

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339774416>

Realistic 3D Virtual Fitting Room

Thesis · March 2020

DOI: 10.13140/RG.2.2.23415.98725

CITATIONS

0

READS

5

1 author:



Morteza Daneshmand

University of Tartu

37 PUBLICATIONS 144 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Virtual Reality [View project](#)

MORTEZA DANESHMAND

Realistic 3D Virtual Fitting Room



MORTEZA DANESHMAND

Realistic 3D Virtual Fitting Room



UNIVERSITY OF TARTU
Press

Institute of Technology, Faculty of Science and Technology, University of Tartu,
Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (Ph.D.) on April 17, 2018 by the Council of the Institute of Technology, University of Tartu.

Supervisors:

Prof. Ph.D. Alvo Aabloo
University of Tartu
Tartu, Estonia

Prof. Ph.D. Gholamreza Anbarjafari
University of Tartu
Tartu, Estonia

Reviewer:

Assoc. Prof. Ph.D. Mark Fishel
University of Tartu
Tartu, Estonia

Opponent:

Prof. Ph.D. Ahmet Kondož
Loughborough University London
London, United Kingdom

The public defense will take place on May 24, 2018 at Nooruse 1 - 121, Tartu, Estonia.

The publication of this dissertation was financed by Institute of Technology, University of Tartu.

ISSN 2228-0855

ISBN 978-9949-77-720-4 (print)

ISBN 978-9949-77-721-1 (pdf)

Copyright: Morteza Daneshmand, 2018

University of Tartu Press

www.tyk.ee

Contents

List of Publications	8
Abstract	12
1 Introduction	18
2 Overview of Virtual Fitting based on Mannequin Robots	21
2.1 Introduction	21
2.2 Overview of the Existing Online Fitting Solutions	26
2.3 The Modified Version of the Virtual Fitting System	28
2.4 Discussion and Ongoing Research	33
2.5 Conclusion	35
3 Real-time Automatic Actuator Position Dictionary Generation for Shape-changing Robots	36
3.1 Introduction	36
3.2 Description of the Robot	37
3.3 The Proposed Method	40
3.4 The Experimental Results and Discussion	44
3.5 Conclusion	46
4 A Survey of 3D Scanning	47
4.1 Introduction	47
4.2 Applications of Photogrammetry	50
4.3 Close-Range Photogrammetry (CRP)	50
4.3.1 Calibrating Cameras	50
4.3.2 Detecting and Matching Lines	51
4.3.3 Monitoring Forests	51
4.3.4 Sardinia Radio Telescope (SRT) Simulation	51
4.3.5 River Surface Topography	51
4.3.6 Modeling Cities	52

4.4	Aerial Photogrammetry	52
4.4.1	The International Society for Photogrammetry and Remote Sensing (RS) (ISPRS) Benchmark	52
4.4.2	Extracting and Matching Features	52
4.4.3	Historical Heritage Monitoring	53
4.4.4	Ground-level Modeling	53
4.4.5	Monitoring Topography	53
4.4.6	Fault Plane Detection	53
4.5	Structure from Motion (SfM)-based Photogrammetry	54
4.6	Terrestrial Photogrammetry	54
4.7	Typical Applications of Laser Scanning (LS)	55
4.8	Mobile LS (MLS)	55
4.9	Terrestrial LS (TLS)	56
4.10	Airborne LS (ALS)	57
4.11	Phase Comparison, Structured Light and Time of Flight (ToF)	57
4.12	Comparison of 3D Scanning Techniques	59
4.12.1	Comparison of 3D Scanning Devices	59
4.12.2	Comparison of ALS and Photogrammetry	59
4.12.3	Comparison of Photogrammetric 3D Modeling Software	59
4.12.4	Analyzing Vegetated Areas and Forests	60
4.12.5	Comparison of Digital Elevation Model (DEM) Methods	60
4.13	Combinational systems	61
4.14	Post-processing	63
4.15	Conclusion	64
5	3D Size-Estimation based on the Geodesic Distance Measured by Photogrammetric Devices	65
5.1	Introduction	65
5.2	The Proposed Size Estimation Algorithm	70
5.3	The Preliminary Tests	73
5.4	The Experiments, Results and Discussion	74
5.5	Conclusion	76
6	3D Object and Scene Reconstruction	78
6.1	Introduction	78
6.2	The Hardware Setup	84
6.3	The Proposed Method	86
6.4	The Experimental Results and Discussion	97
6.5	Conclusion	104

7	3D Garment Reconstruction	107
7.1	Introduction	107
7.2	The Experimental Setup	109
7.3	Segmentation	110
7.4	Point Cloud Post-processing	111
7.4.1	Hole Filling based on Statistical Analysis	111
7.4.2	Point Cloud Smoothing and Surface Approximation	115
7.4.3	Removing Extra Points	119
7.5	Converting the Processed Point Cloud to a Mesh	122
7.6	Approximating Nonexistent Garment Size Categories	127
7.7	Conclusion	129
	Conclusion and Future Work	130
	Bibliography	132
	Acknowledgments	164
	Kokkuvõte (Summary in Estonian)	165
	Appendices	167
	Curriculum Vitae	225
	Elulookirjeldus	230

PUBLICATIONS INCLUDED IN THIS THESIS

Journal Papers

1. Daneshmand, M., Helmi, A., Avots, E., Noroozi, F., Alisinanoglu, F., Sait Arslan, H., Gorbova, J., Haamer, R.E., Ozcinar, C., Anbarjafari, G.: 3D Scanning: A Comprehensive Survey. ArXiv e-prints (Jan 2018).
2. Daneshmand, M., Abels, A., Anbarjafari, G.: Real-time, automatic digitailor mannequin robot adjustment based on human body classification through supervised learning. *International Journal of Advanced Robotic Systems* 14(3), 1–9 (2017).
3. Daneshmand, M., Aabloo, A., Ozcinar, C., Anbarjafari, G.: Real-time, automatic shape-changing robot adjustment and gender classification. *Signal, Image and Video Processing* 10(4), 753–760 (2016).
4. Traumann, A., Daneshmand, M., Escalera, S., Anbarjafari, G.: Accurate 3d measurement using optical depth information. *Electronics Letters* 51(18), 1420–1422 (2015).
5. Daneshmand, M., Aabloo, A., Anbarjafari, G.: Size-dictionary interpolation for robot's adjustment. *Frontiers in bioengineering and biotechnology* 3 (2015).

My contributions to the papers listed above are as follows, respectively:

1. I have summarized some of the reviewed papers, as well as revised and completed the whole text.
2. This research has been done by me. I have conducted all the programming and simulation. Also I have written the paper, and edited the manuscript.
3. This work has been mostly done by me. I have performed all the required programming tasks. I also wrote and revised the whole paper.
4. I have written parts of the introduction and the experimental results, as well as revised the whole manuscript.
5. I have devised, implemented, and verified the proposed method, as well as written the manuscript.

Conference Papers

1. Valgma, L., Daneshmand, M., Anbarjafari, G.: Iterative closest point based 3d object reconstruction using rgb-d acquisition devices. In: Signal Processing and Communication Application Conference (SIU), 2016 24th. pp. 457–460. IEEE (2016).
2. Daneshmand, M., Traumann, A., Anbarjafari, G.: 3d size-estimation based on the geodesic distance measured by photogrammetric scanning device. In: Proc. of 6th Int. Conf. on 3D Body Scanning Technologies. pp. 221–231 (2015).

My contributions to the papers listed above are as follows, respectively:

1. I have helped Lembit with devising the proposed method and producing the experimental results, and revised the whole manuscript.
2. I have written most parts of the paper, and cooperated in doing the experiments.

