A various techniques in the literature have been developed for multiband microstrip antennas as in [4–12]. For instance, a multilayer GaAs is described in [4] to achieve a multiband antenna operating at 35 GHz. In [5], two bands with less than 1.2% bandwidth with gains of -9 dBi and 1 dBi at 24 GHz band and 60 GHz band respectively is presented. A dual band antenna at 41 GHz / 52.2 GHz using a meta-resonator with pair of

split ring resonators is introduced in [6] with bandwidth of 2 %, gain of 3.76 dBi, and efficiency of 71%. In [7], two different modes are obtained to get a dual band at 58 GHz and 77 GHz with gains of -2 dBi and 0.3 dBi, respectively. The achieved bandwidths at both bands are nearly 6%. A dual band centered at 24.5 and 35 GHz has been presented in [8] with only 1% bandwidth and less than 2.8 dBi gain using liquid crystal polymer. A coplanar hybrid dual band antenna at 83 GHz / 94 GHz with a slot in feeding line has been developed in [9]. In [10] an antenna array with Electromagnetic Band Gap [EBG] structure was used to develop a dual band at 28 / 38 GHz with bandwidth of less than 5.8%. A three layers of substrates have been used in [11] to design a Fabry-Perot cavity antenna operates at 36 GHz with high gain. In [12], L-shaped slots have been used to obtain a dual band 28 / 38 GHz slotted patch antenna.

In this paper, the design, optimization, and simulation of a monopole planar antenna are introduced. The antenna is optimized to operate at Ka - band (28 GHz) for Local Multipoint Distribution Service (LMDS) which currently investigated for the fifth generation mobile cellular [13], and the V - band (60 GHz) for Wireless Gigabit Alliance (WiGig) applications [14].

This paper is organized into four sections. Section 1 covers the introduction and literature review. Antenna structure and design is introduced in section 2. Parametric study of a dual band

*Mourad S. Ibrahim, Riyadh 11586, mrizk@psu.edu.sa

omnidirectional pattern circularly polarized wideband antenna is presented in section 3. The simulation results are investigated in Section 4. Finally, the conclusion is presented in section 5.

2. Antenna Geometry and Design

Figure 1 shows the antenna geometry in perspective and top view. The proposed antenna patch consists of a T-folded shape with two rectangular patches and partially grounded [15]. The antenna gives wide bandwidth with improved antenna performance. The patch mounted on substrate FR-4 with a relative permittivity of 4.4. Figure 2 illustrates the whole antenna dimensions. The substrate dimension is W_s by L_s and the partial ground is W_s by L_g . The dimensions of T-folded are L_{top} , L_{fold} , L_t , W_t for top length, folded side length, mid length, and mid width. The two patches have dimensions of W_p by L_p .

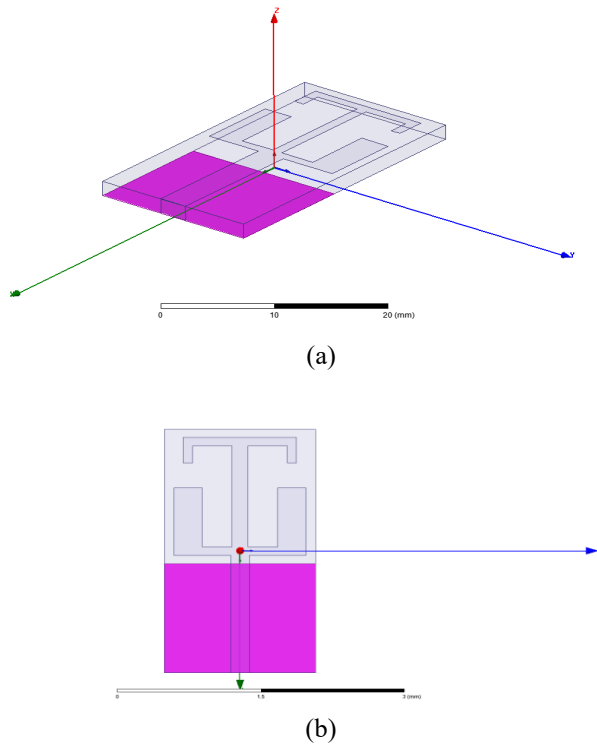


Figure1. Antenna geometry a) Perspective view b) Top view.

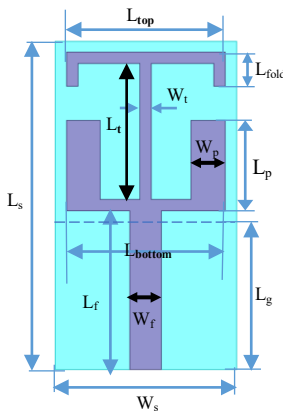


Figure 2. Antenna structure and dimensions.

The antenna feeding transmission line (TL) dimension is W_f by L_f . The TL characteristic impedance can be calculated by the following formulas [16]:

When $\frac{w}{h} \leq 1$, the characteristic impedance is

$$Z_c = \frac{60}{\sqrt{\epsilon_{reff}}} \ln \left[\frac{8h}{w} + \frac{w}{4h} \right] \quad (1)$$

where

$$\epsilon_{reff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \times \left\{ \left[1 + 12 \frac{h}{w} \right]^{-1/2} + 0.04 \left[1 - \frac{w}{h} \right]^2 \right\} \quad (2)$$

while when $\frac{w}{h} > 1$

$$Z_c = \frac{120\pi}{\sqrt{\epsilon_{reff}}} \frac{1}{\frac{w}{h} + 1.393 + 0.667 \ln \left[\frac{w}{h} + 1.444 \right]} \quad (3)$$

where

$$\epsilon_{reff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12 \frac{h}{w} \right]^{-1/2} \quad (4)$$

In the formulas, h represents the substrate height, w represents the TL width, and ϵ_{reff} represents the effective relative permittivity.

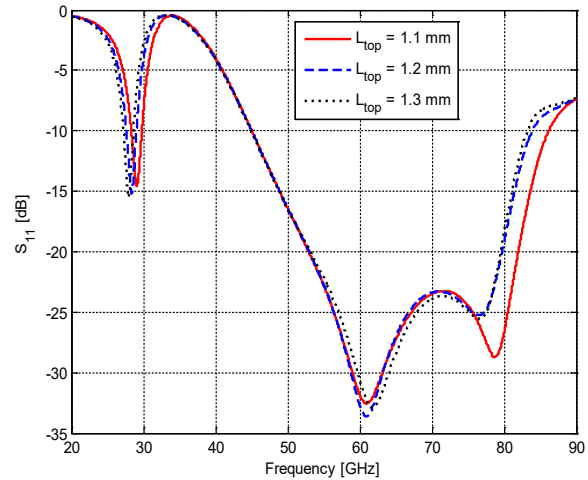


Figure 3. The effect of changing L_{top} on S_{11} .

3. Parametric Study

The effect of varying antenna dimensions L_{fold} , L_{top} , L_{bottom} , L_p , W_p , and L_t using HFSS are shown in figures 3 to 8. The dimensions of the substrate, ground, and feeder are kept unchanged. With increasing L_{top} of the horizontal part of T-shaped, the lower resonant frequency is decreased with same bandwidth whereas the upper bandwidth is decreased as shown in Figure 3. Figure 4 illustrates the effect of varying L_{bottom} on the return loss. As noted from the figure, by increasing the L_{bottom} , the lower resonant frequency almost unaltered whereas the resonant frequency of the

upper band is increased with a little increase in the bandwidth. With increasing W_p , the upper and the lower resonant frequencies are almost unchanged as shown in Figure 5.

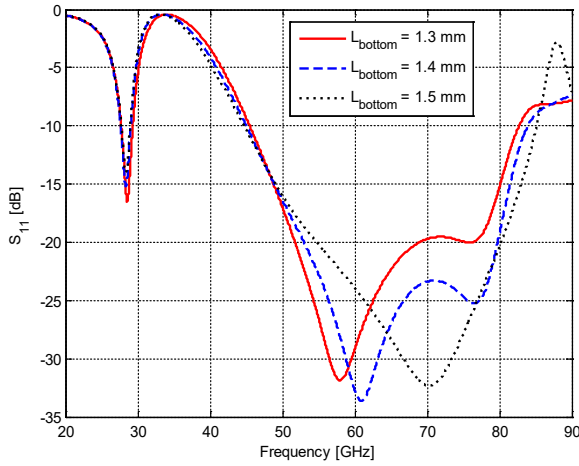


Figure 4. The effect of changing L_{bottom} on S_{11} .

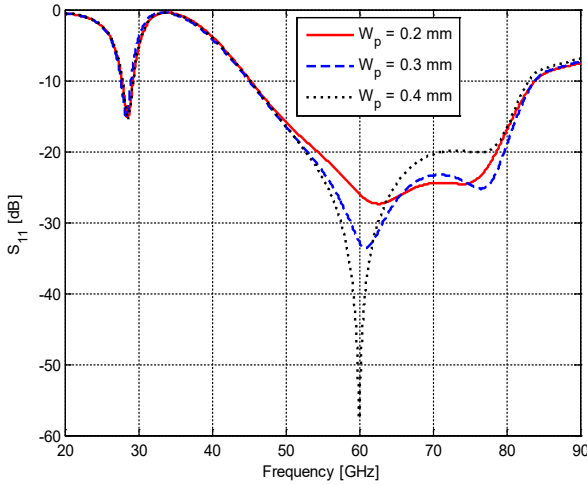


Figure 5. The effect of changing W_p on S_{11} .

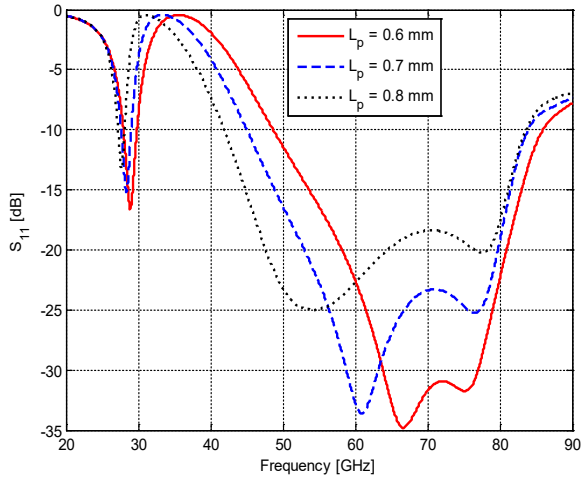


Figure 6. The effect of changing L_p on S_{11} .

As the length of rectangular patch L_p increases, the upper and lower resonant frequency are decreased but the antenna matching becomes worse. Lower bandwidth is increases whereas the upper bandwidth is decreases as shown in Figure 6. Figure 7 illustrates the return loss varies with L_{fold} . As L_{fold} increases the lower

resonant frequency decreases while upper resonant frequency unchanged whereas the lower bandwidth is not affected, the upper bandwidth is a little decreases as shown in Figure 7. With increasing L_t , the lower and the upper resonant frequencies are not affected whereas the upper bandwidth is decreases, the lower bandwidth is unchanged as shown in Figure 8.

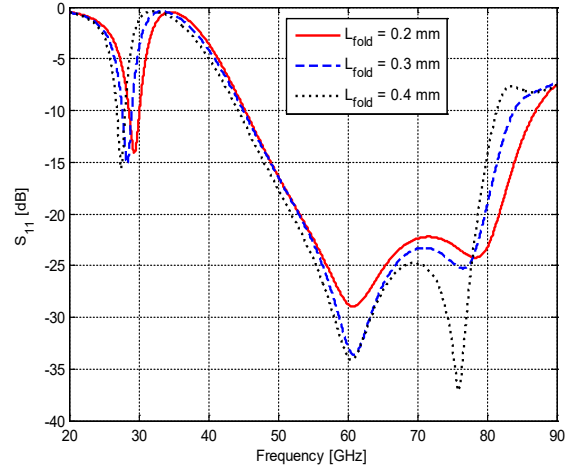


Figure 7. The effect of changing L_{fold} on S_{11} .

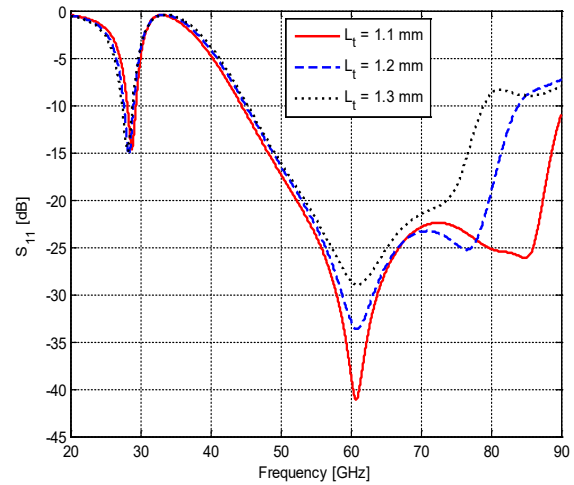


Figure 8. The effect of changing L_t on S_{11} .

4. Antenna Simulation Results

The antenna dimensions are optimized to obtain an Omnidirectional radiation patterns with acceptable performance at 28 GHz and 60 GHz. Table 1 contains the optimized dimensions for the proposed antenna in order to have a dual band.

The optimized antenna is simulated using two simulators based on FIT [17] and FEM [18] in order to validate the results. The S_{11} for the proposed antenna using FEM and FIT is shown in Figure 9. The FEM gives a better matching than FIT and the small deviation in the 60 GHz band is due to the different mesh sizes. As can be noted from Figure 9, the antenna well matched over the two bands and the impedance bandwidth for which $S_{11} \leq -10$ dB in 28 GHz band is extended from 27.52 GHz to 28.96 GHz which

serves the LMDS band and in 60 GHz band is extended from 45.2 GHz to 84.4 GHz which serves the WiGig band.

Table 1: Dimensions of the proposed antenna in mm.

Parameter	Value	Parameter	Value
W_s	1.6	W_p	0.3
L_s	2.9	L_p	0.7
L_g	1.3	W_t	0.1
W_f	0.2	L_t	1.2
L_f	1.4	L_{top}	1.2
L_{fold}	0.3	L_{bottom}	1.4

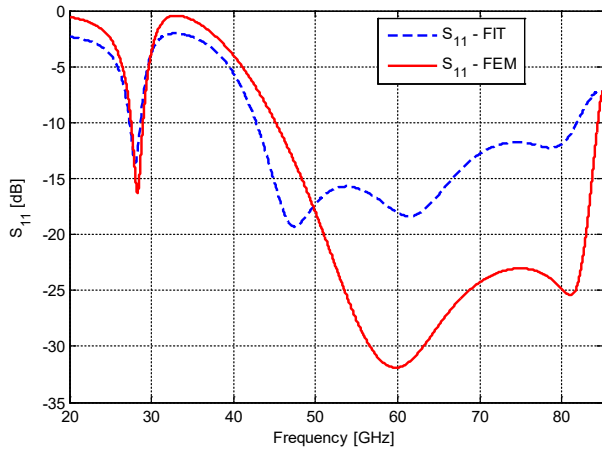


Figure 9. Comparison of S_{11} for the optimized antenna by FIT and FEM.

The antenna resonates at 28.24 GHz with 1.44 GHz bandwidth (5.1 %) and at 64.76 GHz with wide bandwidth of 39.24 GHz (60.6%). The VSWR is shown in Figure 10 with $VSWR \leq 2$ using FIT and FEM in the entire bandwidths of both bands.

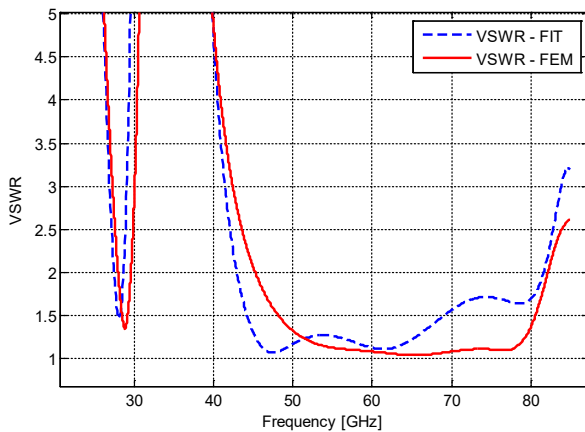


Figure 10. VSWR using FIT and FEM.

The antenna real and imaginary input impedance for the proposed antenna is illustrated in Figure 11. As can be noted from Figure 11, the antenna is well matched to 50 Ω TL since the real part of the input impedance is approximately 50 Ω while the imaginary part tends to zero along the entire bandwidths of both bands.

The radiation patterns at 28 GHz and 60 GHz are illustrated in Figure 12. The directivity pattern at 28 GHz with D_o of 2.28 dBi is illustrated in Figure 12(a) and at 60 GHz with D_o of 3.41 dBi is shown in Figure 12(b). The total antenna efficiency along the lower band is 93 % while along the upper band is 85.5 %.