Other Publications

1. Daneshmand, M., Masouleh, M.T., Saadatzi, M.H., Ozcinar, C., Anbarjafari, G.: A robust proportion-preserving composite objective function for scale-invariant multi-objective optimization. *Scientia Iranica* 24(6), 2977–2991 (2017).
2. Beheshti, I., Maikusa, N., Daneshmand, M., Matsuda, H., Demirel, H., Anbarjafari, G.: Classification of alzheimer’s disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks. *Journal of Alzheimer’s Disease* pp. 1–10 (2017).
3. Daneshmand, M., Bilici, O., Bolotnikova, A., Anbarjafari, G.: Medical robots with potential applications in participatory and opportunistic remote sensing: A review. *Robotics and Autonomous Systems* (2017).
4. Rasti, P., Daneshmand, M., Anbarjafari, G.: Statistical approach based iris recognition using local binary pattern. *DYNA-Ingeniería e Industria* 92(1) (2017).
5. Avots, E., Daneshmand, M., Traumann, A., Escalera, S., Anbarjafari, G.: Automatic garment retexturing based on infrared information. *Computers & Graphics* 59, 28–38 (2016).

6. Daneshmand, M., Masouleh, M.T., Anbarjafari, G.: Kinematic sensitivity analysis of parallel mechanisms by considering the effect of uncertainties in passive joints. *Modares Mechanical Engineering* 15(10), 78–88 (2016).
7. Daneshmand, M., Saadatzi, M.H., Kaloorazi, M.H.F., Masouleh, M.T., Anbarjafari, G.: Optimal design of a spherical parallel manipulator based on kinetostatic performance using evolutionary techniques. *Journal of Mechanical Science and Technology* 30(3), 1323–1331 (2016).
8. Rasti, P., Taşmaz, H., Daneshmand, M., Kiefer, R., Ozcinar, C., Anbarjafari, G.: Satellite image enhancement: Systematic approach for denoising and resolution enhancement. *Dyna-Colombia* 91(3), 326–329 (2016).
9. Rasti, P., Daneshmand, M., Alisinanoglu, F., Ozcinar, C., Anbarjafari, G.: Medical image illumination enhancement and sharpening by using stationary wavelet transform. In: *Signal Processing and Communication Application Conference (SIU), 2016 24th.* pp. 153–156. IEEE (2016).
10. Daneshmand, M., Tale Masouleh, M., Saadatzi, M.H.: Optimization of the kinematic sensitivity and the greatest continuous circle in the constant-orientation workspace of planar parallel mechanisms. *International Journal of Robotics, Theory and Applications* 4(1), 12–21 (2016).
11. Pjatkin, K., Daneshmand, M., Rasti, P., Anbarjafari, G.: Probability distribution function based iris recognition boosted by the mean rule. In: *Intelligent Computing and Internet of Things (ICIT), 2014 International Conference on.* pp. 47–50. IEEE (2015).
12. Laur, L., Daneshmand, M., Agoyi, M., Anbarjafari, G.: Robust grayscale watermarking technique based on face detection. In: *Signal Processing and Communications Applications Conference (SIU), 2015 23th.* pp. 471–475. IEEE (2015).
13. Bolotnikova, A., Rasti, P., Traumann, A., Lusi, I., Daneshmand, M., Noroozi, F., Samuel, K., Sarkar, S., Anbarjafari, G.: Block based image compression technique using rank reduction and wavelet difference reduction. In: *Seventh International Conference on Graphic and Image Processing.* pp. 981702–981702. International Society for Optics and Photonics (2015).
14. Anbarjafari, G., Rasti, P., Daneshmand, M., Ozcinar, C.: Resolution enhancement based image compression technique using singular value decomposition and wavelet transforms. In: *Wavelet Transform and Some of Its Real-World Applications,* pp. 35–52. Intech (2015).

15. Daneshmand, M., Masouleh, M.T., Anbarjafari, G.: Kinematic sensitivity analysis of a 3-dof decoupled translational parallel mechanism with uncertainties in the passive joints. In: Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on. pp. 85–90. IEEE (2015).
16. Rasti, P., Lusi, I., Sahakyan, A., Traumann, A., Bolotnikova, A., Daneshmand, M., Kiefer, R., Aabloo, A., Anbarjafari, G., Demirel, H., et al.: Modified back projection kernel based image super resolution. In: Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on. pp. 161–165. IEEE (2014).

ABSTRACT

Virtual fitting constitutes a fundamental element of the developments expected to rise the commercial prosperity of online garment retailers to a new level, as it is expected to reduce the load of the manual labor and physical efforts required. Nevertheless, most of the previously proposed computer vision and graphics methods have failed to accurately and realistically model the human body, especially, when it comes to the 3D modeling of the whole human body. The failure is largely related to the huge data and calculations required, which in reality is caused mainly by inability to properly account for the simultaneous variations in the body surface. In addition, most of the foregoing techniques cannot render realistic movement representations in real-time. This project intends to overcome the aforementioned shortcomings so as to satisfy the requirements of a virtual fitting room. The proposed methodology consists in scanning and performing some specific analyses of both the user's body and the prospective garment to be virtually fitted, modeling, extracting measurements and assigning reference points on them, and segmenting the 3D visual data imported from the mannequins. Finally, superimposing, adopting and depicting the resulting garment model on the user's body. The project is intended to gather sufficient amounts of visual data using a 3D laser scanner and the Kinect optical camera, to manage it in form of a usable database, in order to experimentally implement the algorithms devised. The latter will provide a realistic visual representation of the garment on the body, and enhance the size-advisor system in the context of the virtual fitting room under study.

ACRONYMS

AFB	Air Force Base
ALS	Airborne LS
ASTER	Advanced Spaceborne Thermal Emission and Reflection
AVA	Absolute Vertical Accuracy
BA	Bundle Adjustment
BKD	Binary Kernel Descriptor
BMI	Body Mass Index
C2C	Cloud to Cloud
CAD	Computer Aided Design
CPU	Central Processing Unit
CRP	Close-Range Photogrammetry
C-SVM	l_1 Classification Support Vector Machine (SVM)
DBH	Diameter at Breast Height
DEM	Digital Elevation Model
DJBF	Directional Joint Bilateral Filter
DLG	Digital Line Graph
DoF	Degree of Freedom
DOM	Digital OrthoImage
DR	Diagnostic-Robust

DRPCA	Diagnostic-Robust (DR) Principal Component Analysis (PCA)
DSLR	Digital Single-Lens Reflex
DSM	Digital Surface Model
DTM	Digital Terrain Model
DVT	Distorting Video Technique
EPnP	Efficient Perspective-n-Point
ESPI	Electronic Speckle Pattern Interferometry
FAST	Features from Accelerated Segment Test
FHD	Full High-Definition (HD)
FMCW	Frequency-Modulated Continuous-Wave
FoV	Field of View
FpS	Frames per Second
GB	GigaByte
GDEM	Global DEM
GIS	Geographic Information System
GPS	Global Positioning System
GPU	Graphics Processing Unit
GSD	Ground Sample Distance
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HD	High-Definition
HDR	High Dynamic Range
HPR	Hidden Point Removal
HR	High-Resolution
ICP	Iterative Closest Point

InSAR	Interferometric Synthetic-Aperture Radar
ISPRS	International Society for Photogrammetry and RS
IR	InfraRed
ITC	Individual Tree Crown
JTF	Joint Trilateral Filter
KDE	Kernel Density Estimation
LAI	Leaf Area Index
LiDAR	Light Detection And Ranging
LPS	Leica Photogrammetry Suite
LS	Laser Scanning
MAD	Median Absolute Deviation
MD	Mahalanobis Distance
MLS	Mobile LS
MMS	Mobile Mapping System
MSAC	M-estimator Sample And Consensus
MSM	Mass Spring Model
Ncut	Normalized cut
OCC	Open CasCade
OI	Orthogonal Iteration
PAR	Perimeter-Area Ratio
PaRS	Photogrammetry and RS
PCA	Principal Component Analysis
PDJBF	Partial Directional Joint Bilateral Filter
PEB	Proportional Error Back-propagation
PSNR	Peak Signal-to-Noise Ratio

PSP	Panoramic Spherical Photogrammetry
RANSAC	RANdom SAmple Consensus
RBA	Reduced Bundle Adjustment (BA)
RD	Robust Distance
REM	Raster Electron Microscope
RF	Random Forest
RGB	Red Green Blue
RGB-D	Red Green Blue (RGB)-Depth
RMSE	Root-Mean-Square Error
RPAS	Remotely Piloted Aerial System
RPCA	Robust PCA
RpM	Revolutions per Minute
RS	Remote Sensing
RTK	Real-Time Kinematic
RVA	Relative Vertical Accuracy
RZ-score	Reference Z-score
SAR	Synthetic Aperture Radar
SDF	Signed Distance Function
SfH	Scanning from Heating
SfM	Structure from Motion
S-GICP	Simultaneous Generalized Iterative Closest Point (ICP)
SI	Site Index
SIFT	Scale-Invariant Feature Transform
SigVox	Significant Voxels
SLAM	Simultaneous Localization And Mapping

SRT	Sardinia Radio Telescope
SRTM	Shuttle Radar Topographic Mission
SVM	Support Vector Machine
TanDEM-X	TerraSAR-X (TSX) add-on for Digital Elevation Measurement
TLS	Terrestrial LS
ToF	Time of Flight
TS	Total Station
TSDF	Truncated Signed Distance Function (SDF)
TSX	TerraSAR-X
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
USGS	United States Geological Survey
VHR	Very High-Resolution (HR)

CHAPTER 1

INTRODUCTION

Although mannequin robots have been widely used in the context of virtual fitting, which is the principal subject of this thesis, most of the modules involved in the process of online fitting of garments still demand manual calculations, operations and adjustments. Therefore, in order to alleviate the foregoing deficiencies, in Chapter 2, notions and practices aimed at the classification of 3D scanning instances of human body are explained.

Chapter 3 introduces a new and rapid automatic approach for accurate estimation of input values of actuators for real-time activation and adjustment of shape-changing robots. The proposed approach operates based on a set of given sizes randomly scattered and generated throughout the whole range of possible actuator positions, in order to produce specific non-existing three dimensional shapes. The estimation process tries to calculate which input values of the moving actuators could result in the closest representation of the desired positions and distances through calculating the mathematical description of a map relating them to each other. The same ratio of the distance between the desired measurement and the surrounding size-template values is maintained by the interpolation between the input control values commanded and the ones associated with the foregoing templates. As a use case, the proposed method is applied to the visual data obtained by scanning mannequin robots being used for the purpose of photo-shooting in realistic virtual fitting rooms, using 3D laser scanners.

Chapter 4 provides an overview of 3D scanning methodologies and technologies proposed in the existing scientific and industrial literature, which is required for figuring out a suitable framework for the remaining aim of the thesis, namely, creating 3D reconstructions of garments to be utilized in the context of virtual fitting. Throughout the chapter, various types of related techniques are reviewed, which consist, mainly, of close-range, aerial, SfM and terrestrial photogrammetry, and mobile, terrestrial and airborne LS, as well as ToF, structured-light and phase-comparison methods, along with comparative and combinational studies,

the latter being intended to help make a clearer distinction on the relevance and reliability of the possible choices when it comes to the underlying purpose of the thesis. Moreover, outlier detection and surface fitting procedures are discussed concisely, which will be necessary as a post-processing stage.

As one of the applications of 3D scanning methods, an innovative approach to estimating important body measurements according to the depth information extracted by photogrammetric scanning devices, such as Microsoft Kinect 2, is introduced in Chapter 5. Its applicability and efficiency will also be verified by applying it to an experimental use-case concerning classification based on size, to be implemented in virtual fitting rooms. The mathematical framework consists, mainly, in constructing a projection from the screen coordinates associated with the pixels onto the real-world ones, and then integrating the geodesic distance gradients throughout a path connecting the beginning and end points to each other. Besides, in order to reduce the inaccuracy through excluding the possible high-frequency noise, and to obtain more realistic measurements by minimizing the discrepancy between the original path and the one taken into account in the integration module, the path is fitted to a smooth curve in a separate stage before performing the geodesic-distance-calculation iterations. In order to verify the accuracy and preciseness of the proposed technique, it is utilized for obtaining numerous body sizes, where for ensuring that the correct measurements are found, imaginary landmarks are placed on the bodies of the subjects. The results show that the proposed system is capable of making measurements with a tolerance of 0.5 cm using Microsoft Kinect 2.

Chapter 6 introduces, and verifies the applicability of a general framework for creating 3D point clouds resembling objects or scenes, while maintaining the color information, based on multiple RGB-Depth (RGB-D) frames having been taken from different viewpoints by the Kinect 2 camera. In order to improve the efficiency of the proposed solution, an automatic turntable-based hardware configuration is proposed, which considerably enhances the speed and reliability of the capturing procedure through the associated circuitry when it comes to reconstructing objects. A fast and efficient variant of ICP is proposed, one of whose major advantages is that the texture belonging to the original object or scene is mapped to the reconstructed point cloud, with a relatively high precision. The underlying notion for achieving the latter texture mapping goal is that the color associated with every point is transferred throughout the merging process iterations, along with the 3D coordinates of the depth frame.

Furthermore, a robust, real-time loop-closure correction technique is proposed for achieving global consistency in 3D reconstruction, which consists in back-propagating the cumulative transformation error appeared while merging the pairs of consecutive frames. The proposed algorithm assumes that the starting frame

and the last frame of the sequence roughly overlap. Moreover, in order to achieve reliable and realistic representations of objects or scenes, the noise involved in the initial result of the reconstruction procedure is removed through employing a low-pass filter mask, being aimed at detecting and excluding the outliers, followed by applying a bilateral filter for the sake of removing the points standing for high chances of being noisy, as well as smoothing the shape of the point cloud returned by the sensor.

The proposed method, based on the experiments whose results are reported, is computationally less costly than the relevant alternatives suggested in the literature heretofore, which is one of its fundamental contributions while dealing with real-time scenarios. Also neither manual input nor interference is required from the user, which renders the whole process automatic. The strengths and weaknesses of the proposed algorithm are discussed according to the results, where objects of various sorts, i.e. with different colors and types of materials and surface, as well as scene sequences encompassing a wide range of experimental conditions, including different scanning trajectories with reversely directed motions within them, are taken into account, and investigated, as case-studies.

Chapter 7 proposes, and implements, a comprehensive automatic system for creating 3D models of garments that could be retextured as desired, using multiple frames taken from different viewpoints by Kinect 2. As the first step, the frames are denoised and refined by means of a novel procedure devised based on numerous sets of experiments aiming at evaluating and improving the quality of the depth information returned by the camera while capturing garments, under various experimental scenarios, involving severe changes of illumination conditions. Next, the refined frames are passed through an extra preprocessing module, which tries to come up with a smooth and consistent representation of the garment surface, using cubic interpolation on the missing points determined based on the median distance between all the points of the depth frame with their closest neighbors. For the latter purpose, as another major contribution of the chapter, a new algorithm is proposed, which alleviates the deficiencies involved in the previous ones suggested in the literature heretofore, via checking for the holes in all the six possible directions around every point separately.

Finally, the thesis concludes by summarizing the contents, providing general hints as to how the ongoing research on realistic 3D virtual fitting rooms could proceed, and suggesting potential applications for the technologies developed throughout the thesis in other contexts.

CHAPTER 2

OVERVIEW OF VIRTUAL FITTING BASED ON MANNEQUIN ROBOTS

2.1 Introduction

Garment shoppers usually find it tedious and annoying to search through the extremely various ranges of models offered by the brands on each clothing item and try every single option on separately, which is conventionally deemed necessary for gathering enough information to make a decision on the choice of the clothes model itself, becoming even worse when it comes to selecting the size category that would fit their body shapes best [294, 344, 158, 20]. The latter complication has encouraged the development of online fitting utilities, which have been gradually finding their way to the garment retail market during the last decades, leading to a chance to virtually try infinite number of clothing choices and combinations on at once, without requiring making efforts beyond the extent of simply providing the online seller with a rough representation of one's body shape sizes and measurements [25, 181, 253, 275].

In fact, the most recent research projects carried out on this subject by the associated industrial and academic parties clearly indicate that the market is experiencing a substantial variation in the trends preferred by the users: Overall, the tendency to have shopping experiences which are more tangible and less demanding at the same time are drastically increasing over time.