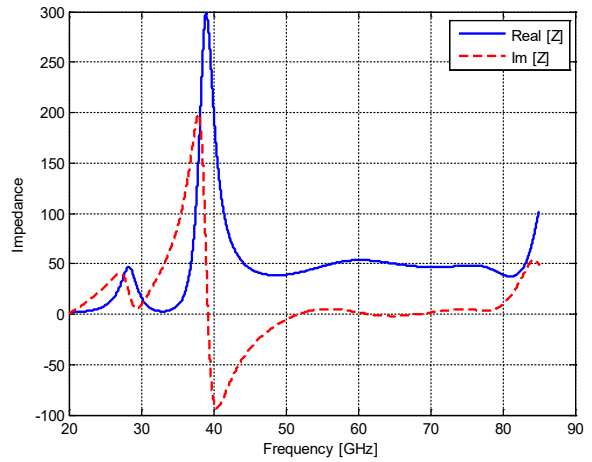
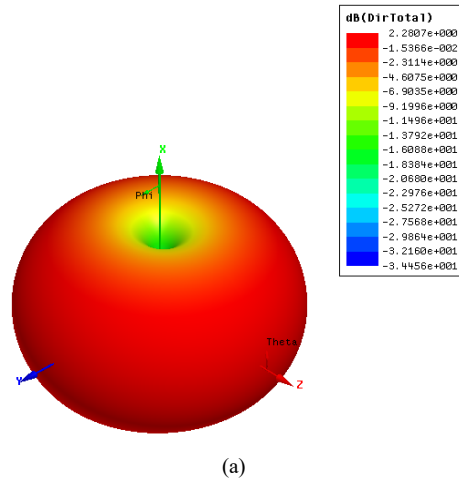
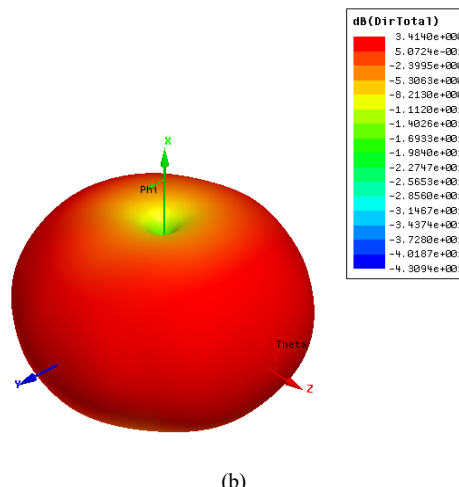


Figure 11. The real and imaginary part of input impedance for the proposed antenna.



(a)



(b)

Figure 12. The directivity patterns at a) 28 GHz ($D_o = 2.28$ dBi) and at b) 60 GHz ($D_o = 3.4$ dBi).

5. Conclusion

In this paper, antenna with two rectangular patches and a T-shaped folded patch is designed, optimized, and simulated. The antenna exhibits two bands resonate at 28.24 GHz and 64.76 GHz with bandwidths of 1.44 GHz (5.1%), and 39.24 GHz (60.6%) respectively. The lower band (Ka-band) is suitable for LMDS while the upper band (V-band) with wideband is suitable for WiGig. The omnidirectional radiation pattern is obtained using partial ground plane with maximum directivities of 2.28 dBi and 3.414 dBi at the two bands respectively. The total efficiency along the entire lower and upper bands exceeds 85% which can be used in the fifth generation applications.

Acknowledgment

The author gratefully acknowledges the support of Prince Sultan University.

References

- [1] M. S. Ibrahim, "Dual-band microstrip antenna for the fifth generation indoor/outdoor wireless applications," in *2018 International Applied Computational Electromagnetics Society Symposium (ACES)*, 2018, pp. 1-2.
- [2] S. S. Jaco du Preez, *Millimeter-Wave Antennas: Configurations and Applications*: Springer International Publishing Switzerland, 2016.
- [3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, vol. 1, pp. 335-349, 2013.
- [4] D. Sanchez-Hernandez, Q. H. Wang, A. A. Rezazadeh, and I. D. Robertson, "Millimeter-wave dual-band microstrip patch antennas using multilayer GaAs technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 44, pp. 1590-1593, 1996.
- [5] H. Jie-Huang, W. Jin-Wei, C. Yi-Lin, and C. F. Jou, "A 24/60GHz dual-band millimeter-wave on-chip monopole antenna fabricated with a 0.13- μ m CMOS technology," in *2009 IEEE International Workshop on Antenna Technology*, pp. 1-4, 2009.
- [6] I. K. Kim and V. V. Varadan, "Electrically Small, Millimeter Wave Dual Band Meta-Resonator Antennas," *IEEE Transactions on Antennas and Propagation*, vol. 58, pp. 3458-3463, 2010.
- [7] T. Y. Lin, T. Chiu, and D. C. Chang, "Design of Dual-Band Millimeter-Wave Antenna-in-Package Using Flip-Chip Assembly," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 4, pp. 385-391, 2014.
- [8] D. Lee and C. Nguyen, "A millimeter-wave dual-band dual-polarization antenna on liquid crystal polymer," in *2014 IEEE Antennas and Propagation Society International Symposium (APSURSI)*, pp. 775-776, 2014.
- [9] S. Agarwal, N. P. Pathak, and D. Singh, "Concurrent 83GHz/94 GHz parasitically coupled defected microstrip feedline antenna for millimeter wave applications," in *2013 IEEE Applied Electromagnetics Conference (AEMC)*, pp. 1-2, 2013.
- [10] N. Ashraf, O. Haraz, M. A. Ashraf, and S. Alshebeili, "28/38-GHz dual-band millimeter wave SIW array antenna with EBG structures for 5G applications," in *2015 International Conference on Information and Communication Technology Research (ICTRC)*, pp. 5-8, 2015.
- [11] G. N. Tan, X. X. Yang, and B. Han, "A dual-polarized Fabry-Perot cavity antenna at millimeter wave band with high gain," in *2015 IEEE 4th Asia-Pacific Conference on Antennas and Propagation (APCAP)*, pp. 621-622, 2015.
- [12] H. Aliakbari, A. Abdipour, R. Mirzavand, A. Costanzo, and P. Mousavi, "A single feed dual-band circularly polarized millimeter-wave antenna for 5G communication," in *2016 10th European Conference on Antennas and Propagation (EuCAP)*, pp. 1-5, 2016.
- [13] S. Hur, S. Baek, B. Kim, Y. Chang, A. F. Molisch, T. S. Rappaport, K. Haneda, and J. Park, "Proposal on Millimeter-Wave Channel Modeling for 5G Cellular System," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 454-469, 2016.
- [14] C. J. Hansen, "WiGiG: Multi-gigabit wireless communications in the 60 GHz band," *IEEE Wireless Communications*, vol. 18, pp. 6-7, 2011.
- [15] W. H. Yang J., Lv Z., Wang H., "Design of miniaturized dual-band microstrip antenna for WLAN application," *Sensors*, vol. 16, pp. 1-15, 2016.
- [16] C. A. Balanis, *Advanced Engineering Electromagnetics*, Second Edition ed.: JohnWiley & Sons, New York, 2012.
- [17] (2015) CST Microwave Studio. Available: <https://www.cst.com/products/cstmws>
- [18] High Frequency Surface Structure (HFSS) (15 ed.). Available: <http://www.ansys.com>

Building an Online Interactive 3D Virtual World for AquaFlux and Epsilon

Omar Al Hashimi*, Perry Xiao

London South Bank University, School of Engineering, SE1 0AE, UK

ARTICLE INFO

Article history:

Received: 31 October, 2018

Accepted: 16 December, 2018

Online: 23 December, 2018

Keywords:

3D modelling

Virtual Reality(VR)

AquaFlux and Epsilon

3ds Max

Web 3D applications

Virtual User Manual (VUM)

ABSTRACT

In today's technology, 3D presentation is vital in conveying a realist and comprehensive understanding of a specific notion or demonstrating certain functionality for a specific device or tool, especially on the World Wide Web. Therefore, the importance of this field and how its continuous enhancement has become one of the dominant topics in web development research. Virtual Reality (VR) combined with the use of a 3D scene and 3D content is one of the best delivering mechanisms of this realist ambience to users. AquaFlux and Epsilon are clinical instruments that were built, designed, and developed at London South Bank University as research projects for medical and cosmetic purposes. Currently, They have been marketed and used in almost 200 institutions internationally. Nevertheless, considering the type of these tools, they often involve on-site thorough training, which is costly and time-consuming. There is a real necessity for a system or an application where the features and functionalities of these two instruments can be illustrated and comprehensively explained to clients or users. Virtual User Manual (VUM) environment would serve this purpose efficiently, especially if it is introduced in 3D content. The newly created system consists of a detailed virtual guide that will assist users and direct them on how to use these two devices step-by-step. Presenting this work in a VR immersed environment will benefit clients, user and trainees to fully understand all the features and characteristics of AquaFlux and Epsilon and to master all their functionalities.

1. Introduction

This paper is an extension of work originally presented in *Advances in Science and Engineering Technology International Conferences (ASET) 2018* [1]. The current research paper presents the development process of a web-based interactive 3D virtual world that illustrates and covers all steps of how AquaFlux and Epsilon operate. VR collectively used with 3D objects presentation will efficiently serve the purpose of illustrating all features and functions of any newly purchased device in an immersed world that gives viewers a real feel of the experience. The popular change to VR in online training or education is going to promote a new learning concept, where clients, trainees or even students not only gain knowledge but also communicate with each other by changing content in a variety of ways. The main feature of VR is the prospect of social interaction, providing the ability for immediate actions and reactions in real time. VR environment has constantly been associated with 3D modelling; it is by far one of the best ways to illustrate and show any 3D content, object, scene, model, etc. The word virtualisation generally depicts the separation of a resource or request for a service from the underlying physical delivery of that service. An additional factor that makes virtualisation very

practical and useful is the interactivity that 3D multimedia applications can provide it, particularly when the whole project displayed on-line using the WWW [2]. Moreover, users of the current Internet age are ready to shift from 2D online presentations to 3D. The arrival and the broad use of 3D web materials have amplified the necessity to acquire improved 3D technologies and generate an extremely advanced and practical product. Therefore, it is suitable to upgrade and improve the usage of those 3D technologies and interactive environments to new fields, like e-learning, Virtual Learning Environment (VLE), museums, e-commerce, online training, Learning Management System (LMS), tourism, health and the government part. Adapting these enhanced 3D virtual interactive apparatus into those latest fields has immensely improved and upgraded the user understanding and became stimulating. Online 3D VR worlds and contents have amplified users' perceptions as they deliver and comprise physical world appearances and attributes to users and permit them to connect to it resulting in immersing users into their environments. 3D VR world, used through the WWW, would be as viable as using an application locally. Therefore, the concept of VR can be incorporated with social media to additionally increase its attraction. Nonetheless, a new VMware report explains that the promising future of 3D/VR technology in our daily applications

*Omar Al Hashimi, London South Bank University, alhashio@lsbu.ac.uk

can be entirely used solely if joined with the improvement of a competent and simple to use ways of building, controlling, search and demonstration of interactive 3D multimedia content, that can be practised by skilled and novice users [3]. In this context, London South Bank University's engineering lab has created and designed AquaFlux and Epsilon, which are clinical instruments. Displaying these medical devices in the VR approach would demonstrate their capabilities as being practised in the real world. Xu (2015) states AquaFlux as:

a new condenser based, closed chamber technology for measuring water vapour flux density from arbitrary surfaces, including in-vivo measurements of transepidermal water loss (TEWL), skin surface water loss (SSWL) and perspiration. It uses a cylindrical measurement chamber [4].

Compared to other technologies, AquaFlux has greater sensitivity, greater repeatability, and most importantly, the measurement results are independent of the external environment. Biox (2014) describes Epsilon as:

a new instrument for imaging dielectric permittivity (ϵ) of a wide variety of soft materials, including animal and plant tissues, waxes, fats, gels, liquids, and powders. Its proprietary electronics and signal processing transform the sensor's native non-linear signals into a calibrated permittivity scale for imaging properties like hydration. [5].

The Epsilon user manual describes that "The system consists of a hand-held probe, a parking base, and an in-vitro stand, securely stored in a purpose-designed case" [6]. Both devices are shown in Figure 1.



Figure. 1. Medical instruments AquaFlux (L), Epsilon (R)

2. The utilisation of 3D contents and related work

After the advent of Web 2.0, there has been a major improvement in web applications. Web 2.0 supports and encompasses various functions, like team working, communicating, and the connection between computer and Internet users. 3D immersive virtual worlds (3DVW) are one of the important applications of Web 2.0, which are computer-generated, virtual, online, graphics, multimedia and 3D worlds [7]. The major notion of building such an object in 3D is to illustrate to users the reality of the content. 3D modelling is an essential element in the area of VR technology [8]. The development of a 3D content can be divided into three main steps:

- 1- 3D modelling
- 2- Layout and animation
- 3- Rendering

www.astesj.com

Normally, 3D models are created using 3D modelling software, like 3ds Max, or Maya, or their open source equivalents – although open source tends to be less complex with fewer advanced features. A 3D scene model is created from geometrical shape objects, for example, rectangles, triangles, circles, cones, etc. e-learning contents are placed into hypermedia documents, causing difficulty in 3D integration into HTML files. Web browsers are not yet designed to deal with the (normally large) 3D data files which require a large number of computational resources. It takes time and effort to digest such proposals. Therefore, readily available Mozilla and Google proposals can be adopted as standards. Mozilla and Google use the same technologies: OpenGL interfaced with JavaScript [9].

3D objects are widely used in science, technology, engineering, health, education, cosmetics, simulation, e-learning etc. It has been used in the areas named above joined with VR to stimulate the displayed content as it is applied and practised in the physical world. Another aspect that 3D contents and VR play a very active role in, is the Vocational Education System. It is a system that is specifically designed to support the industry and manufactures by offering vocational and technical courses delivered in VR ambience and PLE (Personal Learning Environment) which allows users, learners to control their learning and manage their own learning experience (distance learning). Kotsilieris, Dimopoulou (2013) point out:

The development of 3D Virtual Worlds plays an important role in e-learning and distance learning. Through three main features: I) creating the illusion of a 3D environment, II) support the application of avatars as virtual representations of human users, III) offer communication and interaction tools to their users. The evolution of virtual worlds is a result of a rapidly evolving field of electronic games. In brief, Virtual Worlds are designed to offer real-time communication tools, interaction capabilities and collaboration [10].

Some up-and-coming technologies will overpower some difficulties that are faced currently in areas like education, technology, health, engineering and others. These include computer graphics, augmented reality, computational dynamics and virtual worlds. Lately, we have witnessed a number of novel thoughts emerging in the literature related to the future of education. As has been pointed out by Potkonjak, Gardner, Callaghan, Mattila, Guetl, Petrović, Jovanović (2016):

Technological examples mostly related are e-learning, virtual laboratories, VR and virtual worlds [11].

Accordingly, the benefit of adopting VR in the medical and healthcare sectors is to teach and train medical students, trainees and clients on how to use medical devices and instruments and how to conduct some medical tasks. Web-based and online 3D objects used in medical training tools and environments showed to develop the educational process. This web-based virtual medical system of devices for cardiac diagnostic and monitoring functionalities has been created and built to assist in the process of training medical students, qualified health personnel and non-medical staff to carry out an Electrocardiogram (ECG), an Automatic External Defibrillator (AED) and a blood pressure device [12].

Those applications guarantee an interactive e-learning experience in the medical field. Also, considering the main objective to emulate real patients, anatomic regions, and clinical

tasks and to represent real-life conditions in which this medical tool is built for. These virtual environments allow interaction between users and the system as well as manipulation with very sensitive reactions similar to that of real-life objects. This type of system will promote learning by practising, which makes the whole experience straightforward and enjoyable [12]. Using VR in such projects will help to achieve an extremely immersive experience [13].

Another related work in this regards is the ARCO project (Augmented Representation of Cultural Objects). The ARCO project is specifically built for the tourism and heritage industry. Its main objective is to build up an entire virtual museum that contains a collection of technologies for producing, changing, controlling and demonstrating cultural objects in VR environment that are accessible globally via the Internet [14].

3. The implementation of the interactive Virtual User Manual (VUM) of the medical instruments

The design process of AquaFlux and Epsilon medical devices was slightly difficult and experienced the array of choices for selecting the suitable 3D modelling software; in this project, the software modelling tool used is 3D Studio Max (3ds Max). In addition, adding interactivity between users (clients) and the devices, the software Adobe Flash CS6 is used. Other modelling software products were used at the early stages of this project, like Google Sketchup, Blender and Unity were used as a final result of the research was to create, build and develop the objects and all 3D scenes in 3ds Max for its professionalism and the complete collection of functions and options that it had to offer.

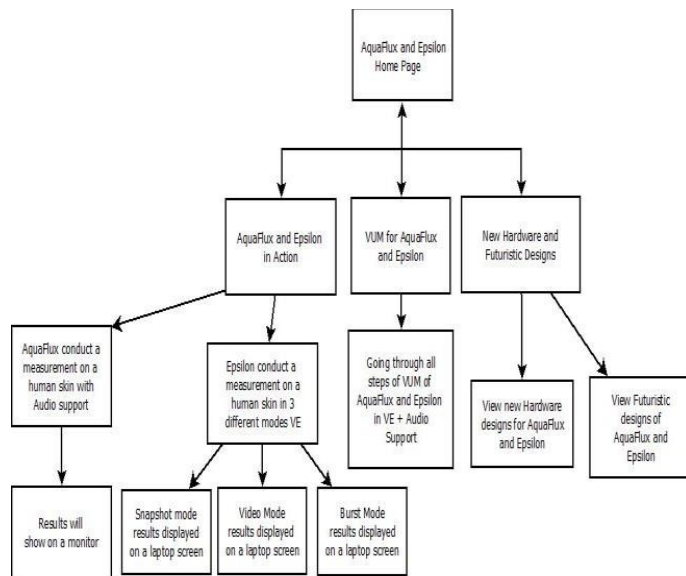


Figure 2. Flowchart of AquaFlux and Epsilon online 3D Environment

AquaFlux and Epsilon are clinical devices used for skin treatment, belonging to the health and medical sector. Figure 2 shows the workflow diagram that illustrates all the steps of the AquaFlux and Epsilon 3D virtual environment system and describes the development processes of the Virtual User Manual (VUM). At the home page, users have three options: AquaFlux and Epsilon in Action, users can perform virtual skin measurements using the instruments. For VUM of AquaFlux and Epsilon, users can go through all the training steps, with audio and illustrative text support. For New Hardware and Futuristics

Designs, users can view the new hardware designs and futuristic concepts for both medical instruments.