First and foremost, the users wish for a shopping procedure which would alleviate the cumbersomeness of the task of selecting and verifying the suitability of the shopping items. The conventional process involves visiting shops with enormous numbers of different types of items, which are, sometimes, not even organized in an understandable manner. Thus the user will have to physically go to all the shops possibly offering the intended shopping item, keep in mind various specifications of different variants of the item offered in every shop, and finally compare

them to come up with a conclusion as to which variant and from where to buy. The above task entails numerous drawbacks which render the resulting experience slow, inefficient and unreliable. As the first issue, in order to ensure that the best choice is made, finding out which shops offer the desired item is necessary, which is time-consuming, and usually there is not a certain solution to obtain a list containing the foregoing information as a whole. Besides, visiting different shops, which may be scattered throughout the neighboring area, may be exhaustive, especially in metropolitan areas, where commuting within the city for visiting multiple shops might be extremely challenging. Thus it could be inferred that the task of choosing the shopping item using the conventional practice is considerably inefficient and slow.

Moreover, looking for the desired item in the classic form is significantly inflexible, both for the sellers and the shoppers. On the seller side, in the case multiple variants of a given item are available, making all of them visible to the eyes of window shoppers is usually impossible due to the physical constraints, and the considerations which need to be taken into account while deciding which variants of each item to keep in the stock and which ones to move to the store. On the shopper side, trying out different variants of a given item, e.g. different colors or sizes, is impractical, as obtaining a realistic impression of whether the intended item suits the surrounding context would require physically checking it out every time, which is not feasible.

Even though online shopping solutions of numerous sorts have been suggested and implemented heretofore, they still suffer from serious limitations. They bring about the advantage of overcoming the above problems by integrating the information about various shops and the items they are offering into a unified system, which makes it possible to achieve an overall understanding of the possible choices in shopping a certain item. However, the most fundamental shortcoming associated with them is that they cannot provide the user with a tangible and realistic perception of the properties of the item, which is the main motivation for developing VR-based 3D online shopping applications, and will be discussed in what follows.

Principally, even assuming that the online shop provided by the seller contains a comprehensive database of the information required for making a reasonable choice on the purchase in all the aforementioned terms, still the visual information usually available on their websites is not enough for grasping a clear gauge of the visual and material-based characteristics of the item. In other words, from the visual perspective, the images taken from the objects show them from a particular view point, and under specific lighting conditions, which may vary significantly from other perspectives, or under different lighting conditions. Furthermore, the absence of a possibility to touch the item and feel the properties of the material

from which it is made, as well as the pattern present on its surface considerably limits the depth and authenticity of the perception made by the user. Finally, yet importantly, as the object cannot be seen through a contextualized visualization, i.e. where the objects which are going to surround it in the designated location, as well as the specific lighting conditions of the room, have not yet been incorporated into the scene, the suitability of the item for the intended settings cannot be evaluated confidently.

Among other applications portraying either virtual or physical representations of the prospective impressions of the human bodies when wearing a specific garment or set of garment pieces proposed in the literature so far, one of the most successful approaches, which has demonstrated outstanding practical and commercial performance, is the solution employed by online garment retailers making use of mannequin robots as a means to photoshoot the closest possible resemblance of the associated shape, being accomplished through adjusting them based on major body specifications such as girths and the geodesic distances between pairs from a set of decisive landmarks [34]. More precisely, the underlying notion is making photos of a mannequin robot capable of producing body shapes possessing the sizes falling within the desired ranges while putting clothes from different categories of the garments with the models sought from the outset on. The next step is preparing template-based visualizations standing for all the possible combinations of the sizes spanning the whole intervals of the measurements that could be possibly produced, with steps small enough to create representations sufficiently close to all the shapes going to be reconstructed by the mannequin robot.

However, one of the problems usually involved in purchasing garments online is that they cannot be seen from different viewpoints, which may play a paramount role in evaluating the extent to which the garment suits the body. More clearly, from a given point of view, the garment's size, pattern, material, etc. may look pleasing, but not necessarily from all the others. Besides, as the pose of the garment while being captured by a camera or sensor in the most cases affects its appearance, a single shot or sketch may not clearly reflect all the necessary information required for deciding whether the garment is a suitable one or not. In other words, although the physical and mechanical forces and interactions available in the scene, such as gravity, friction, momentum, etc., each influence the associated shape of the garment, in numerous occasions, the garment may be deemed favorable by the customer from a certain perspective, but unfavorably different from others.

Thus it will be substantially helpful to let the user evaluate the appearance of the garment not only from different viewpoints, but also while taking different poses, so that it would be ensured that the fit is evaluated fairly and reliably, thereby increasing the chances of successful online purchases. From a broader perspective,

handling the above issue is also important for enhancing the users' trust in online garment purchases as a whole, which is essential for the service to survive in the market. Lacking confidence in the latter utilities has always been one of their fundamental drawbacks when it comes to competing with the classical, physical alternatives.

From a technical point of view, the construction of the mannequin robot itself should be optimized in terms of electrical and mechanical structures, in order to make it capable of providing flexibility and manipulability levels required for producing the widest possible ranges of the shapes and body-size combinations at a satisfactory degree of ease of use. One of the challenging steps required for achieving the latter goal is to figure out a design strategy that could bear the flexibility to accommodate the details of each particular body. This issue was investigated in [33], by means of a shape-optimization problem. The mannequin robots may represent different gender and size categories, where the ones utilized in the context of the experiments whose results are reported in this chapter, include male and female instances of small, medium and large types.

The process of choosing the most suitable mannequin robot and making it mimic the intended body shape for photoshooting while wearing the desired clothes demands, first of all, classifying the input human body model, which has been obtained by a laser scanner. In other words, the appropriate mannequin robot standing for the right gender and size [103] is determined according to a selection of the ratios between the body size measurements, using maximum likelihood function, along with SVM. Then the closest template from the existing size-dictionary is chosen, which may be further improved in the sense of accuracy by creating a new template via interpolating the templates containing the values immediately preceding and succeeding the ones aimed at [101]. Finally, the corresponding mannequin robot is activated using the associated set of actuator positions, playing the role of control commands.

Being broadly categorized as a completely different approach, one yet commercially immature idea consists in making separate 3D models of the garment and the human body, and then visualizing the approximate appearance of the garment while being virtually worn by the human body. From the list of the most prominent studies having investigated the foregoing concept in the literature heretofore, [177] could be mentioned, which proposes making use of constrained texture mapping for incorporating the color and shading information into the created 3D model of the garment. In another major set of relevant studies, such as [277, 212], on the other hand, the problem of parametrizing the results of triangulating point clouds, which is required for building surface meshes, is addressed. Moreover, the task of superimposing the garment on the human body may be fulfilled using dense point clouds, which has been discussed in [83, 159].

Similarly to most other virtual-reality-based utilities aimed at different academic and commercial settings, due to the technical complexities of the task of devising the modules involved in creating models of the garment and the human body, as well as superimposing and visualizing the results, although numerous shortcomings of the early applications developed on the basis of the foregoing notion have been overcome gradually, still the existing virtual-reality-based applications demand enormous amount of effort and attention in order for them to be deemed proficient and realistic enough to make the majority of the customers consider them trustworthy alternatives to the real-world, physical experience. For example, applications utilizing dense point clouds are usually unable to process sets of points numbering higher than a certain extent, which is normally the case when dealing with 3D point clouds standing for garments and human bodies.

Therefore, in this chapter, rather than the above virtual reality methods making use of 3D models, the focus is on optimizing the performance of the mannequin-robot-based approach in terms of both time-consumption and the average expenses needed to be made per a garment-mannequin combination. Although both design and control problems have been exhaustively analyzed in [34, 33], and the commercial justifiability of the whole application has been successfully demonstrated, the user-friendliness of the service is not still been enhanced to the greatest possible extent, i.e. most components of the online try-on utility yet necessitate manual operation, which might cause slowness, inaccuracy and inconsistency. For alleviating the foregoing problems, this chapter proposes a new version of the solution, which could enhance the reliability and speed of each element, and as a result, the desirability of the whole package, through automating and incorporating it into the procedure, while paying due attention to the physical compatibility considerations.

The rest of this chapter is structured as follows. In the next section, the pick of the bunch from the list of studies dealing with virtual fitting applications and approaches reported in the existing literature is reviewed so as to provide a descriptive background to the current state-of-the-art. Afterward, the properties and specifications of the earlier version of the mannequin-robot-based solution are explained. Next, the technical details of the modifications made in order to improve the efficiency of the system are elaborated, and their impact on the practical desirability of the application, as a whole, are discussed and presented using the experimental results. Lastly, the materials presented in the chapter are outlined and concluded.

2.2 Overview of the Existing Online Fitting Solutions

As a matter of fact, the purpose of a virtual try-on application is to approximate the appearance of the garment while being put on the human body, and visualize the resulting model, possibly from different points of view, to the customer. The latter could be realized through a variety of approaches, which are briefly reviewed in this section, along with the main advantages and disadvantages of each.

Virtual-reality-based utilities aimed at online garment fitting usually make use of a hybrid avatar, which is matched to, and then superimposed on, a 3D model of the human body. The main steps constructing the structure of such methods could be outlined as follows:

- Specifying the associated 3D models and reference points for both the body and the garment upon scanning and obtaining a set of preliminary measurements;
- Creating a segmented garment model;
- Matching the garment model to the 3D model of the body by physical geometric transformation, and then superimposing and refining it;
- Modeling, testing and visualizing the 3D garment model on the body in the virtual fitting room, while noticing the compatibility requirements.

Numerous studies taking the above strategy suggest employing the Kinect™ v2 sensor [12], which does not basically possess enough precision for creating satisfactory representations of the garment while being virtually put on the human body. In [328], the depth distribution accuracy of Kinect 2 has been investigated, and questioned, through a cone model. The depth distribution variance is analyzed based on the entropy. A trilateration procedure is proposed in order to improve the depth accuracy while making use of multiple sensors.

[238] proposes a method for estimating original Kinect depth sensor noise, both axial, i.e. distance from sensor along z axis, and lateral, i.e. along x and y axes. They derive a mathematical model for estimating standard deviations for both types of noises, and conclude that the lateral noise increases linearly with the distance from the sensor, while axial noise increases quadratically. Additionally, they analyze the noise dependency on the surface angle, and infer that it is constant up to about 70° , after which it increases rapidly. They use the noise models for improved filtering of the Kinect data, and in order to enhance the KinectFusion [236] pipeline in terms of reconstruction quality and pose estimation accuracy. A couple of experimental results are compared to those of KinectFusion, and some improvements have been observed.

In [197], a scheme has been proposed to reduce the noise and interpolate the missing values in Kinect depth images. They detected the pixels representing edges based on the RGB information. The existing data were then denoised using Joint Trilateral Filter (JTF) [173] in the case of non-edge data. They also proposed a Directional Joint Bilateral Filter (DJBF) for denoising the edges. With the latter, the direction of the edge was considered when filtering noisy data. Filling the missing non-edge points was performed using a Partial Directional Joint Bilateral Filter (PDJBF), as some of those pixels lay near edges. It should be noted that stronger denoising although leads to less amounts of noisy points, reduces the density of the final point cloud, and decreases the level of details visible from the result. Using the aforementioned type of classification and treating pixels differently enabled them to achieve a denoising approach that rather than losing data would highlight it, which is also illustrated by a much higher Peak Signal-to-Noise Ratio (PSNR) in comparison to other methods.

Many projects intending to develop virtual try-on applications have built and utilized 2D garment models, which is not enough for objectively evaluating the prospective appearance of the garment when put on the body, since they do not possess detailed information about the properties of the cloth, being needed for adjusting and adapting it properly on the human body model. The foregoing drawback has led the associated researchers to make the effort to create 3D clothes models instead [91, 248].

As an instance of the above type of online fitting solutions, the one introduced in [73] could be mentioned. It involves a graphical user interface responsible for creating 2D panels. Moreover, the Miralab [309] has offered an application called MIRACloth. In [253], the main focus is to build 3D garment models that could most properly react to the physical conditions and mechanical forces. A number of studies in the current literature, such as [90, 50, 213, 206], have concentrated on enhancing the compatibility of virtual fitting utilities to be incorporated into the Worldwide Web. Many others, including [118, 84, 93, 233], pay particular attention to building 3D models of human bodies, which is handled by specifying the clothes geometric shape through providing the application by 2D sketches in projects such as the one reported in [311].

Apart from the mathematical and computational complexities and obstacles involved in 3D modeling side, most of the methodologies proposed and implemented heretofore entail the serious drawback that the result is still not realistic and reliable enough to provide an accurate gauge of the prospective appearance of the user while actually wearing the cloth. In contrast, the design and development of physical digi-tailor mannequin robots, although incurs the necessity of designing, controlling and maintaining the robot, is one of the rare feasible solutions for fit advising, as a solid alternative to overcome the problematic nature of virtual

reality and 3D modeling.

Due to the fact that the mannequin-robot-based methodology has to satisfy the specifications of a practical facility, the challenges reported by the online retail sector should be considered, and the actual desires of the costumers have to be taken into account in order for the whole package to meet the requirements and standards of the international markets. Including among others, one of the most important considerations playing roles in the satisfactoriness of the virtual fitting system to the eyes of online garment retailers is the amount of time and finance spent on creating photo shoots of the clothes while put on the body.

For example, creating the body-shape templates is considered a time-consuming and costly task, due to the amount of manual adjustment effort it requires for finding out the appropriate combination of the actuator positions within their feasible ranges, which has been considerably improved through proposing and implementing the interpolation procedure introduced in Chapter 3.10. However, further enhancements of many sorts have to be made to the system, so that optimum performance would be secured by an automatic process of modeling an impression of the clothes using the most suitable mannequin robot, which is dealt with and reported in this chapter.

2.3 The Modified Version of the Virtual Fitting System

In this section, the major modifications which have been made to the virtual fitting system through this thesis are presented, along with discussion on the resulting enhancements and contributions.

The first module of the proposed system handles the task of processing a set of principal measurements of the intended human body. It is needed for determining both the type of the mannequin robot to be utilized and the values specifying the actuator positions making the mannequin robot imitate the target shape. In the context of the experiments presented in this chapter, for the foregoing goal, a 3D laser scanner offered by Vitronic [1] is employed, which comes along a software developed by Human Solutions. It consists of a cabin darkened by means of impermeable curtains covering its surroundings, with four laser sensors located at the corners, such that while moving back and forth through the vertical sliders, they always look at the center of the cabin. Calibrating the scanner might be prompted by the software when it is required, and is conducted using a metallic bar placed vertically in the middle of the scanner in a certain pose. On average, it takes around 29 seconds, in total, to perform a full iteration of scanning. More precisely, it takes the scanner approximately 15 seconds to move the lasers from one end to another on the linear modules, together with 14 seconds spent on initializing and finalizing the procedure.

Since the ultimate goal of the utility is to find out and visualize an estimation of the appearance of the garment while being physically tried on by the customer, as aforementioned, the fundamental aim guiding the process of designing and controlling the mannequin robots utilized in the context of this chapter is to provide the manipulability required for most closely resembling the intended body specifications and detailed characteristics, which is the main focus of the analyses described in [33].

The mannequin robots are constructed in a way that a metallic skeleton supports the whole body, to which a set of actuators are connected. The latter are positioned such that they could push an elastic cover outward, or pull it inward, in order to make the mannequin robot take the desired shape through creating the intended combination of the crucial body sizes and girths. As a characteristic common with reconfigurable robots, the mannequin robots have the capability of changing their shapes. They could also be categorized as humanoid robots, which do resemble humans, but despite other types of the latter sort of robots, are supposed to mimic different human shapes, rather than their actions.

For classification of the input body shape based on gender and size, a set of informative sizes, along with certain relationships between them, should be selected, which would, desirably, maximize the between class distances, and minimize the within class ones. In the case of classification based on gender, according to the results of the investigation of the possibilities conducted in [103], the ratio between the bust girth and under-bust circumference satisfies the foregoing condition to the greatest extent. On the other hand, for classification based on size, taking the fact that the most extreme body sizes have to fit into the intervals sustainable by the garment into account, cross shoulder and hip circumference have been reported to be the most reliable ones in [103].

The above conclusions have been made via drawing scatter plots each representing the values associated with a pair of the measurements returned by the laser scanner, which would provide an illustration for comparing them against each other and obtaining a gauge of how successfully they could be expected to distinguish the data classes. Examples of the foregoing plots, which represent the parameters chosen to be used for the classification algorithm, are shown in Figs. 2.1 and 2.2.