The health sector is largely connected with the utilisation of leading technologies like 3-dimensional objects and virtual presentation for demonstration, education, testing and performing a number of medical operations and routines. 3D Virtual Worlds (3DVWs) have been used in a variety of applications in the medical sector and applied in health-related activities [7].

3.1. The development and design of AquaFlux and Epsilon using 3ds Max modelling software

3ds Max modelling tool can swiftly and professionally produce 3D scenes and objects [15]. Like in various 3D development tools, initially, we have to create and design the medical tools. AquaFlux and Epsilon contain a base part and a probe; on the base part, there are little design obstacles, e.g. cabling input port and buttons that have to be switched left or right in AquaFlux's case. The probe and the base have been calculated with an actual ruler to let us build the models in an extremely realistic form. Additionally, considering some photos to get an accurate proportion and size for every model, for instance, how large the probe would be in comparison to a human hand.



Figure 3. AquaFlux probe measured by a ruler

Once the developer assesses the accurate size of every object, we use 3ds Max to design the medical devices. Modelling software is shown in the figures below for AquaFlux and Epsilon:

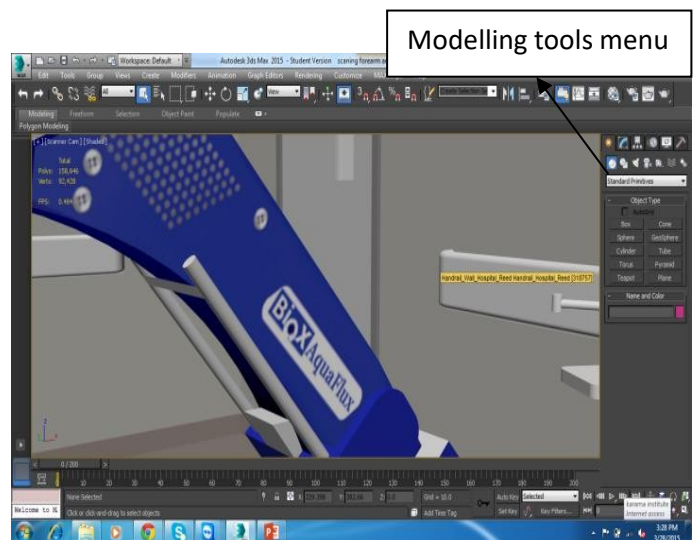


Figure 4. Tools box menu 3ds Max software used to model objects

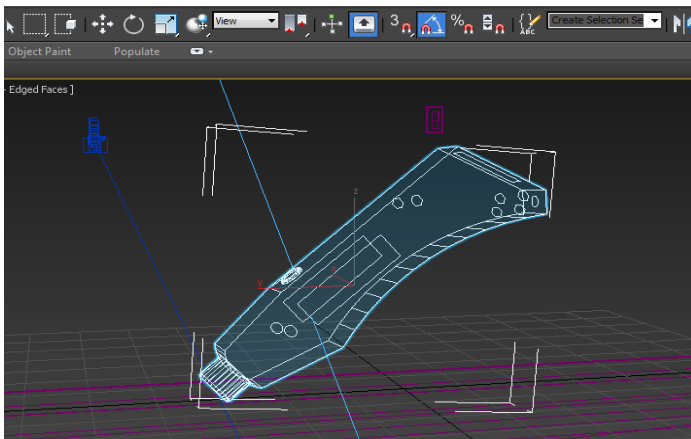


Figure 5. AquaFlux probe modelling in 3ds Max

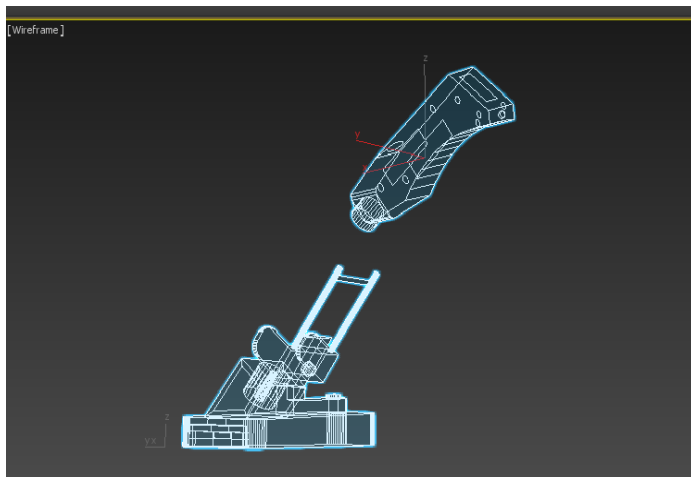


Figure 6. Snapshot of AquaFlux probe and base design in 3ds Max

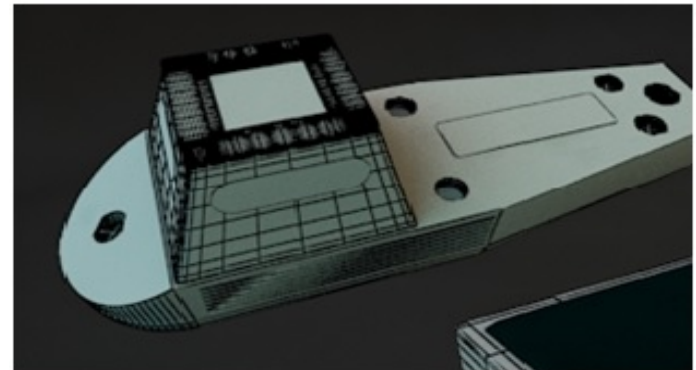
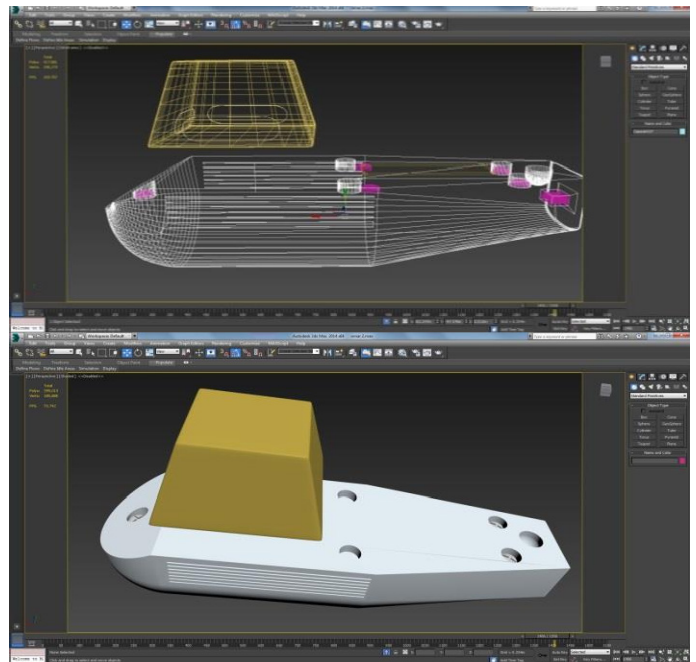


Figure 7. Different stages of modelling Epsilon's probe in 3ds Max

The designer is able to choose a diverse object's shapes from the modelling menu e.g. a box, cone, and sphere, tube and cylinder and gives names for all parts. As the process of modelling is progressing, the developer has to construct a real reference to check how the recently created model is comparable to the real object via situating the real model's image close to the newly designed one as shown in Fig 8:

3.2. Adding materials to AquaFlux and Epsilon

Materials and colours are considered early to befit the model reference if it is really sufficient to satisfy the client's demands. The selection process of the material is closely examined to achieve a minimum rendering time and to present a very real product. Materials that are rich in glossiness and reflection will consume most of the processing (rendering) time, however, without materials, the model would be very unreal to viewers. Vray materials are used in this project, and the third-party renderer is Vray for the whole work. Vray is one of the industry standards for generating specialised pictures. For the Epsilon Probe and other objects, a semi-gloss with a whitish material is used with 0.6 glossiness and a subdivision of 20. Those settings were chosen to reduce the time it usually takes during the rendering process, as rendering is a crucial part of any 3D creation projects.

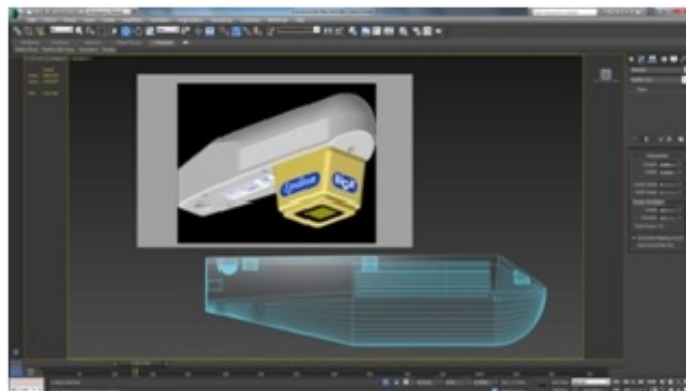


Figure 8. The original picture of Epsilon compared to an Epsilon object.

In this project, High Definition Range Imaging (HDRI) is used for the background to make it slightly more practical for the lights and shadows. HDRI is absolutely valuable to emulate shadows and lightings of a real environment that will be projected on the objects in the 3D scene. Although using HDRI will add some extra time during rendering each scene, the glossiness and reflectivity of the model are changed slightly low, to improve the overall execution-time process.

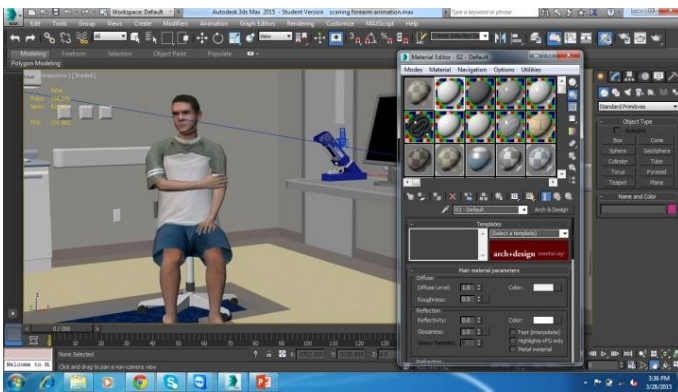


Figure 9. Colours added to AquaFlux using the material box menu

animation is slightly easier as it does not require bones and additional complicated rigging tools. Animating objects in this work has been accomplished in two ways: attaching an object into a group of objects and using the collapse utility to cave in multiple selected objects into a single object. The necessary points for all movements must be thoroughly calculated prior to the animation phase. In the animation stage, the Timeline is set to 1000 frames and 15 frames per second because this work is specifically designed for online environment only. The number of frames per second is decreased to suit the file size of the images that will be rendered to be used on the WWW. Applying curve editor menu options for animating all objects and scenes. When modifying curve points, all animated models can be managed via (speed, timing, and movements options).

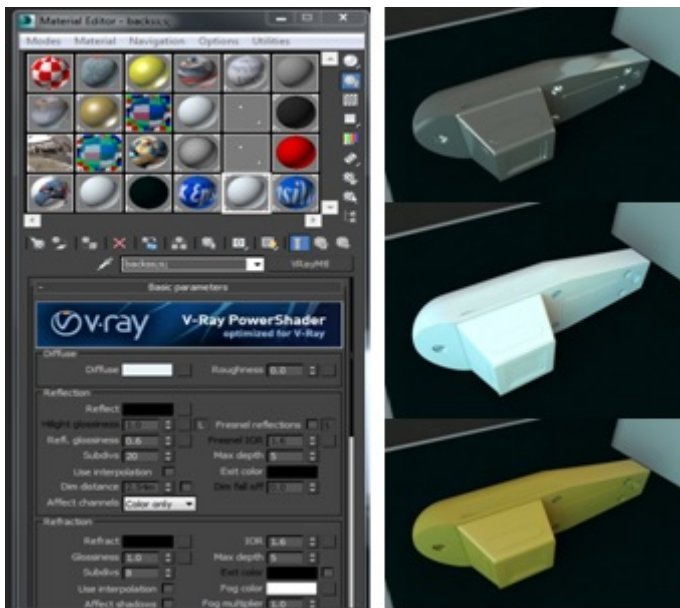


Figure 10. Materials menu in 3sd Max (L), Materials added to Epsilon probe (R)

Once the materials eventually determined and added taking into account the time and quality of the work, we arrange all 3D contents for the next process: Animation. In order for all the objects in our 3D scene to be animated correctly, all objects in the selected scene must be grouped together via animation attach buttons in the command panel under edit geometry menu. This means all objects should be grouped, attached or sometimes called fused together. If any object, for any reason, has broken up and not been linked and left out of the attached objects group, the keyframes associated with that model will disappear after this stage and the model will no more be animated.

3.3. The process of animating AquaFlux and Epsilon

Animating 3D objects is an essential phase of our project's design. Almost a third of the time of this project was used on the animation process of our objects that were modelled and built earlier. Planning is critical at this stage for each planned scenario. Planning should be prepared and anticipated for every object in our scene: how the object will interact with other objects, the proposed movement and the path of that motion that the individual object will experience in the virtual world. It is a somewhat prolonged process. Contrary to character's animation, this type of

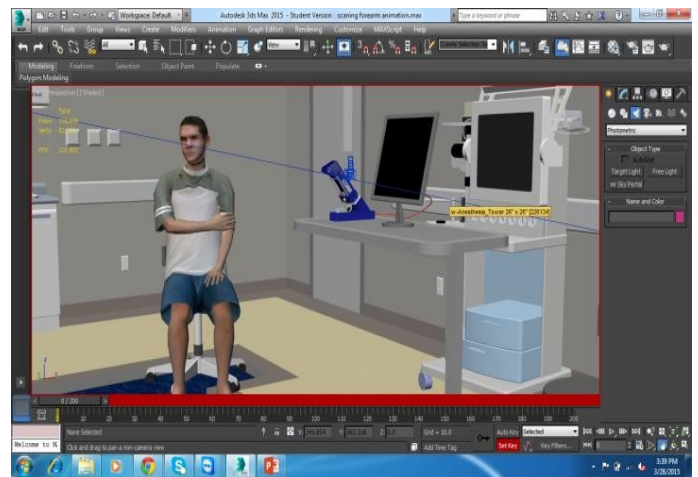


Figure 11. Adding keyframes (redline) in 3ds Max

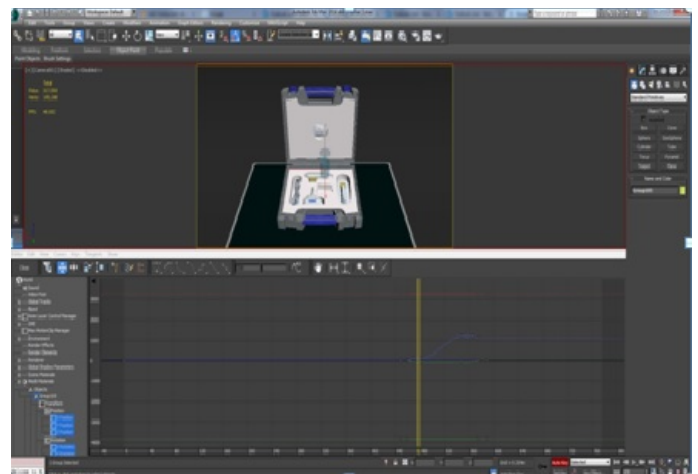


Figure 12. Using the Curve editor in 3ds Max

3.4. The process of rendering AquaFlux and Epsilon

After the animation is completed and setting keyframes from a specific position into where we want our object to stop at a certain point. Currently, all objects are now prepared for the last phase of the modelling process, which is rendering. In rendering the developer has to monitor all frames being rendered. Nevertheless, at this point in our work, it is extremely vital to monitor the speed and time that each object takes to complete rendering. Every frame is closely checked if it is in the correct

position (number) and whether all models have the exact shadows and lighting. Occasionally models are spotted floating that are unnoticed in the animation process. In rendering, a small window will pop up showing how rendering each square pixel of the object, calculating the objects materials, colours, shadows, light reflections, etc. In the rendering options box, it is possible to select the format of the rendered images and their file locations that will be exported into Adobe Flash CS6 later on for adding interactivity via linking all rendered images, scenes and objects.

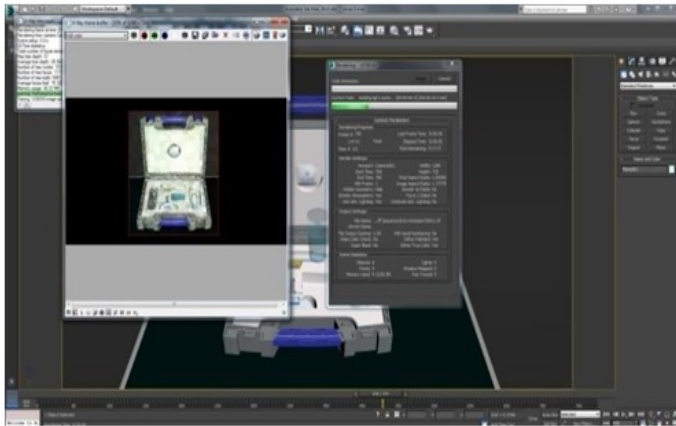


Figure 13. Rendering window after all keyframes and files setup completed

3.5. Creating links and adding interactivity to AquaFlux and Epsilon

All rendered image sequences of the medical devices AquaFlux and Epsilon have to be exported from 3ds Max software into Adobe Flash CS6 in order to add interactivity to the project via creating interactive buttons.