The scatter plots are helpful in terms of diagnosing and configuring the classifier through analyzing the possible causes of misclassification. In other words, as every misclassified sample significantly differs from the norm in the sense of the considered criterion or ratio, by looking at the properties of their measurements, one could decide whether the classification principle has been devised improperly. It should be noted that in some cases, as the main cause of misclassification, the sample has been too irregular, which should be removed from the database.

As the points corresponding to all the available scanning instances are shown in

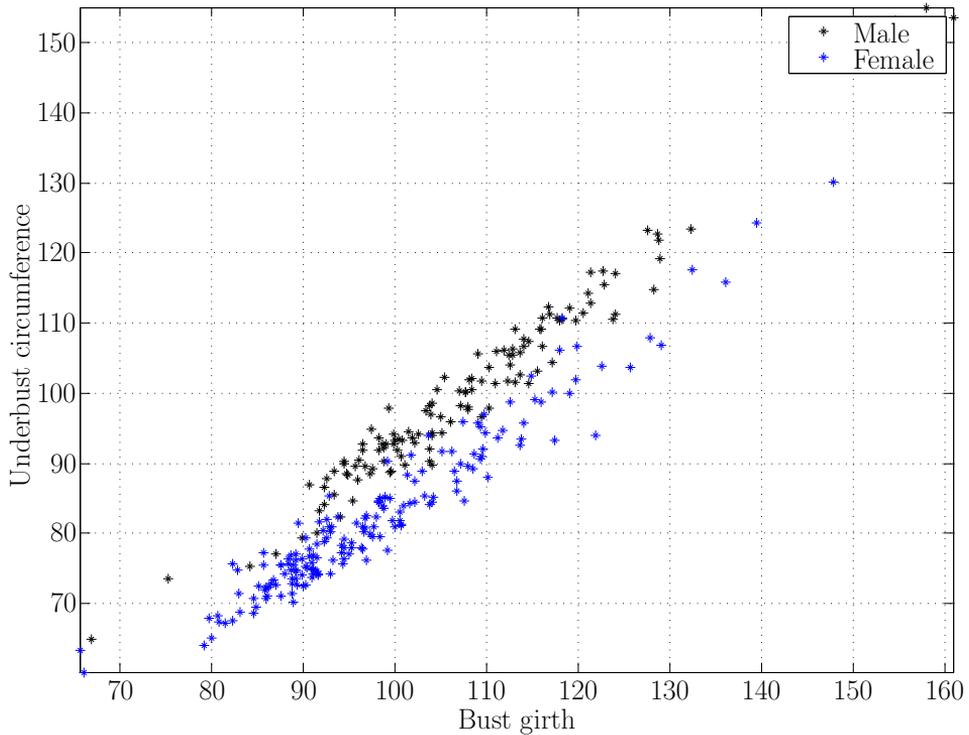


Figure 2.1: A scatter plot showing the measurements used for classification based on gender.

the scatter plots, where the desired class regions could be sought according to the colors associated with each class, the aforementioned figures could greatly help see which one of the criteria stands the highest chance to classify the data accurately. The latter virtue arises from the fact that the ideal criterion would expectedly lead to an illustration in which the colored regions are completely separate and distinguishable, where the least possible confusion or overlap is apparent between them.

In the proposed framework, two measurements are considered for the classification based on size. In other words, the classifier demands simultaneous satisfaction of the conditions associated with the two measurements. The foregoing notion has resulted from the investigation of the scatter plots, which have shown that no single parameter could handle the aforementioned task efficiently.

The classification thresholds are determined through an exhaustive search trying out all the possible combinations of the size steps, and finding the optimum choice, considering minimizing the percentage of the misclassified instances as the criterion. In total, four thresholds are required for the classification based on

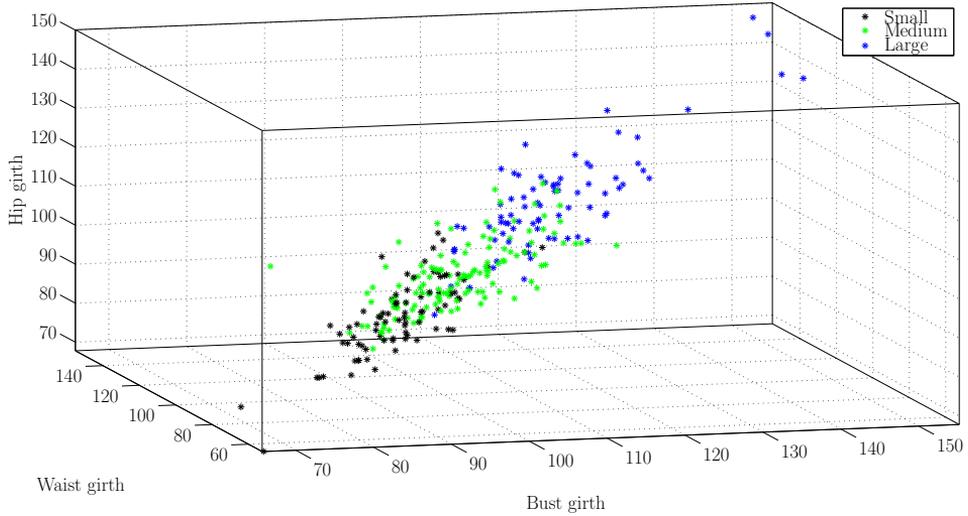


Figure 2.2: A 3D scatter plot representing the measurements utilized for classification based on size.

size, each pair of which specifies the boundaries between the small, medium and large classes for each gender.

Although the classification procedure proposed in this chapter has proved its efficiency in terms of recognizing the real gender and size classes accurately, errors might occur, which usually take place when the body shape is abnormal, either originally or due to the noise affecting the 3D laser scanner software. Furthermore, the true size category itself is not obvious in some cases, which exposes the task of evaluating the performance of the process to human judgment subjectivity. As aforementioned, from mathematical point of view, the classification algorithm is based on the maximum likelihood function, using SVM. Various forms of SVM have been introduced and implemented in studies such as [123, 174, 333, 186, 210, 183]. For the purpose of this chapter, l_1 Classification SVM (C -SVM), being described in [123], is utilized, which is reviewed in what follows. With the assumption that the training data is represented as (\mathbf{x}_i, y_i) , $y_i = \pm 1$, the task of SVM is to calculate the specifications of the hyperplane $\mathbf{w}^T \Phi(\mathbf{x}) + b = 0$ that best separates the two classes involved, i.e. to maximize the distance between the hyperplane and the closest point from each class to it, and minimize the within-class distances through considering a cost function, which is tantamount to the

following optimization problem:

$$\begin{aligned}
& \min_{\mathbf{w}, \boldsymbol{\xi}} \quad \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \mathbf{1}^T \boldsymbol{\xi}, \\
& \text{subject to} \quad y_i (\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \text{ and } \boldsymbol{\xi} \geq 0, \\
& \text{where} \quad \Phi^T(\mathbf{x}_i) \Phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j),
\end{aligned} \tag{2.1}$$

standing for the misclassification penalty coefficient.

After deciding on which mannequin robot to use for producing the intended body shape, a combination of actuator positions should be determined that could make the mannequin robot resemble it as closely as possible. For every mannequin robot, a size-dictionary exists, which contains the set of templates each of which maps a certain configuration of the actuator positions to a shape comprising a specific combination of the geodesic distances between the landmarks on the body. The templates are created through either manually adjusting the actuator positions by trial-and-error, and aiming at producing typical sizes or scanning the mannequin robot while randomly taking different shapes.

For creating the desired shape, the templates containing the values surrounding that of the intended sizes have to be interpolated, so that the actuator positions required for the latter purpose could be found out. The templates utilized for interpolation are determined by solving an optimization problem minimizing the Euclidean norm of a vector standing for the difference between the vector containing the measurements of the desired shape and those of each of the existing templates. The whole interpolation procedure, along with the mathematical expressions, is presented and evaluated in Chapter 3.10.

In order to avoid the waste of time incurred by the need to manually deliver the template positions decided by the classification process to the mannequins, we devised a virtual link between MATLAB and the software package controlling the positions of the mannequins. To be more clear, with assistance of the module developed in our software, after deciding on the mannequin to be activated and the position template to be used, MATLAB automatically displays the results of its calculations, and sends the activation information to the mannequin, which lets the whole process work automatically. In the conclusion of this part of the project, after scanning each person, and obtaining the calculated measurements, the classification and mannequin activation processes operate instantly, where a set of (mannequin, template) data is determined, and the corresponding mannequin has taken the position template chosen in real-time.

As the last module of the proposed automatic virtual fitting system, the intended mannequin robot has to be activated using the actuator positions resulted from the interpolation process. For the sake of reducing the amount of manual labor, the computer is connected to several mannequin robots simultaneously. Therefore, a combinatory circuitry has been devised, which has been adopted from [30], which

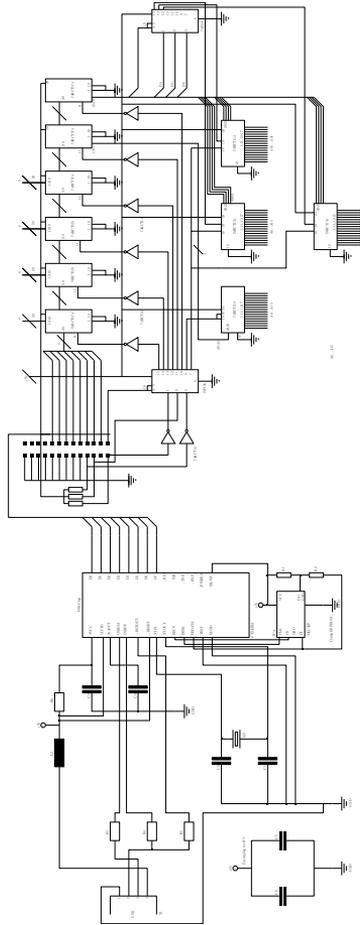


Figure 2.3: A schematic resembling the circuitry being used for activating the mannequin robots after classification [30].

is shown in Fig. 2.3. A pseudo-code summarizing the whole virtual fitting system is presented in Algorithm 1.

2.4 Discussion and Ongoing Research

The application designed and implemented in the context of this chapter has been successfully verified through actual usage, which has demonstrated its effectiveness in terms of improving the time efficiency and ease of use of online fitting utilities. Nevertheless, there are still a bunch of issues to be further investigated and dealt with in the future, which are briefly discussed in what follows.

First, since the use-case is such that the data are mixed, without any prior knowl-

Algorithm 1: A pseudo-code summarizing the virtual fitting process.

```
Data:  $bg$   $\leftarrow$  The bust girth
Data:  $bgu$   $\leftarrow$   $bg$  boundary of medium and large classes
Data:  $bgl$   $\leftarrow$   $bg$  boundary of small and medium classes
Data:  $uc$   $\leftarrow$  The under-bust circumference
Data:  $gt$   $\leftarrow$  The threshold for the gender classification
Data:  $hg$   $\leftarrow$  The hip girth
Data:  $hgu$   $\leftarrow$   $hg$  boundary of medium and large classes
Data:  $hgl$   $\leftarrow$   $hg$  boundary of small and medium classes
Data:  $stm$   $\leftarrow$  The size-template measurements
Result:  $mi$   $\leftarrow$  The ID of the mannequin robot to be activated
Result:  $ap$   $\leftarrow$  The set of actuator positions

1 Function main
2   | go to 3
3 Function gender_classification
4   | if  $\frac{bg}{uc} > gt$  then
5     |    $gender \leftarrow$  female
6   | else
7     |    $gender \leftarrow$  male
8   | return gender
9   | go to 10
10 Function size_classification
11  | if  $bg > bgu$  or  $hg > hgu$  then
12    |    $size \leftarrow$  large
13  | else
14    |   if  $bg < bgl$  and  $hg < hgl$  then
15      |      $size \leftarrow$  small
16    |   else
17      |      $size \leftarrow$  medium
18  | return  $size$ 
19  |  $mi \leftarrow$  combine( $gender, size$ )
20  | go to 21
21 Function interpolation
22  |  $body\_sizes \leftarrow [bg \quad uc \quad hg]^T$   $size\_templates \leftarrow$  minimize(norm( $stm-body\_sizes$ ))
23  |  $ap \leftarrow$  interpolate( $body\_sizes, size\_templates$ )
24  | go to 24
24 Function robot_activation
25  | send( $mi, ap$ )
```

edge of the gender, the algorithm is supposed to handle the task of classifying the bodies based on gender, as well as the size. However, noticing that most retailers have different stores designated for male and female users, the algorithm would have been faster in the case the gender were known from the outset, in which case, one of the classification layers could be rendered unnecessary. The latter, as far as the author is concerned, has not yet been surveyed in the literature heretofore, and is expected to be analyzed in more details within the forthcoming studies.

Besides, the size classes considered throughout the chapter are limited to the three main categories, namely, small, medium and large, where in reality, other levels, such as extremely small or large instances, also exist, which originates from the fact that current mannequin robots, due to mechanical and electrical constraints,

cannot handle such bodies. It is worth mentioning that in the case new robots imitating shapes belonging to the aforementioned extreme classes of sizes are developed, the proposed classification and activation algorithms are still supposed to be flexible and general enough to assimilate their properties appropriately. Last but not least, even though the proposed solution leads to significantly more reliable performance, compared to the classical methodology, especially if accompanied by the real-time, automatic robot adjustment procedure suggested in [183], the time-consumption of the proposed solution could be further reduced via making use of other devices or approaches leading to quicker responses. For instance, as per the current practice, scanning takes 29 seconds, being followed by around 7 seconds of data preprocessing, which could be accelerated by employing photogrammetric scanners instead [299, 114]. Besides, the robot activation could be made more accurate and dexterous by means of optimizing its kinetostatic design [110].

2.5 Conclusion

This chapter described the overall structure of real-time fully automatic adjustment of digi-tailor robots to show the actual shape of a scanned body for a realistic virtual fitting room. The proposed algorithm has been tested on 320 persons scanned by the 3D laser scanner and the automatically adjusted digi-tailor robots have been compared with the corresponding manually adjusted ones. The proposed automatic adjustment is exactly the same as the ones that are manually adjusted, resulting in saving a significant amount of time. The results obtained were illustrated, and the difficulties encountered were discussed. The proposed algorithm automatically classified the gender of the scanned person and estimated the initial size of the body. The experimental results showed that the proposed method for mimicking the body by means of a digi-tailor robot was as accurate as the manual adjustment of robot could be.

CHAPTER 3

REAL-TIME AUTOMATIC ACTUATOR POSITION DICTIONARY GENERATION FOR SHAPE-CHANGING ROBOTS

3.1 Introduction

Design and control of robotic systems, during the last decades, have attracted the attention of numerous researchers [304, 141, 140, 80]. They normally operate based on non-linear relationships between the input control commands and the end-effector pose [249], which are, in the most cases, considerably demanding to formulate, but have been considered the mere solution to the problem of manipulating the resulting movement. Even if successfully figured out, the foregoing relationships necessitate time-consuming and computationally-costly calculations, which undesirably slow down the functional efficiency of the whole system. For example, mechanical manipulators bear significant complexity when it comes to formulating their kinematic, static or dynamic input-output relationships [89, 86, 74, 115].

This chapter proposes an automatic algorithm for facilitating and accelerating the above procedure upon estimating the control inputs leading to approximations of the desired poses, sizes and distances, on the basis of given size-dictionary templates having been obtained via scanning the robot while taking randomly-chosen combinations of the input commands.

The implementation of the aforementioned interpolation process also brings about significant advantages compared to another traditional size-dictionary preparation algorithm, consisting in manually investigating the effect of the changes in the input commands on the resulting output variables, e.g., poses, sizes and distances, which has been used for a long time. In what follows, the foregoing shortcomings, as well as the improvements resulted from the methodology suggested in

this chapter, are concisely explained.