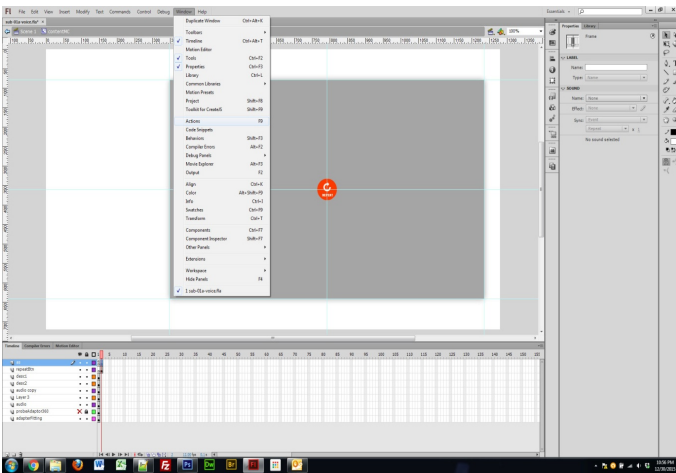


Figure 14. Adding ActionScript into a scene in Flash CS6

AquaFlux VUM is a completely interactive website that demonstrates all functions, features and a sample skin measurement process conducted on a human.

The AquaFlux VUM is an interactive system. Users can click on the device's probe to command it to go towards a human's forearm to perform a sample skin measurement. Each step of the VUM is assisted with audio instructions recorded earlier to assist in exemplifying the operational process for this medical device.

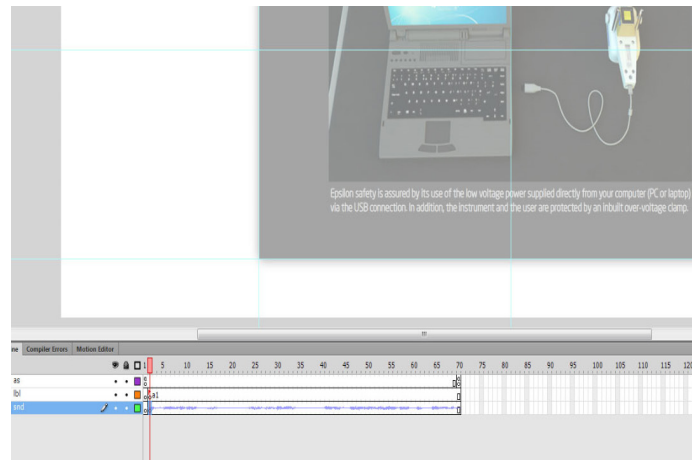


Figure 15. Syncing a voice clip with the text frame in Epsilon scene

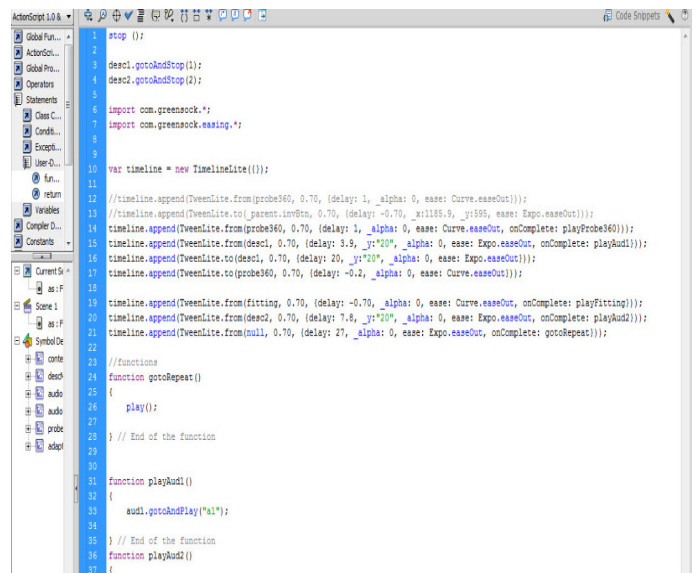


Figure 16. ActionScript code for linking the button to play, replay and stop the audio clip

The above steps (Figures 14, 15 and 16) can be repeated to all the 3D scenes and objects that required some form of connectivity and interactivity with the user in a VR atmosphere by using buttons in Flash CS6.

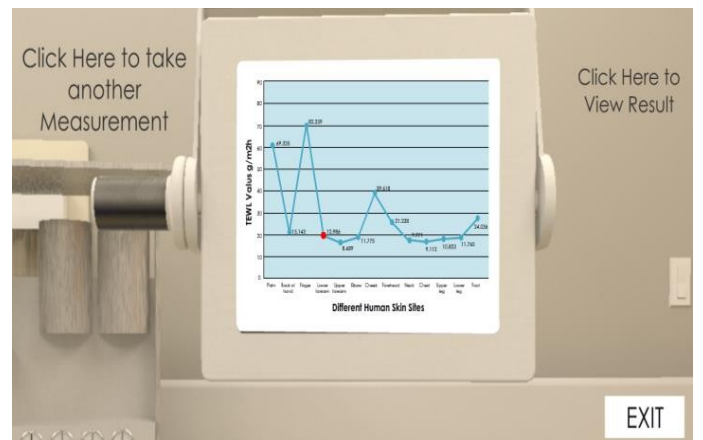


Figure 17: The result of the process of skin measuring will be shown on a screen

Epsilon VUM follows a similar approach to conducting a human skin measurement procedure for medical purposes. Figure 18 shows the Epsilon instrument in action.



Figure. 18. Skin measurement process taking place on a patient using Epsilon

Once carrying out the measuring process, the scene's camera will move towards the laptop's screen to display the scan results as shown in Figure 19.

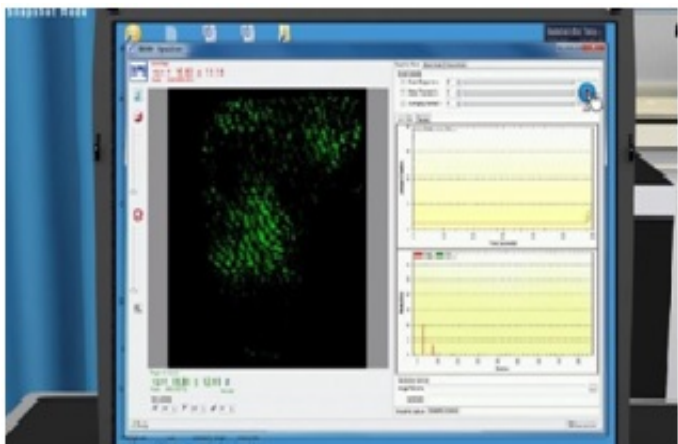


Figure. 19. The result of the Epsilon scanning process is displayed once clicked on the blue button, top right corner.



Figure. 20. Epsilon protective case and all its components are shown in a 3-dimensional interactive atmosphere

In Epsilon VUM, users are able to perform human skin scanning in three various approaches, snapshot, burst and video mode. All results of scans (images and videos) are stored in a folder on the laptop's hard disc. Scan buttons and tabs are all interactive.

4. Designing and introducing new accessory, hardware and holders of AquaFlux and Epsilon

During the final stages of this project, a fresh design plan has evolved, which will contribute positively to the entire skin measurement process. Moreover, it will make a great addition to the practicality of the two devices. The notion was to develop and build a novel part (holder) for both tools. It will contribute massively to the benefit of the entire process making it more effective and competent by allowing the simultaneous measurement of several patients. The newly designed holder for Epsilon was thoroughly considered and particularly built to suit the Epsilon structure and patient's easement. The material added to the newly developed holder was a cloth semblance material to emulate the real world object.



Figure. 21. Epsilon newly designed holder strap in action

In relation to Aquaflux, it was rather sensitive to create the novel holder due to the shaped probe. After a thorough analysis of the structure of the device, AquaFlux's new accessory was completed. The new holder already considered both patients relive and client's ease. AquaFlux probe is made of metal and fairly heavier in comparison to Epsilon. Thus, the black frame around the border of the holder is recommended to be manufactured of the magnet material providing AquaFlux with further support, firm placing, and accurate reading.



Figure. 22. AquaFlux newly designed holder strap with magnet head

Conclusively, the novel ideas of designing and building new hardware for these medical tools has optimised the functionality of both devices. Furthermore, to provide a broad number of people concurrently with the service, futuristic concepts and designs have evolved. The innovative designs of AquaFlux allow it to rotate in all directions giving it extra flexibility and ease of use for all parties involved in the process as shown in Figure 23 below:

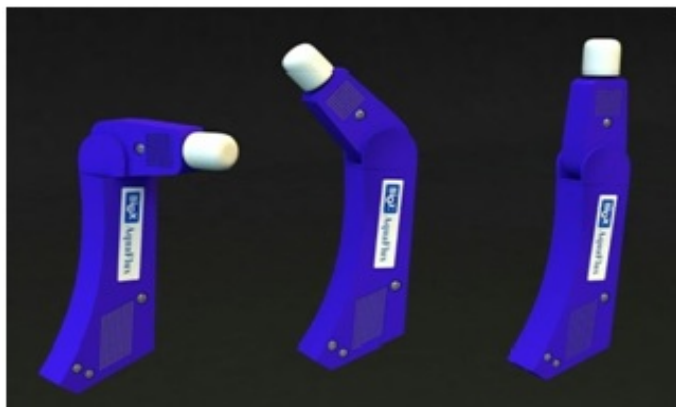


Figure. 23. Rotatable AquaFlux for further flexible positions

The AquaFlux arm-band is an entirely new idea, extremely lightweight and it allows further ease of use. It is padded and soft which can be used almost everywhere on the human's body. It has a Bluetooth capability to connect to a PC or laptop wirelessly and spare the trouble of cabling that can be distracting.

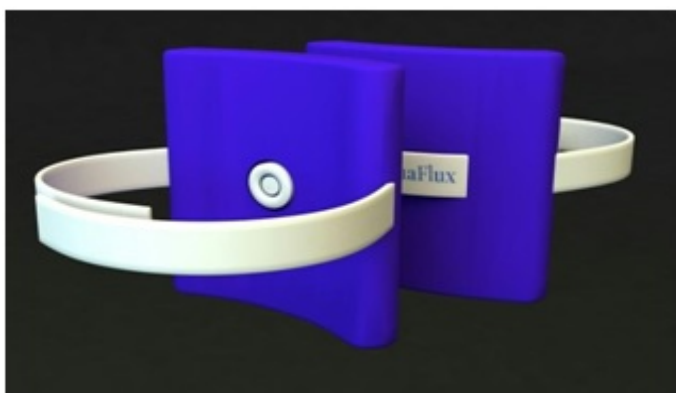


Figure. 24. AquaFlux wireless stripe, newly designed hardware

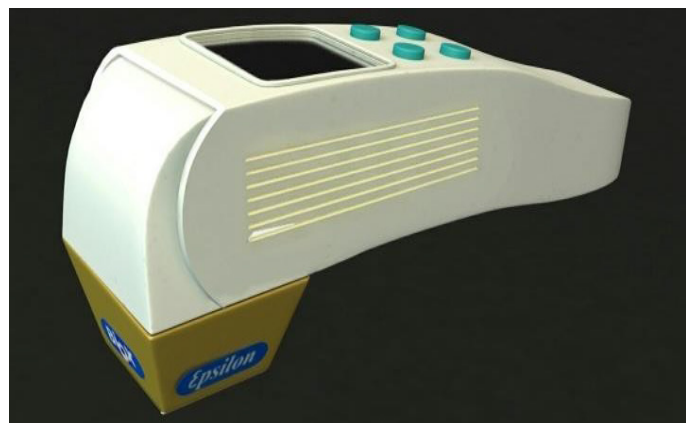


Figure. 25. Epsilon turning head, on probe monitor and operational buttons

Epsilon's turning head assists to carry out skin measurements in vertical/horizontal positions. A fresh screen is supplemented to the initial probe for the purposes of accessibility, simplicity and comfort. Additionally, Functional buttons have been supplemented to the top surface of the handle to offer on probe actions, e.g. on/off, scan, reset. Epsilon device can be wholly moveable.

5. Virtual Reality creation

Virtual Reality is a second world (simulated environment) accessed by various users, and it resembles our world in most of its characteristics depending on what is required from that virtual world to present. Users in virtual worlds can be presented by avatars or walkthrough scenes to conduct a variety of activities.

There are currently two methods to create a VR environment:

- 1- Using a 360-degree video: this method is easier than the second one, as you can record real-life video footages and use those footages in building the content. This method needs another tool at the later stages to view the 3D content, HMD (Head Mounted Display) unit. With this HMD unit, users can experience the new world of computer simulated reality that imitates physical existence in the actual world.
- 2- Using 3D animation: this method is when all selected objects will be created and developed in 3D modelling tool e.g. 3ds Max, Blender, Maya and many more. It is more expensive as it requires more tools and software to be involved in the construction process as well as requiring the expertise to develop, create, animate, render and subsequently publish the product online. Another way of creating VR is by using game engine software such as Unity. This method is mainly used for creating and developing games.

From the two methods mentioned above, it is clear that the second method does require a huge amount of effort and experience. All aspects of the second method need to be carefully and professionally designed and planned. No real images or footages that can help the creator to replace any physical existence in the virtual world; all have to be created, built and designed from scratch. In addition, it is the method that has been adopted in this work.



Figure. 26. AquaFlux interactive user manual website contains three links

6. Results

6.1. AquaFlux Interactive User Manual

AquaFlux VUM online system is a completely interactive website that displays a sample skin measurements conducted on a human and illustrates all functions and features of the medical tool. The following Figures 26, 27, 28, 29, 30, 31, 32 and 33 present few steps on how to use the VUM.

By clicking on the button labelled Aquaflux in Action, will present a sample of skin measuring process conducted on a client at a clinic.



Figure 27. AquaFlux in its base, waiting for a user to click the link above it



Figure 28. AquaFlux in action, moving to a patient's forearm for a measurement

After clicking on the link in Figure 28, the AquaFlux probe will move towards a patient's forearm to conduct a sample skin measurement, by placing its head that has a cap into the patient's skin.

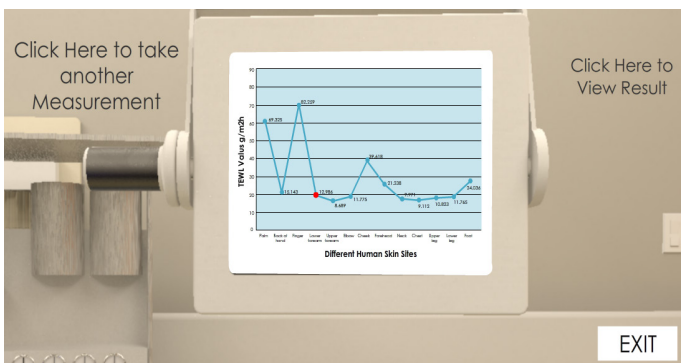


Figure 29. AquaFlux monitor shows measurement results and other user's options www.astesj.com

Once clicked on view results link (top right corner), the reading will show on the screen with an accurate reading of the forearm. A user, a trainee or a client also has another option to take another measurement (top left) or to exit to the main page via the exit button (bottom right).



Figure 30. The 2nd option link to view the VUM

From the main page, clicking the second option takes the user into the AquaFlux VUM page. Here, users can click on the VUM (top right) link to view AquaFlux setting up the process, calibration, caps, AC adaptors, connectors, probe parking, holding the probe, software familiarisation. Also, the process of how a user can fit and slide the probe into the parking base and an explanation of the AquaFlux rear panel unit.



Figure 31. AquaFlux 3D components illustrated in the protective case

As shown previously in Figure 31, once the mouse is placed over one of the other six links, it will popup the part and rotate it in 3D mode; this gives a chance to the user to know all the right names for the AquaFlux device components.



Figure 32. Clicking on the VUM menu, the background becomes blurry

In Figure 38, users can click on VUM's options' menu to display all list items of AquaFlux components. The menu is designed to illustrate all parts of the medical device as well as demonstrating the role of each part in the measuring process displayed in a 3D environment.

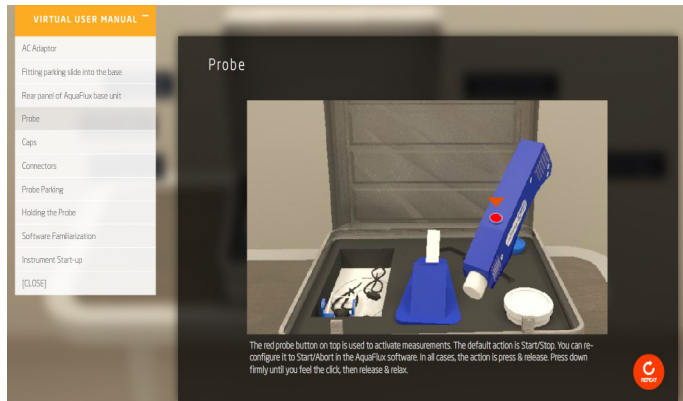


Figure 33. Clicking on one of the options in the VUM menu (probe).

In Figure 33, the user clicked on probe item in the VUM's drop options box, a window appeared to the right of the screen containing instructions and info regarding the AquaFlux's probe with brief descriptions at the bottom of the page. Users can start again the step by clicking on the repeat button (bottom right corner). To go back, users need to click on the [CLOSE] link first then clicking on the link back to the home page.



Figure 34. The main page of Epsilon VUM contains three main links



Figure 35. Clicking on the 1st link to see Epsilon in action

6.2. Epsilon Interactive User Manual

Epsilon VUM online system is a completely interactive website that displays a sample skin measurement conducted on a human and shows all functions and features of the device.

After clicking on the Epsilon probe (see Figure 35) the probe will move towards the patient's hand and placed there. Then the scene will switch into the laptop. After that, the camera moves towards the screen to see the result of skin measurement taken. All the above steps are conducted with the assistance of audio feature.



Figure 36. Epsilon probe moving towards the patient's hand for a measurement

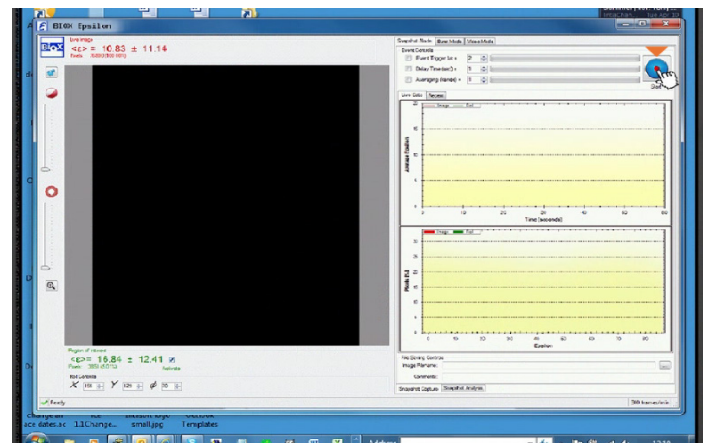


Figure 37. Clicking on the button (top right) to start scanning human skin

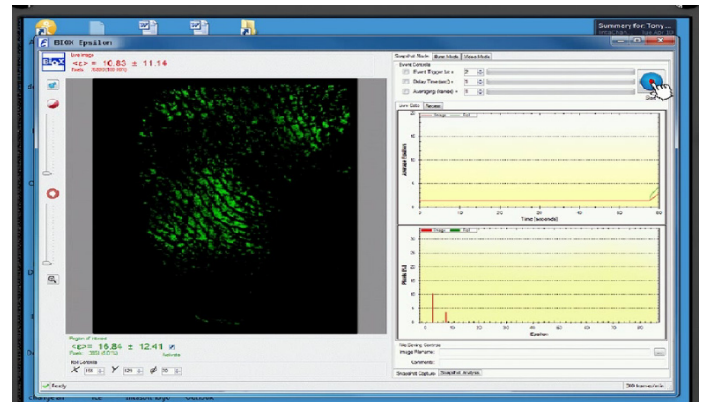


Figure 38. Scanning in a snapshot mode

After clicking on the scan button, scanning will begin in a snapshot mode, in Epsilon users can scan in three various approaches (snapshot mode, burst mode and video mode).

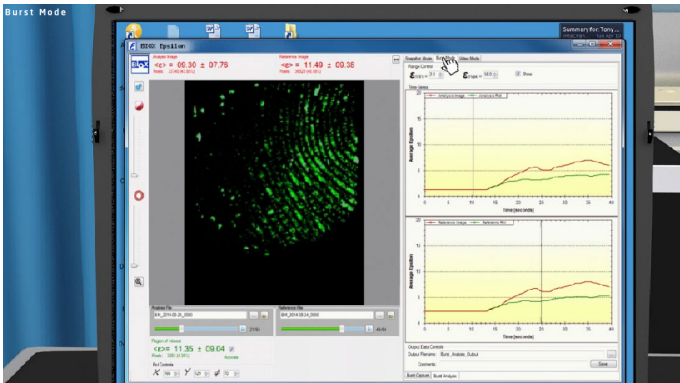


Figure 39. Scanning in a burst mode (clicking the tab at the top to switch mode)

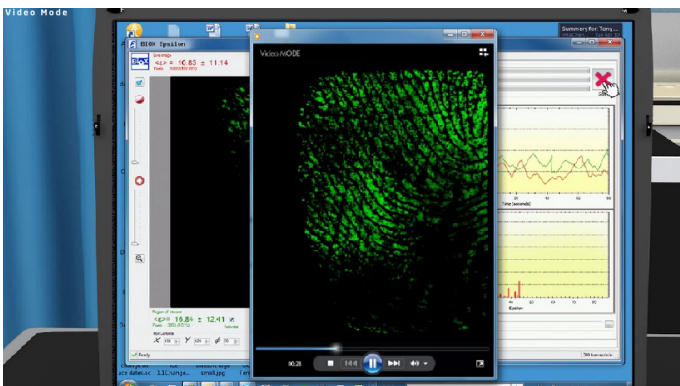


Figure 40. Scanning in a video mode

All files, images and videos of the three different scanning modes saved in their respective folders, snapshot, burst and video folders, users can go to their folders and view, replay or send the captured files. This VUM is assisted with an audio feature to clarify any unclear steps throughout the demonstration process.



Figure 41. Epsilon 3D components in a sturdy case

As previously explained in the AquaFlux VUM (Section 7.1), the second link in Epsilon's main page as shown in Figure 34, will take the user to the Epsilon's VUM menu. Every part or

component located in the protective case will pop up once the user places the mouse on it and start rotating in 3D 360 degree with a brief thumbnail defining that part. The use of mixed technologies of 3D designs and VR results in achieving an incredible engulfing feeling as if being at the real location of the demonstration process [13].

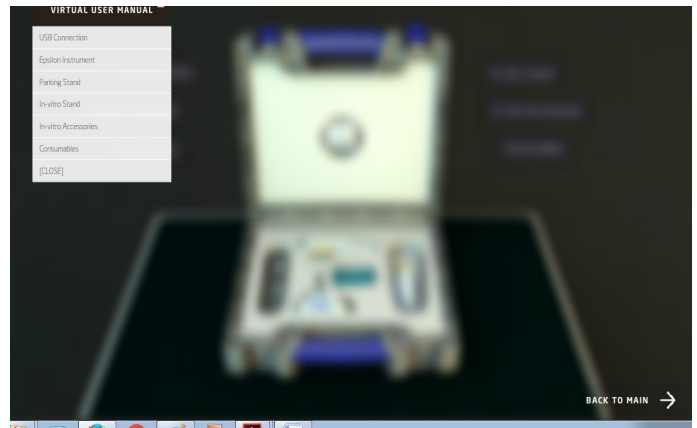


Figure 42. Epsilon VUM drop down menu

In a similar manner to the AquaFlux pervious VUM, Epsilon's VUM contains interactive links to demonstrate USB connection, Epsilon instrument, parking stand, in-vitro stand, in-vitro accessories and consumables.



Figure 43. Epsilon probe sliding into the stand in the VUM menu options

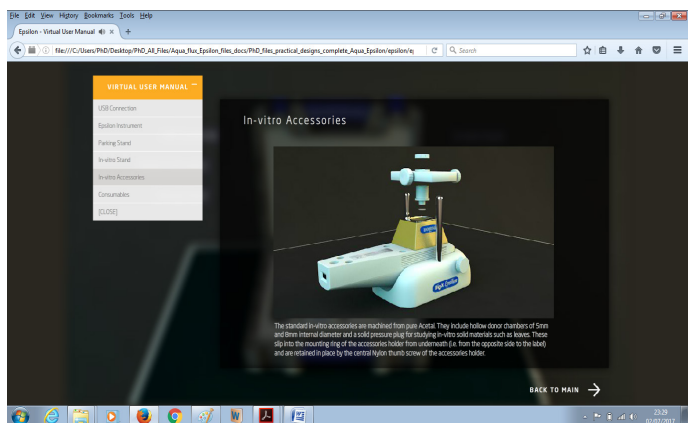


Figure 44. Epsilon in-vitro stand and accessories demonstrated with the aid of audio feature and 3D ambience in the VUM menu options

Figure 44 shows the VUM of Epsilon in operation from the World Wide Web, with the help of audio and illustrative text displayed in a virtual environment and user interaction, the demonstration process will be very smooth, efficient and comprehensive for all users.

7.3 Epsilon and AquaFlux new holders

The third and final button on the AquaFlux and Epsilon main pages will take the user into the new addition that evolved during the process of working on this project, which is adding a new set of holders to each instrument (AquaFlux and Epsilon) that will certainly enhance the measurement process and will provide a more efficient and accurate reading.

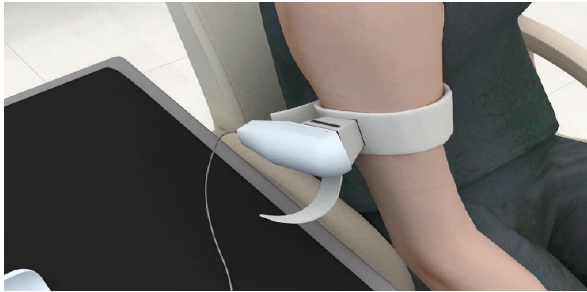


Figure. 45. Epsilon's holder in operation

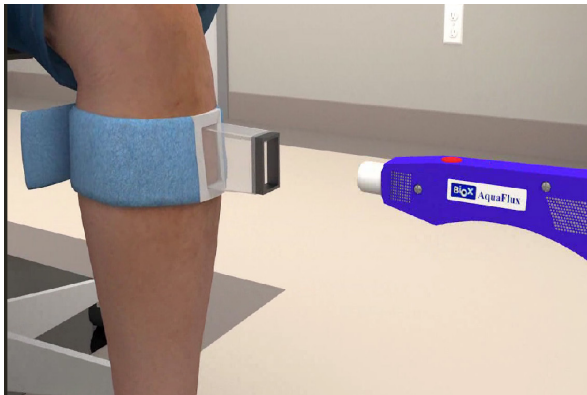


Figure. 46. AquaFlux's holder in operation

7. Evaluating AquaFlux and Epsilon VUM

The VUM of AquaFlux and Epsilon was intended to serve clients, users, and trainees who are interested in purchasing AquaFlux and Epsilon medical instruments and wanted to have a clear and detailed illustrative idea on how these two devices function, and what their features are. From a methodological and marketing point of view, it was vital to conduct a usability study that will show the products' advantages, disadvantages and point out any areas of excellence and parts that require further improvements.

The usability study was conducted by a variety of users, mainly people from a non-IT background. A questionnaire was designed to tackle the most obvious and fundamental questions that could arise while using such a system, moving on to more technical questions. The following table shows the user's feedback on the survey questions that involved 12 independent participants:

	Was the service provided online smooth and error free?	Did you enjoy your experience using the VUM?	Were you successful using the VUM?	Were you able to control the VUM?	Was the audio support feature clear and useful?	Do you think that the VUM will be sufficient to learn how to use the medical devices comprehensively?
1-	Agree	Agree	Neutral	Agree	Agree	Neutral
2-	Neutral	Neutral	Neutral	Neutral	Agree	Neutral
3-	Agree	Agree	Agree	Agree	Agree	Agree
4-	Neutral	Neutral	Neutral	Disagree	Agree	Disagree
5-	Agree	Agree	Agree	Agree	Disagree	Neutral
6-	Agree	Agree	Neutral	Agree	Agree	Neutral
7-	Agree	Neutral	Agree	Agree	Agree	Agree
8-	Agree	Agree	Agree	Agree	Agree	Agree
9-	Agree	Neutral	Agree	Neutral	Agree	Neutral
10-	Agree	Agree	Neutral	Neutral	Agree	Neutral
11-	Agree	Agree	Agree	Neutral	Agree	Disagree
12-	Agree	Neutral	Agree	Agree	Agree	Agree
Agree	10	7	7	7	11	4
Disagree	0	0	0	1	1	2
Neutral	2	5	4	4	0	5

Figure. 47. Survey's questions with user's feedback. 12 users participated

The following illustrative charts show the results of the individual survey questions.

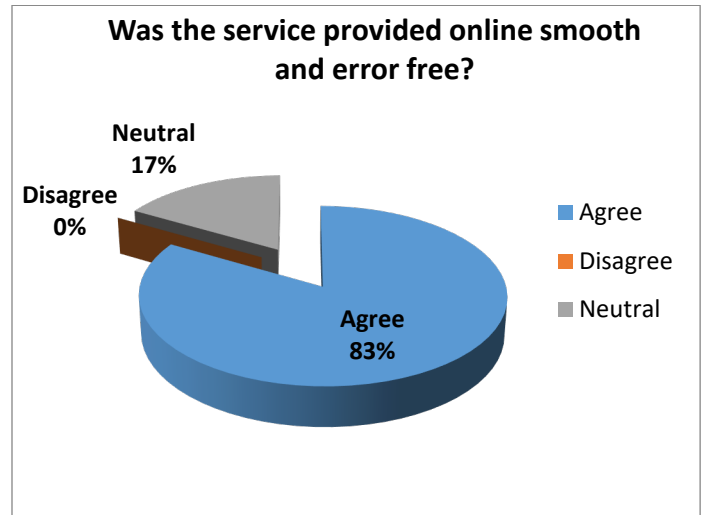


Figure. 48. Shows the success rate of the VUM system being smooth and error-free

The above chart shows that 83% of the participants found that the service provided online was smooth and error-free, whereas 17% were neutral. The result indicates the strength of the virtual system.

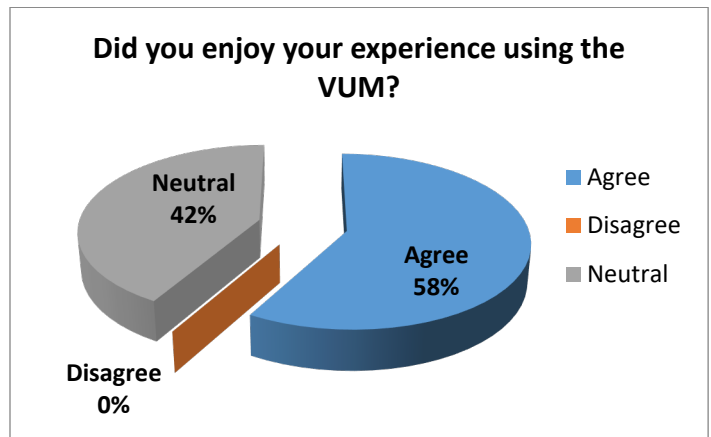


Figure. 49. Shows that the user experience of using VUM was enjoyable

Figure 49 Shows that 58% of the participants enjoyed their VUM experience, 42% were neutral, and 0% disagreed. The result suggests that the system provides an enjoyable experience.

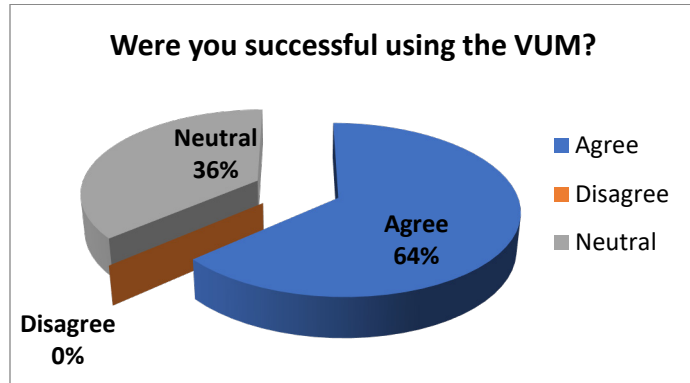


Figure. 50. Shows the success rate of using the VUM

The above chart shows 64% of the participants were successful in using the VUM while 36% were impartial and non disagreed. The chart suggests that the VUM system is a success.

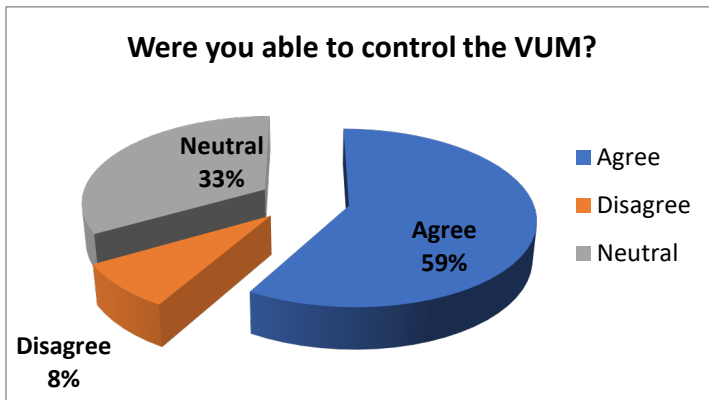


Figure. 51. Shows a high rate of users controlling the VUM

The chart demonstrates that 59% of the participants were able to control the VUM system, whereas 33% were neutral and 8% disagreed. This shows that although a high percentage was able to control it, a minority found it difficult. The result suggests more improvement is required to achieve a higher control success rate.

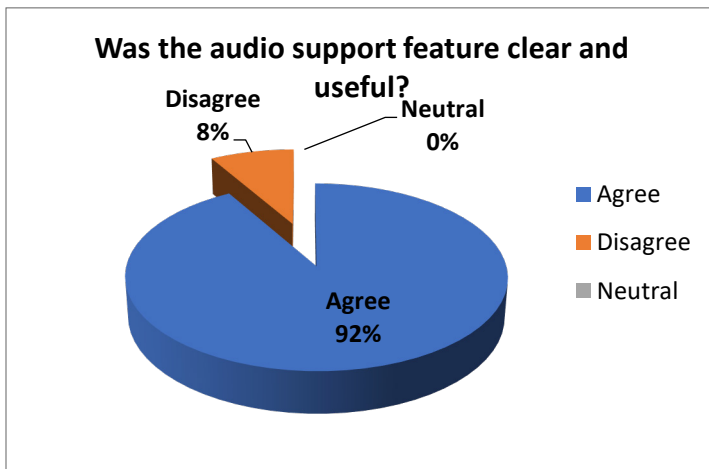


Figure. 52. Shows the success rate of applying the audio feature into VUM

Figure 52 shows that 92% of the participants found that using the support of the audio feature is clear and useful, whereas only 8% disagreed. This is strong evidence to support the excellence of audio feature.

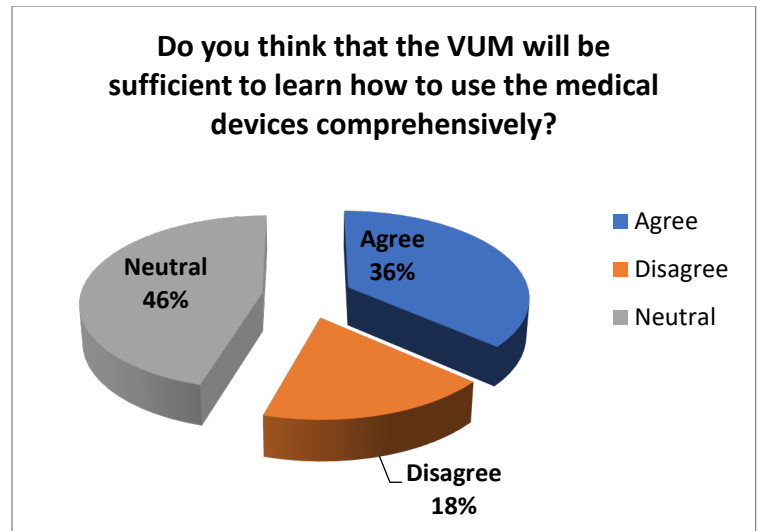


Figure. 53. Shows some users agreed that VUM system is comprehensive

Figure 53 shows that 36% of the participants thought that the VUM system is sufficient and comprehensive, whereas 46% were neutral and 18% disagreed. Although there was a high percentage of neutral responses, a higher percentage agreed rather than disagreed. However, this indicates that there is room for further improvement perhaps by introducing further feature at a later stage.

From the previously illustrated charts and results in relation to all survey's questions, there are two main issues to discuss:

- 1- Aspects of excellence: presented by
 - The service is smooth and free of errors.
 - Users successfully used the VUM system.
 - The audio feature provided was clear and useful.
- 2- Aspects of improvements: presented by
 - The VUM is a sufficient tool to teach users how to use the medical devices comprehensively.

The VUM or the interactive virtual environment of both AquaFlux and Epsilon showed great ease of use, efficiency and practicality. With this virtual environment demonstration at hand and its availability online, clients, trainees, and users who are interested in purchasing one of those skin measurement instruments or wanting to see how they operate by obtaining accurate results. In addition, the VUM demonstration process will help immensely in conveying an accurate thought and a precise sense and enjoyable experience of these two devices and how they operate and what their main functions are. Future ideas, improvements, and upgrades are well on the horizon, Augmented Reality (AR), for example, is a fascinating notion that can support this research and enhance it further. Virtual Reality and AR are opening new ways of interactivity between users and the constructed environments [16]. Moreover, Holograms Technology (3D Hologram) provides a remarkable interactivity surface in team works ambience [17].

The results of the survey strongly suggest that the system is operating effectively although there is room for slight improvements.

8. Conclusions

The objective of the research paper was to introduce a total understanding of an online 3D contents and utilising it in an effective way and not limited to displaying it on a webpage. The main contribution to the field is using an interactive virtual environment to demonstrate the process of using a medical device in conducting a human skin measurement for medical purposes with the aid of audio feature. The new web-based 3D interactive VUM for AquaFlux and Epsilon will replace the old method of illustrating how these medical devices work and what are all their features and functionalities via providing each client with an accessible URL after purchasing the instruments. The web-based virtual system is self-explanatory and easy to use as well as the availability of audio support that guides users step by step throughout the process. Novel holders and futuristic concepts have been introduced to display the capabilities of the new addition and to utilise the marketing process using 3D contents with the interactive virtual world to deliver real novel ideas globally.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

This research work is part of a postgraduate degree. I would like to take this opportunity to thank London South Bank University for their continuous support and giving us all access needed to use, test, and, experiment these two medical instruments. The authors would like to thank the reviewers for their invaluable comments and recommendations.

References

- [1] Al Hashimi, O. and Xiao, P. (2018). Developing a web-based interactive 3D virtual environment for novel skin measurement instruments. *2018 Advances in Science and Engineering Technology International Conferences (ASET)*, Abu Dhabi, 2018, pp. 1-8. doi: 10.1109/ICASET.2018.8376823
- [2] Cellary, W. and Walczak, K. (2012). *Interactive 3D Multimedia Content*. London: Springer London
- [3] VMware Inc, V. (2006). *Virtualization Overview White Paper*. Palo Alto CA: VMware.
- [4] Anon, Instrument Development and Digital Signal Processing in Skin Measurements Zhang Xu School of Engineering London South Bank University Supervisors, Oct 2015.
- [5] Xiao, P. (2014). Biox Epsilon Model E100: Contact Imaging System. Available at: www.biox.biz/home/brochures.php
- [6] Biox (2013). *Epsilon E100 Manual*. London: Biox.
- [7] Ghanbarzadeh, R., Ghapanchi, A.H. & Blumenstein, M., 2014. Application areas of multi-user virtual environments in the healthcare context. *Studies in Health Technology and Informatics*, 204, pp.38–46.
- [8] Shen, W. & Zeng, W., 2011. Research of VR Modeling Technology Based on VRML and 3DSMAX. , pp.487–490.
- [9] Neto, C., 2009. Developing Interaction 3D Models for E-Learning Applications.
- [10] Kotsilieris, T. & Dimopoulou, N., 2013. “ The Evolution of e-Learning in the Context of 3D Virtual Worlds ” Department of Health & Welfare Units Administration, Technological Educational. , 11(2), pp.147–167.
- [11] Potkonjak, V., Gardner, M., Callaghan, V., Mattila, P., Guetl, C., Petrović, V. and Jovanović, K. (2016). Virtual laboratories for education in science, technology, and engineering: A review. *Computers & Education*, 95, pp.309-327.

- [12] Violante, M.G. (2015). Politecnico di Torino Porto Institutional Repository Design and implementation of 3D Web-based interactive.
- [13] Naber, J., Krupitzer, C. and Becker, C. (2017). Transferring an Interactive Display Service to the Virtual Reality. Mannheim: IEEE, pp.1-4.
- [14] Walczak, K., White, M. & Cellary, W., 2004. Building Virtual and Augmented Reality Museum Exhibitions *. , 1(212), pp.135–145.
- [15] Yang, X. et al., 2008. Virtual Reality-Based Robotics Learning System. , (September), pp.859–864.
- [16] Voinea, A., Moldoveanu, F. and Moldoveanu, A. (2017). 3D Model Generation of Human Musculoskeletal System Based on Image Processing. In: *21st International Conference on Control Systems and Computer Science*. Bucharest: IEEE, pp.1-3.
- [17] Wahab, N., Hasbullah, N., Ramli, S. and Zainuddin, N. (2016). Verification of A Battlefield Visualization Framework in Military Decision Making Using Holograms (3D) and Multi-Touch Technology. In: *2016 International Conference on Information and Communication Technology (ICICTM)*. Kuala Lumpur: IEEE, pp.1-2.

PID-Type FLC Controller Design and Tuning for Sensorless Speed Control of DC Motor

Abdullah Y. Al-Maliki^{*1,2}, Kamran Iqbal¹

¹Department of System Engineering, University of Arkansas-Little Rock, 72204, USA

²State Company for Oil Project, Iraqi Ministry of Oil, Iraq

ARTICLE INFO

Article history:

Received: 20 December, 2018

Accepted: 21 December, 2018

Online: 27 December, 2018

Keywords:

DC Motor control

Sensorless speed control

PID

FLC

FLC-PID

Genetic Algorithm

Lookup-Table

ABSTRACT

This article examines the use of non-ideal current and voltage sensors for sensorless speed control for a fixed field DC motor. A PID type speed controller with KF estimator was applied to control the DC motor and IAE, settling time, and peak overshoot were taken as performance indices. However, KF facilitated the noise reduction. After tuning controller gains through MATLAB yielded high peak overshoot as well as IAE with an extended settling time. When we applied, a PID-Type FLC tuned by means of GA (genetic algorithms) caused a 75.98%, 97.89% and 56.2% cut in settling time, maximum overshoot and IAE correspondingly. The FLC-PID fundamentally enhanced sudden load changes disturbance rejection and the reference command speed tracking for the dc motor design in comparison to the conventional PID with no KF. This study was also able to replace the designed FLC-PID with linear lookup-table while achieving the same performance improvements.

1. Introduction

This paper is continuation for our work presented in [1]. DC motors are applied in one way or the other in factories, home appliances, computers to robots, airplanes, and cars. They are more widely used than the related machines, the AC motors, owing to their diverse favorable characteristics. These characteristics some of which are, linear speed control properties and high starting torque. There are more than one types DC motors and all these types have numerous benefits over AC motors which include: less heat production, simpler controllers used, have higher efficiency, can offer precise position control, can produce very close to constant torque and they are easily controllable [2-11]. For that reason, the adoption of DC motors will reduce the amount of energy consumed and improve the efficiency of the machines they are installed. The improvement of DC motors' control arrangement to enhance their response characteristics is one way of achieving these. Is so doing, they will be able to accomplish their work efficiently without the necessarily increasing the capacities of motors alongside their control circuits [12-16].

Researchers have previously done a lot of work to design DC motor control circuits with or without the use of Kalman Filter (KF) and/or Fuzzy Logic Controller (FLC). In [2] the researchers introduced a feedback and feedforward system controllers to

control DC motor. While the feedforward loop was just a gain represented by K_f , the feedback loop was PD controller with gains denoted by K_d and K_p . The gains were optimized using fuzzy logic and genetic algorithm and all were constant. A PC was used in [3] to implement a DC motor PID controller, it was configured by trial and error to attain the desired performance. This study also applied a photo sensor to estimate the motor speed and transmit it to the personal computer. In [17] the researchers came up with a torque estimator for DC motors by means of adaptive Kalman filter consisting of two sections; estimation and extraction. Each of these parts is a DC motor's mathematical model and the first one had one more PID controller to guarantee the accuracy of the estimation. Although it required current and speed sensors, this estimator was designed with no torque sensors. A state feedback controller was developed by the researchers in [4] with the use of Extended Kalman Filter (EKF) for induction motor and KF for DC motor to evaluate the states. A DC motor was controlled using a PID controller in [5], K_p K_i K_d were dynamically modified online by three distinct FLC controllers, one for each gain. KF was utilized for tuning the Membership Functions (MFs) of such FLCs. The FLC controllers, in [7,10], were intended to control DC motors. Microcontrollers was used to implement both FLCs and both made use physical sensors ,that is, no estimators were introduced in these studies. In [9], the researchers made use of a Kalman filter, torque sensor and current sensor to measure the

*Abdullah Al-Maliki, University of Arkansas-Little Rock, ayalmaliki@ualr.edu

speed of a DC motor and thereafter fed it back to a PI controller to enhance the motor's speed estimation. In [11], the DC motor speed was estimated using Kalman filter and was controlled using both Linear-Quadratic Regulator (LQR) and PID controllers. In [18], the authors adjusted the widths of FLC membership functions by means of arithmetic averaging technique and feed forward to run the DC motor without the use of feedback to minimize online computational needs.

The objective of this study is to realize better transient and steady-state responses for a DC motor's speed control (separately excited brushed type) and efficient reference speed tracking irrespective of load and alteration of reference speed. The voltage will only be used as the output of the controller, while the current will not be controlled but restricted within the rated motor current. Then the results from the PID controller and conventional PID controller will be compared with the PID-type FLC controller (FLC-PID) and KF. In addition to the classical time response characteristics; the performance index of IAE is selected to illustrate the achievement in the motor speed response. IAE displays the total error for the entire run period of the simulation to increase the size of the built controller in tracking command speed, notwithstanding its variations and changes in load. In Section 2 below, the DC motor model is developed, Section 3 covers the KF design, Section 4 addresses the different controller designs, and Section 5 carries the conclusion.

2. DC Motor Model

The most common type of DC motors which are broadly applied in robotic and industrial functions is the separately excited DC motor with constant excitation field can be represented by state-space model [4,13]. This model comprises two differential equations representing the mechanical and electrical responses of the DC motor in both static and active operating states. Equations (1) and (2) illustrate this model [4,19].

$$\begin{bmatrix} \dot{I}_a \\ \dot{\omega} \end{bmatrix} = \begin{bmatrix} -\frac{R_a}{L_a} & -\frac{K_b}{L_a} \\ \frac{K_t}{J_r} & -\frac{B_m}{J_r} \end{bmatrix} \cdot \begin{bmatrix} I_a \\ \omega \end{bmatrix} + \begin{bmatrix} \frac{1}{L_a} & 0 \\ 0 & -\frac{1}{J_r} \end{bmatrix} \cdot \begin{bmatrix} V \\ T_{load} \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} I_a \\ \omega \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} I_a \\ \omega \end{bmatrix} \quad (2)$$

The chosen motor specification are listed in Table 1 [20].

The values from Table I are substituted in (1) and (2) and a sampling time (T_s) of 1 millisecond was chosen. Then, using a MATLAB (c2d) function, the continues state-space equations are changed to discrete. The discrete-time state-space motor model is shown in (3) and (4):

$$\begin{bmatrix} i_a(t+1) \\ \omega(t+1) \end{bmatrix} = \begin{bmatrix} 0.9526 & -0.0231 \\ 0.0153 & 0.9998 \end{bmatrix} \cdot \begin{bmatrix} i_a(t) \\ \omega(t) \end{bmatrix} + \begin{bmatrix} 0.0210 & 0.0002 \\ 0.0002 & -0.0143 \end{bmatrix} \cdot \begin{bmatrix} v(t) \\ t_{load}(t) \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} i_a(t) \\ \omega(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} i_a(t) \\ \omega(t) \end{bmatrix} \quad (4)$$

Where (t) represents the current sample, and (t+1) denote the next sample.

Table 1 Chosen motor specifications [20]

Parameter/Specification	Value	Unit
R_a (Armature resistance)	2.25	Ohm
L_a (Armature inductance)	$46.5 \cdot 10^{-3}$	H
K_b (Back EMF constant)	1.1	Volt · sec/rad
K_t (Torque constant)	1.1	N · m/A
J_r (Rotor inertia)	$7 \cdot 10^{-2}$	Kg · m ²
B_m (Mechanical damping factor)	$2 \cdot 10^{-3}$	N · m · sec/rad
V_R (Rated Voltage)	220	V
T_R (Rated Torque)	4.7495	N · m
I_R (Rated current)	4.8	A
ω_R (Rated angular speed)	157.08	rad/sec
P_R (Rated power)	1	HP

3. Kalman Filter Observer

As mentioned in [21,22] Kalman filter is considered as optimum observer for the state variables and the outputs of linear time-invariant (LTI) systems. It has two significant characteristics which made it our choice to estimate the states of the DC motor under study; one it is able to exploit the already known state-space model of the DC motor, and the second reason because it has the capacity to filter diverse sensor noises [23]. The Kalman filter system discrete equations can be broken into two categories and are illustrated in (5)-(10) [23].

A. Time update equations

$$\hat{x}_t = A \cdot x_{t-1} + B \cdot u_t + W_t \quad (5)$$

$$\hat{P}_t = A \cdot P_{t-1} \cdot A^T + Q_t \quad (6)$$

B. Measurement update equations

$$K_t = \frac{\hat{P}_t \cdot H^T}{H \cdot \hat{P}_t \cdot H^T + R} \quad (7)$$

$$y_t = C_{KF} \cdot x_{m_t} + z_k \quad (8)$$

$$x_t = \hat{x}_t + K_t \cdot (y_t - H \cdot \hat{x}_t) \quad (9)$$

$$P_t = (I - K_t \cdot H) \cdot \hat{P}_t \quad (10)$$

In which \hat{x}_t is the estimated current state vector; \hat{P}_t is the estimated process error covariance matrix; W_t is the estimated noise matrix; Q_t is the system noise covariance matrix; K_t is Kalman gain; H: is the transformation matrix; R is sensor noise covariance matrix; y_t is measurement of the state; C_{KF} is Kalman filter C matrix for state-space model of the system under study (in this case the DC motor); x_{m_t} is the measured states vector; z_k is measurement noise; x_t is corrected state vector; and, P_t is corrected process error covariance matrix

Kalman filter estimates the states of present system and the error in such estimates by means of the time update equations. It then rectifies these errors using the measurement update calculations based on the Kalman gain and the actual inaccurate measurement. This done in a repetitive way for all sample times (T_s) or time steps [23]. Let $\Delta\omega$ and ΔI_a represent the error in speed and current calculation/measurement correspondingly.

Then accurate T_{load} computation needs the integration of ω and $d\omega/dt$ as illustrated in (11), which is solving the second row of (1) for T_{load} :

$$T_{load} = -\frac{d\omega}{dt}J_r + K_t \cdot I_a - B_m \cdot \omega \quad (11)$$

As $|K_t \cdot I_a| \gg |-B_m \cdot \omega|$ and $|K_t \cdot I_a| \gg \left|-\frac{d\omega}{dt}J_r\right|$, and the term $-\frac{d\omega}{dt}J_r$ has noise because of the differentiation, The approximate formula below can be used to solve T_{load} :

$$T_{load} \cong K_t \cdot I_a \quad (12)$$

Equation (13) is derived by solving the first row of (3) for ω , is used to calculate the motor speed:

$$\omega = -\frac{L_a}{K_b} \frac{dI_a}{dt} - \frac{R_a}{K_b} I_a + \frac{1}{K_b} V \quad (13)$$

Because $\left|-\frac{R_a}{K_b} I_a + \frac{1}{K_b} V\right| \gg \left|-\frac{L_a}{K_b} \frac{dI_a}{dt}\right|$, ω can be calculated with high enough accuracy using (14) below:

$$\omega \cong -\frac{R_a}{K_b} I_a + \frac{1}{K_b} V \quad (14)$$

The sensors and process noise covariance matrices are articulated in terms of the stochastic characteristics of their respective noises and they can be either static or dynamic [24]. Q elements were found to be directly proportional to the inverse of the estimated armature current and motor speed and therefore was chosen to be dynamic as shown in (15). Since the sensor error is fixed by the manufacturer in its datasheet, R was selected to be static. In (16) the calculations of R are shown and the datasheet of a sample current sensor is found in [25].

$$Q = \begin{bmatrix} \frac{1}{I_a^2} & \frac{1}{I_a} \cdot \frac{1}{\omega^2} \\ \frac{1}{I_a} \cdot \frac{1}{\omega^2} & \frac{1}{\omega^2} \end{bmatrix} \quad (15)$$

$$R = \begin{bmatrix} 2.25 & 0 \\ 0 & 6.25 \end{bmatrix} \quad (16)$$

4. Controller Design, Simulation and Results

In this section, three potential designs of a controller for DC motor speed control are discussed. Figure 1 shows the general system architecture under study.

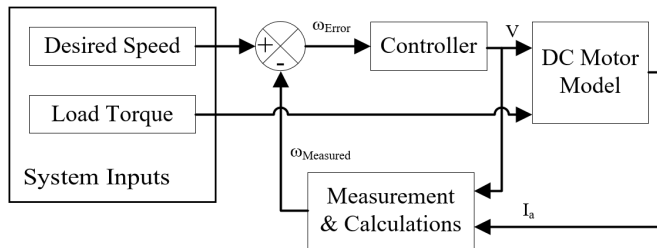


Figure 1: The DC motor with the PID controller and speed feedback

4.1. Discrete PID Controller

Initially, discrete PID was used to control the DC motor. The transfer function of the controller is presented in (17). Figure 2 shows the command speed signal to the motor, the motor will experience an uncontrolled torque load variation illustrated in Figure 3. In the rest of this work, Figure 2 and Figure 3 will be considered as the running conditions for the DC motor under study.

Using MATLAB/Simulink PID tuner with “design focus” option set to “Reference tracking”, a discrete PID controller was

tuned to control the motor’s speed by means of changing its supply voltage. The gains after MATLAB/Simulink PID tuner are: $K_P=2.51$, $K_I=9.724$, $K_D=-0.19185$ and $N=12.89$.

$$U(Z) = K_P + K_I \cdot T_s \cdot \left(\frac{1}{z-1}\right) + K_D \cdot \left(\frac{N}{1+N \cdot T_s \cdot \left(\frac{1}{z-1}\right)}\right) \quad (17)$$

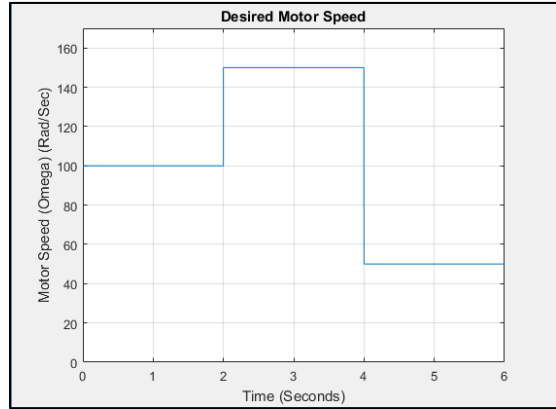


Figure 2 Desired motor speed profile

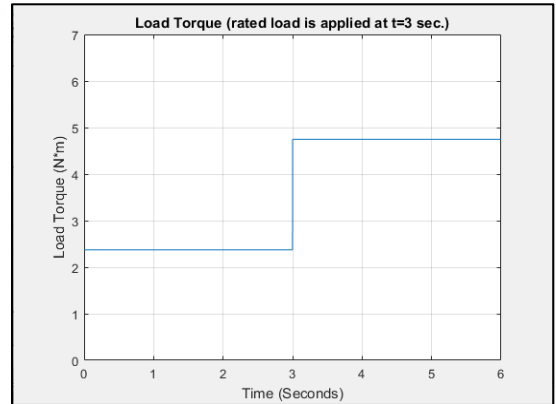


Figure 3 Load torque (disturbance) profile

To simulate current and voltage sensors precision tolerances, arbitrary noise of 2% was added to voltage and current [25]. The voltage (controller output) was kept at or below the nameplate voltage that is 220 V, and this limitation is kept for the FLC PID in part C and throughout this work. The measurement of current was restricted certain upper limit to mimic the current sensor magnetic limitation characteristics. In the measurement block, the speed is determined using (13) only and no more conditioning is required. Figure 4 demonstrate the system response in which the system experienced an IAE of 47.52, settling time (t_s) of 1.08 second and peak overshoot of 25.7%. Using (14) to obtain ω in the subsequent run lowered the peak overshoot to 23.7% and lowered the IAE a little making it 45.89. The settling time experienced a slight fall to 1.07 seconds. Such a small enhancement in system response is associated to the cancelation of the noisy derivative section in (13). The measured and actual motor speed using (13) and with no Kalman filter are displayed in Figure 5.

4.2. PID with KF

Kalman filter was applied in the next run in an effort to predict the motor’s speed, Figure 6 shows the whole system. When Kalman filter was used and everything else as from the previous run was unchanged, Figure 7 displays clear falls in the settling time, IAE and peak overshoot which are now in order, 0.635, 36.71

and 8.01%. The KF estimated speed of the motor vs the actual and measured speeds are shown in Figure 8.

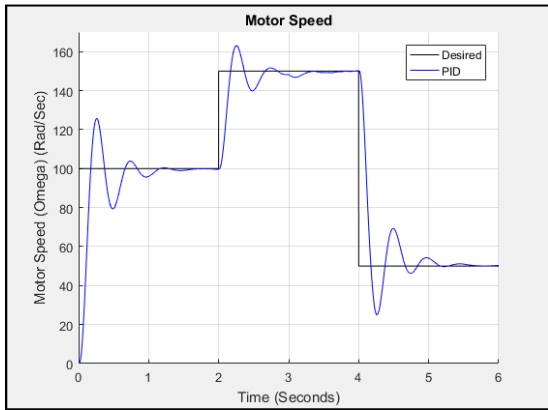


Figure 4 motor speed response with PID and unit feedback

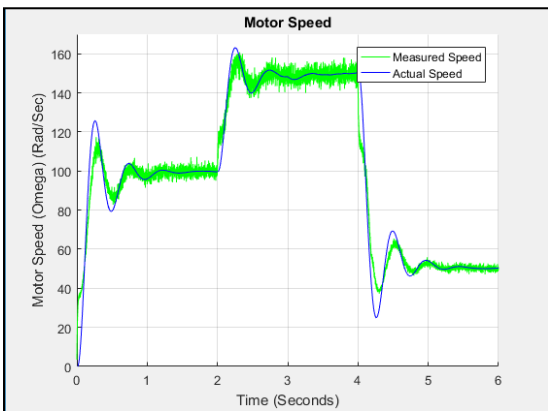


Figure 5 actual motor speed and measured motor speed with added noise

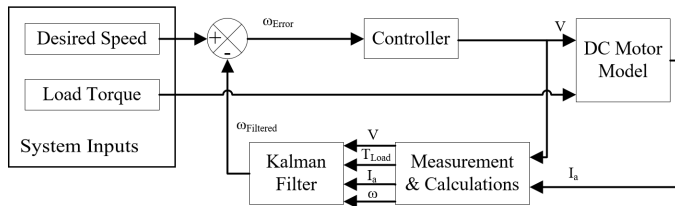


Figure 6 The DC motor with the PID controller and Kalman filter

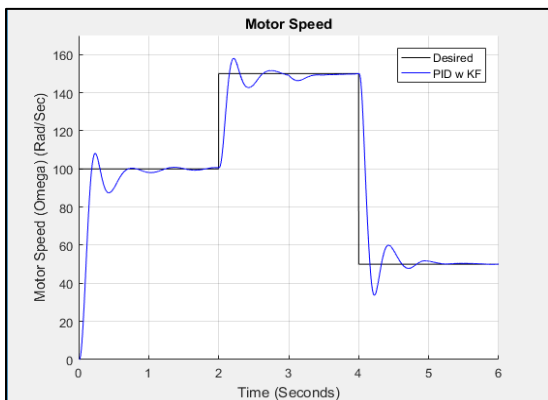


Figure 7 motor speed response with PID and Kalman filter estimator

While it showed some improvement in the system response, the PID with the help of Kalman filter was unable to fully remove

the overshoot of the motor’s speed response which introduced unwanted ringing in the motor’s speed. This limitation can be overcome by designing a PID-type fuzzy logic controller with PID properties and the additional variable gains throughout the full range of the controller’s operation [26,27].

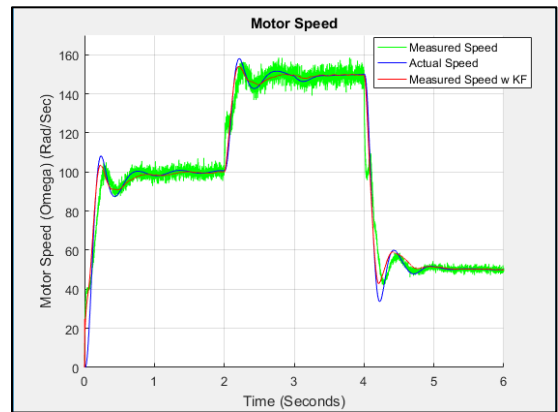


Figure 8 actual motor speed and measured motor speed with added noise with and without Kalman filter

4.3. FLC-PID Controller

A Takagi-Sugeno type FLC-PID controller is chosen in this work as it is “more diverse in gain variation characteristics” than the Mamdani type FLC-PID controller [26]. Aside from few changes in the rule base; the main difference in design between Takagi-Sugeno type FLC-PID controller developed in this work and the one utilized in [28,29] is the different and distinct gains for PI and PD controllers, offering an additional degree of design freedom. Figure 9 represents the block diagram of the FLC-PID developed for this work.

For both controllers (FLC-PI and FLC-PD), the input and output signal ranges were normalized to [-1,1]. Doing so will enable this FLC-PID controller to be applied for various motor sizes by just tuning its gains. Figure 10 shows the MF’s for error as well as rate of change of error, Gaussian type was chosen to guarantee smooth control action. The MF’s for the controller’s output are presented in Figure 11 which are uni-valued MF’s. one benefit of using FLC controller is that we can make all the equivalent MF’s similar for both FLC-PD and FLC-PI, and the unique action variation between the controllers is achieved only by altering their rule base. Which is what we did in this work.

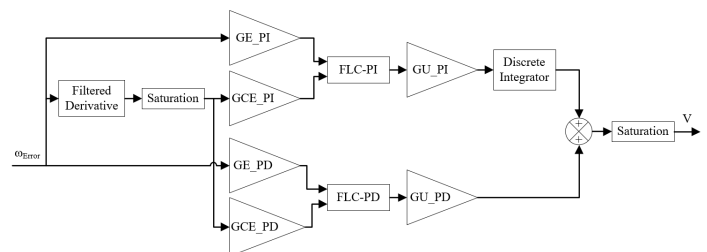


Figure 9 FLC-PID block diagram

Where GE_PI is FLC-PD speed error gain; GCE_PI is FLC-PD change of rate of speed error gain; GU_PI is FLC-PD output gain; GE_PD is FLC-PD speed error gain; GCE_PD is FLC-PD change of rate of speed error gain; and, GU_PD is FLC-PD output gain.

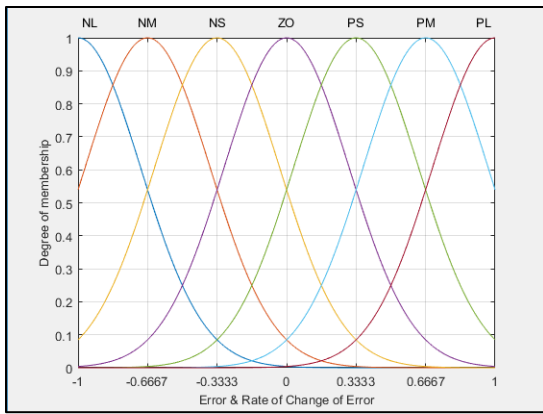


Figure 10 input membership functions for FLC-PI and FLC-PD

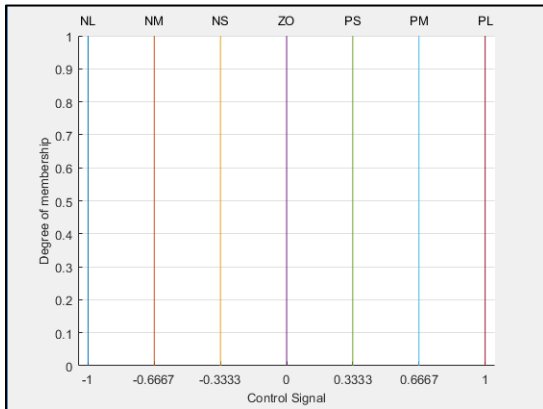


Figure 11 output membership functions for FLC-PI and FLC-PD

The rule base for FLC-PD and FLC-PI are shown by Tables 2 and 3 respectively. The FLC-PID applies these rules in the arrangement illustrated in (17) [30] to produce the control surfaces presented in Figure 12 and 13. Such rules simulate the performance of the common PI and PD controllers to develop FLC-PI and FLC-PD controllers in that order.

Table 2 FLC-PD rule base

Control Output UPD		Error (e)						
		NL	NM	NS	ZO	PS	PM	PL
Rate of Change of Error (ce)	PL	PS	PS	PM	PL	PL	PL	PL
	PM	PS	PS	PS	PS	PL	PL	PL
	PS	PM	PS	PS	PS	PM	PL	PL
	ZO	NL	NM	PS	ZO	PS	PM	PL
	NS	NL	NL	NM	NS	PS	PS	PM
	NM	NL	NL	NL	NS	PS	PS	PS
	NL	NL	NL	NL	NS	PS	PS	PS

Table 3 FLC-PI rule base

Control Output UPI		Error (e)						
		NL	NM	NS	ZO	PS	PM	PL
Rate of Change of Error (ce)	PL	ZO	PS	PL	PL	PL	PL	PL
	PM	NS	ZO	PS	PM	PL	PL	PL
	PS	NL	NS	ZO	PS	PM	PL	PL
	ZO	NL	NM	NS	ZO	PS	PM	PL
	NS	NL	NL	NM	NS	ZO	PS	PL
	NM	NL	NL	NL	NM	NS	ZO	PS
	NL	NL	NL	NL	NL	NL	NS	ZO

Where PL is Positive Large; PM is Positive Medium; PS is Positive Small; ZO is Zero; NS is Negative Small; NM is Negative Medium; and NL is Negative Large.

$$\text{if } \underbrace{\text{Error}}_{\substack{\text{1st} \\ \text{input} \\ \text{MF}}} = \underbrace{\text{NL}}_{\substack{\text{AND} \\ \text{Fuzzy} \\ \text{Logic} \\ \text{Operator}}} \text{ AND } \underbrace{\text{Rate of Change of Error}}_{\substack{\text{2nd} \\ \text{input} \\ \text{MF}}} = \underbrace{\text{PL}}_{\substack{\text{2nd} \\ \text{input} \\ \text{MF}}} \text{ Then } \underbrace{\text{UPD}}_{\substack{\text{Output} \\ \text{MF}}} = \underbrace{\text{PS}}_{\substack{\text{Output} \\ \text{MF}}} \quad (17)$$

Antecedent Consequent

When we have (17), Tables 2 and 3, Figure 10, 11, 12 and 13 into perspective; it would be obvious that both FLC have more than 49 equations and more than 10 parameters (comprising membership functions parameters) to adapt its output. These add to diverse design flexibility for the preferred control surface that the common PID controller simply cannot provide.

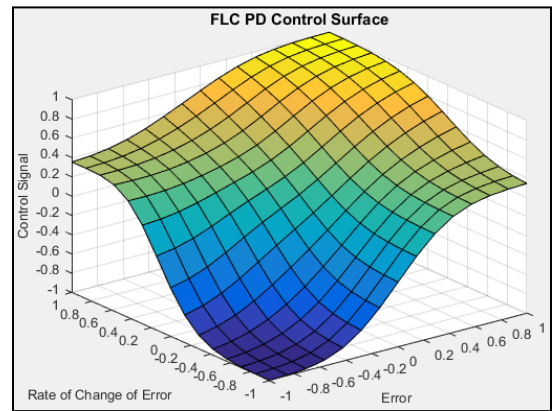


Figure 12 FLC-PD control surface

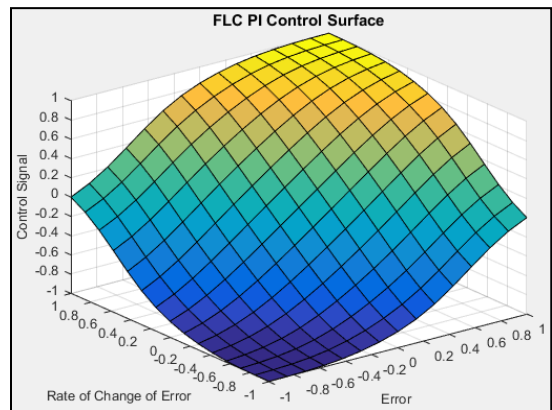


Figure 13 FLC-PI control surface

The next five steps briefly describe how an FLC works [30,31,32]:

1. The input signals are mapped to their corresponding MF's and assign degree of membership ranging between [0,1] for each input signal. (Fuzzification stage)
2. Finding the degree of firing of each rule by applying fuzzy logic operation (e.g. AND, OR, ... etc.) to the antecedents. (Inference mechanism matching stage 1).
3. FLC determines which rule is to be fired by checking the result from step 2, each rule having a nonzero result will be fired. (Inference mechanism matching stage 2).
4. Multiply the consequent of each rule by its firing degree (the corresponding result from step 3). (Inference setup/aggregation stage).

- Apply the selected defuzzification method (for this work weighted average method was selected, which is shown in equation 18). (Defuzzification stage).

$$u = \frac{\sum_{n=1}^{\text{no.of fired rules}} \mu_{\text{rule}_n} \cdot U_{\text{rule}_n}}{\sum_{n=1}^{\text{no.of fired rules}} \mu_{\text{rule}_n}} \quad (18)$$

where:

u: is FLC output

μ_{rule_n} : is the firing degree for each rule (step 2 above)

U_{rule_n} : is the consequent of rule_n.

The selection of FLC-PID gains was informed by the following: because the rated speed of the motor is 157.08 rad/sec, speed error gains were calculated as $\frac{1}{157.08} = 0.006367$. A range of [-1000,1000] was taken for derivative of error to neutralize high values due to the derivative term; consequently, it's gain was chosen to be $\frac{1}{1000} = 0.001$. Lastly, the output gains is set to be $1.5 \times$ rated voltage to put an accelerated control action to the motor's speed. FLC-PID gains first estimation may as well be acquired from an adjusted PID as stated in [28].

4.4. Gains' Value Optimization Using GA

Genetic algorithm (GA) is an evolutionary algorithm; it uses the process of natural selection to solve optimization problems; it works to find the fittest value in the range of search regardless of that range's nonlinearity. Simple GA can be summarized in the following steps [33]:

- Convert the variable values into binary. (Initialization 1)
- Initiate random population of 50 individual per variable. (Initialization 2)
- Check the fitness of each individual. (Fitness 1)
- Select the best-fitted individuals. (Fitness 2)
- Generate new individuals (children) by combining two individuals (parents) from step 4. (Crossover)
- Generate new individuals (children) by making random changes to individuals (parents) from step 4. (Mutation)
- Replace the current population with the children from steps 4 and 5. (Initiate new generation)
- Repeat from step 3 until at least one stopping condition (number of generations, the value of the relative change in the fitness function, ... etc.) is satisfied.

Table 4 tuned controller gains

Gain	Value
GE_PI	0.1
GCE_PI	$8.51 \cdot 10^{-4}$
GU_PI	225
GE_PD	$2.43 \cdot 10^{-2}$
GCE_PD	$1.134 \cdot 10^{-3}$
GU_PD	300

MATLAB optimization tool with built-in GA optimizer has been used in this project to fine tune the FLC-PID gains. The selected fitness function was the IAE of the speed, to ensure that the tuning is aimed at faster and improved speed command tracking with an all-round load change rejection. Because GA is a stochastic search tool, various runs may produce varying results,

but each result must satisfy a minimum of one stopping condition. Indeed, the remaining runs with various gains can result in similar system response. Table 4 shows the FLC-PID gains upon tuning.

The FLC-PID successfully removed the peak overshoot (less than 0.5%), decreased the settling time to 0.257 seconds and reduced the IAE to only 20.1. To make this clearer, Table 4 illustrates a percentage contrast between FLC-PID with KF, common PID without KF and PID with KF, choosing common PID without KF as a 100% base. Figure 14 is a graph of the motor speed response for all scenarios mentioned above versus the desired speed. It is worth noting that lower value for every performance index used in this paper means improved system response.

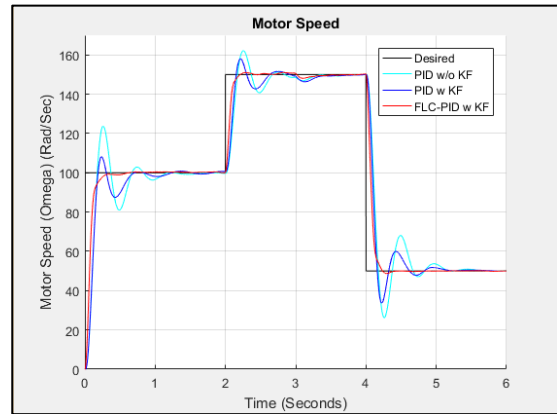


Figure 14 compare motor speed for all three controllers used in this work

Table 4 comparison of system response for all three controllers used in this work

Performance Indices	PID W/O KF	PID With KF	FLC-PID With KF
Settling time (seconds)	1.07 100%	0.635 59.35%	0.257 24.02%
Peak Overshoot (radians)	23.7 100%	8.01 33.8%	<0.5 <2.11%
IAE	45.89 100%	36.71 80%	20.1 43.8%

4.5. Converting FLC-PID to lookup Table

In order to reduce the computational requirement of FCL-PID to a real word controller e.g. microcontroller or field programmable gate array (FPGA), thus making it easier for the controller to response in real time, the FLC-PI and FLC-PID control surfaces will be stored in a linear lookup-Table format. This can be done by dividing the two inputs of both FLC controllers (FLC-PI and FLC-PD) into 21 breakpoints (since both inputs range is normalized to [-1,1], 21 breakpoints mean 0.1 spacing) and evaluate the FLC controllers at each of the breakpoints. The total evaluations will be $21 \times 21 = 441$ for each FLC controller. A linear interpolation based lookup Table was put in place of each FLC controller in the controller block as shown in Figure 15, and they were loaded with the corresponding 441 evaluations. The motor speed response with FLC-PID and lookup Table FLC-PID are shown in Figure 16. Both responses are actually overlapping as the difference between them is minimal. A zoomed view of the speed response from the time 0.3to 0.7 seconds is shown in Figure 17 to magnify the difference between these responses, which has a maximum value of less than 1%. As a

result, the DC motor with the lookup Table had almost identical performance indices (the difference is less than 0.1%) as compared to the response with the actual FLC-PID controller.

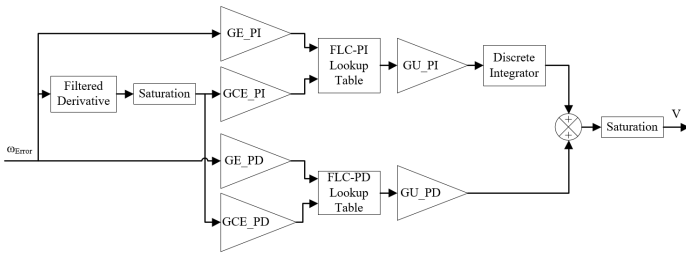


Figure 15 lookup table FLC-PID block diagram

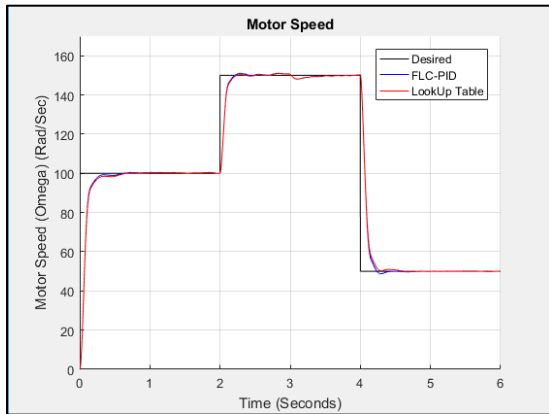


Figure 16 lookup table FLC-PID versus actual FLC-PID speed response

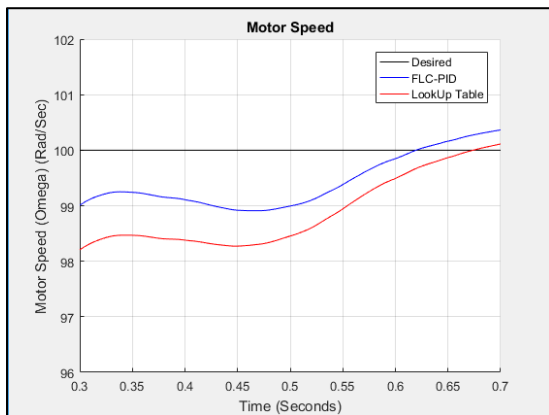


Figure 17 zoomed view of lookup Table FLC-PID versus actual FLC-PID speed response

5. Conclusion

This work investigated the sensorless speed control of fixed field DC motor in a noisy environment mimicking real-live operating environment. As a result of noise in this environment, a common PID controller yielded high peak overshoot from the reference and took significant time to settle. That was true even after removing the noisy derivative terms in T_{load} and ω in their respective estimation equations. When using Kalman filter, the PID controller was able to lower the peak overshoot by good margin, nonetheless it stayed there with long ringing in the speed response. To minimize these undesirable system behaviors, an FLC-PID controller was applied to the motor. This controller was tuned by GA. This new controller design end up reducing the peak

overshoot to less than 1% and the settling time by 75.98%. The IAE was also reduced to only 56.2% of what it was. Finally, the FLC-PID was converted into linear interpolation-based lookup-Table to reduce the computational complexity of the FLC-PID and make it easier to realize in real world applications while maintaining the same response as with the actual FLC-PID. Achieving these enhanced time response and speed tracking improvement will ensure smoother, more robust and more power saving operation of the motor. The FLC-PID gave the DC motor “immunity” to reference speed and load changes as long as these changes are within its operating capacity.

In the future we intend to implement the lookup-Table by using a specialized lookup-Table Integrated Circuit (IC) or a generic Field Programmable Gate Array (FPGA) IC and to connect it to a physical separately excited DC motor in an experimental setup. Variable load, variable reference speed and current limiting protection must be included in this setup. This implementation will require retuning the FLC-PID gains, but in return it will solidly verify the proposed controller.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] A. Y. Al-Maliki and K. Iqbal, “FLC-based PID controller tuning for sensorless speed control of DC motor;” Proc. IEEE Int. Conf. Ind. Technol., vol. 2018–February, pp. 169–174, 2018.
- [2] S. Chowdhuri and A. Mukherjee, “An evolutionary approach to optimize speed controller of DC machines;” Proc. IEEE Int. Conf. Ind. Technol. 2000 (IEEE Cat. No.00TH8482), vol. 2, pp. 682–687, 2000.
- [3] G. Huang and S. Lee, “PC-based PID speed control in DC motor;” 2008 Int. Conf. Audio, Lang. Image Process., pp. 400–407, 2008.
- [4] G. G. and P. Siano, “Sensorless Control of Electric Motors with Kalman Filters: Applications to Robotic and Industrial Systems;” Int. J. Adv. Robot. Syst., vol. 8, no. 6, p. 1, 2011.
- [5] M. Shadkam, H. Mojallali, and Y. Bostani, “Speed Control of DC Motor Using Extended Kalman FilterBased Fuzzy PID;” Int. J. Inf. Electron. Eng., vol. 3, no. 9, pp. 1209–1220, 2013.
- [6] J. Jacob, S. S. Alex, and A. E. Daniel, “Speed Control of Brushless DC Motor Implementing Extended Kalman Filter;” Int. J. Eng. Innov. Technol., vol. 3, no. 1, pp. 305–309, 2013.
- [7] A. A. Thorat, S. Yadav, and S. S. Patil, “Implementation of Fuzzy Logic System for DC Motor Speed Control using Microcontroller;” J. Eng. Res. Appl., vol. 3, no. 2, pp. 950–956, 2013.
- [8] J. N. Rai and M. Singhal, “Speed Control of Dc Motor Using Fuzzy Logic Technique;” IOSR J. Electr. Electron. Eng., vol. 3, no. 6, pp. 41–48, 2012.
- [9] P. Deshpande and A. Deshpande, “Inferential control of DC motor using Kalman Filter;” in 2012 2nd International Conference on Power, Control and Embedded Systems, 2012, pp. 1–5.
- [10] H. R. Jayetileke, W. R. De Mel, and H. U. W. Ratnayake, “Real-time fuzzy logic speed tracking controller for a DC motor using Arduino Due;” 2014 7th Int. Conf. Inf. Autom. Sustain. “Sharpening Futur. with Sustain. Technol. ICIAfS 2014, 2014.
- [11] T. Abut, “Modeling and Optimal Control of a DC Motor;” Int. J. Eng. Trends Technol., vol. 32, no. 3, pp. 146–150, Feb. 2016.
- [12] A. O. Al-Jazaeri, L. Samaranayake, S. Longo, and D. J. Auger, “Fuzzy Logic Control for energy saving in Autonomous Electric Vehicles;” in 2014 IEEE International Electric Vehicle Conference (IEVC), 2014, pp. 1–6.
- [13] A. Watanabe, S. Yuta, and A. Ohya, “A new method for efficient drive and current control of small-sized brushless DC motor: Experiments and its evaluation;” in IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society, 2010, pp. 735–741.
- [14] L. Varghese and J. T. Kuncheria, “Modelling and design of cost efficient novel digital controller for brushless DC motor drive;” in 2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), 2014, pp. 1–5.
- [15] Y.-P. Yang and M.-T. Peng, “A Surface-Mounted Permanent-Magnet Motor With Sinusoidal Pulsewidth-Modulation-Shaped Magnets;” IEEE Trans.

Magn., vol. PP, pp. 1–8, 2018.

- [16] S. Noguchi and H. Dohmeki, "Improvement of efficiency and vibration noise characteristics depending on excitation waveform of a brushless DC motor," 2018 IEEE Int. Magn. Conf., pp. 1–5, 2018.
- [17] S. Lee and H. Ahn, "Sensorless torque estimation using adaptive Kalman filter and disturbance estimator," in Proceedings of 2010 IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications, 2010, pp. 87–92.
- [18] J. K. Satapathy, S. Das, C. J. Harris, and P. Misra, "Indirect tuning of membership function in a fixed fuzzy structure for efficient control of a DC drive system," in SMC 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics. "Cybernetics Evolving to Systems, Humans, Organizations, and their Complex Interactions" (Cat. No.00CH37166), 2000, vol. 5, pp. 3790–3793.
- [19] S. P. Paul C. Krause, Oleg Wasynczuk, Scott D. Sudhoff, Analysis of Electric Machinery and Drive Systems, 3rd ed. Piscataway, NJ, USA: John Wiley & Sons, Inc., 2013.
- [20] F. Rashidi, M. Rashidi, and A. Hashemi-Hosseini, "Speed regulation of DC motors using intelligent controllers," Control Appl., pp. 925–930, 2003.
- [21] F. L. Lewis, L. Xie, and D. Popa, Optimal and Robust Estimation With an Introduction to Stochastic Control Theory, Second Edi. CRC press, 2008.
- [22] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., vol. 82, no. 1, p. 35, 1960.
- [23] R. G. Brown and P. Y. C. Hwang, Introduction to Random Signals and Applied Kalman Filtering WITH MATLAB EXERCISES, FOURTH EDI. John Wiley & Sons, Inc., 2012.
- [24] Texas Instruments and T. I. Europe, "Sensorless Control with Kalman Filter on TMS320 Fixed-Point DSP," Control, no. July, 1997.
- [25] "Split Core Current Transformer ECS1030-L72." ECHUN Electronic Co., Ltd, Dongguan City. Guang Dong. China, pp. 1–3, 2017.
- [26] H. Ying, "Constructing nonlinear variable gain controllers via the Takagi-Sugeno fuzzy control," IEEE Trans. Fuzzy Syst., vol. 6, no. 2, pp. 226–234, 1998.
- [27] B. M. Mohan and A. Sinha, "Analytical structure and stability analysis of a fuzzy PID controller," Appl. Soft Comput., vol. 8, no. 1, pp. 749–758, 2008.
- [28] A. Noshadi, J. Shi, W. Lee, and A. Kalam, "PID-type fuzzy logic controller for active magnetic bearing system," in IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society, 2014, vol. 27, no. 7, pp. 241–247.
- [29] J. X. Xu, C. C. Hang, and C. Liu, "Parallel structure and tuning of a fuzzy PID controller," Automatica, vol. 36, no. 5, pp. 673–684, 2000.
- [30] K. M. Passino and S. Yurkovich, Fuzzy Control. Addison Wesley Longman, Inc., 1998.
- [31] G. S. Sandhu, T. Brehm, and K. S. Rattan, "Analysis and design of a proportional plus derivative fuzzy logic controller," in Proceedings of the IEEE 1996 National Aerospace and Electronics Conference NAECON 1996, 1996, vol. 1, pp. 397–404.
- [32] S. Z. Hassan, H. Li, T. Kamal, F. Mehmood, "Fuzzy embedded MPPT modeling and control of PV system in a hybrid power system" in 12th International IEEE Conference on Emerging Technologies (ICET) 18-19-October 2016
- [33] D. A. Coley, An Introduction to Genetic Algorithms for Scientists and Engineers. WORLD SCIENTIFIC, 1999.