First, due to the fact that the manual size-dictionary preparation is solely based on trial-and-error, it is extremely time-consuming. In other words, for generating each template, one has to randomly enter control values, and manipulate them until achieving the desired sizes. Besides, as changing each distance or size might affect the other ones as well, it further increases the time required for measuring and generating each template. For example, while adjusting the waist girth for producing size-dictionary templates of mannequin robots, the bust and hip girths also change, and *vice versa*. Numerous attempts have to be made in order to achieve the desired combination of the sizes for a single template.

Second, in the case of the traditional approach, only prepared templates are used without further manipulation. Therefore, even assuming that the prepared size-dictionary could contain a substantial number of templates, still they are limited to a certain amount. The latter results in the limitation that, in the best case, the template having been considered the closest one to the desired sizes and distances will be taken into account for mimicking the shape.

The proposed method, on the other hand, provides an opportunity to produce all the desired combinations of the sizes with a maximum tolerance of 2 cm. It is worth noticing that the possible error is due to the aforementioned non-linear nature of the mathematical relationships governing the behavior of the output shape upon changing the input control command. Besides, the foregoing inaccuracy could be overcome to a satisfactory level through increasing the number of the prepared templates, which leads to a closer gauge of the consecutive interpolation values.

The remaining sections of this chapter are structured as follows. First, a general description of the use-case, namely, mannequin robots used for photo-shooting in virtual fitting rooms, is provided. Afterward, a mathematical description of the estimation process for approximately adjusting robots mimicking the desired distances and sizes, based on the prepared templates, is presented. Next, the results of the implementation of the proposed method on the aforementioned case-study are presented and discussed. Finally, the chapter concludes.

3.2 Description of the Robot

This section outlines the general specifications of the shape-changing mannequin robot to be used as a case-study for implementing and verifying the interpolation process. It should be noted that further information about the robot can be found in [34, 191].

The main motivation to design and develop mannequin robots in the context of virtual fitting rooms is to provide a manipulable system being able to take ar-

bitrary shapes and sizes. It could enable the on-line try-on system to adjust the visual data according to the details of the body of the person having been scanned, which means that a single mannequin robot could represent numerous people with different body shapes.

The mannequin robot utilized in this chapter is constructed in a way that a metallic skeleton supports its weight, and an elastic cover, which is adjusted by changing the positions of a couple of layers of linear actuators attached to the skeleton, mimics the desired human body shape.

It should be noted that female mannequin robots, compared to the male ones, possess wider ranges of movement, aiming at accommodating the particularities of female bodies, which, in mathematical terms, could be interpreted as a higher covariance in each dimension.

To the breast part of the female mannequin robots, due to its importance in the visual representation, especial attention has been paid, by making use of breast prostheses, in order to achieve a visual rendering which is as realistic as possible. From a geometric viewpoint, the structure of mannequin robots is similar to reconfigurable ones, since they also consist of a number of blocks, which can produce different shapes through being displaced [323, 140].

Basically, a mannequin robot is a particular type of humanoid robot that mimics shapes instead of behavior. From another perspective, they can be considered parallel mechanical manipulators [64, 45, 180, 72, 230, 111] with the plastic cover as the moving platform. Fig. 3.1 shows a sketch of one of the mannequin robots.

The actuators of the upper part of the mannequin robot are placed such that they could be manipulated in order to adjust the pose and length of the shoulder as desired. The girth of the bust and waist parts, as well as chest and stomach, can be configured in a similar manner.

From a geometric point of view, due to the high level of flexibility resulted from the optimization of the number of the actuators, and their arrangement within the layers, having being reported in detail in the literature [34], not only the body sizes and girths, but also cross-section shapes could be configured, which provides a chance to most closely resemble the body, and if needed, represent it while taking different poses.

The actuators are intended to control the size and shape of either rigid or soft parts of the body, and the aforementioned arrangement optimization has taken such a goal into account as one of its objectives. It should be noted that the actuators can be controlled with a maximum tolerance of 0.25 mm.

Segments of the Pedilin[®] [26], which is typically used for prosthetic device development purposes, as well as actuators using special joints, are considered for building the internal level of the cover. The main properties of the foregoing material include its flexibility, non-stretchiness and manipulability under heat, which

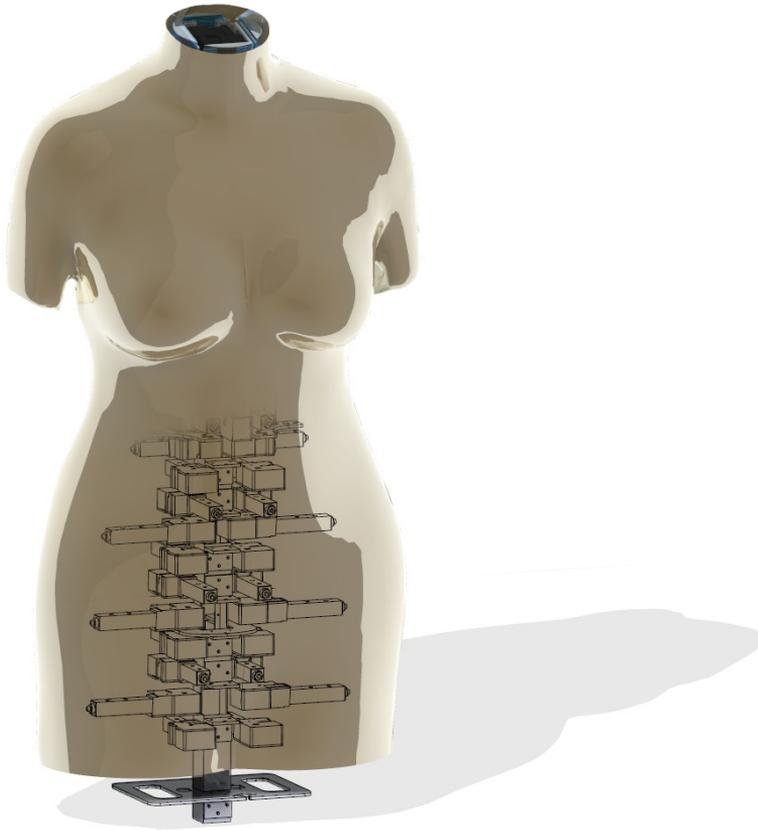


Figure 3.1: A sketch of a female mannequin robot, with the placements of some of the actuators shown.

qualify it as a suitable choice for the application being discussed. The actuators can have from two to four Degree of Freedom (DoF)s.

In [34], a 3D mesh-based model for simulating the body shape according to the input actuator positions has been suggested, which has also been used for the purpose of the above optimization procedure. In the foregoing study, a special case of Mass Spring Model (MSM) [41, 85, 143, 278] has been utilized for developing the 3D model, having successfully simulated the flexibility of the joints attaching the cover segments to each other, as well as to the actuator end points.

In [34], an approach has been proposed for imitating the body shape, which considers the weighted mean square Euclidean distance between the landmarks on the scanned body and the physical mannequin robot as the objective function. It iterates over a linearized relationship describing the movements of the cover resulted from the changes in the positions of the actuators. The latter leads to similar

results for on-line fitting applications.

The main deficiency of the above approach compared to the one introduced in this chapter is that the latter simplifies the methodology, and speeds up the calculations upon constructing new shapes according to a couple of prepared size-dictionary templates, which could be generated fast and in a random manner, instead of modeling overcomplicated mathematical relationships governed by both the geometric construction and the forces being applied from each of the actuators to the cover segments.

It is worth noticing that a virtual fitting cabin [91, 118, 184] needs to be equipped with realistic 3D visual models of mannequin robots [203, 204, 205, 246, 325, 310, 315] capable of producing representations of the costumers based on the prediction of their prospective appearance after actually wearing the piece of cloth.

3.3 The Proposed Method

The purpose of the proposed procedure is to save time through avoiding having to prepare size-dictionaries manually, and to increase the accuracy and flexibility of the robots in terms of mimicking specific shapes and sizes [101]. It should be noted that shape generation and interpolation procedures deal with the deformation-registration problem [247, 270, 339, 239, 164, 57, 226].

In the case of the classical manual approach of preparing the templates, they had to be carefully adjusted to predetermined shapes, which could guarantee the maximum difference from frequent combinations of the sizes. However, one of the most important advantages of the above interpolation process is that the placement of the size-dictionary templates' sizes no longer matters, which significantly enhances the speed.

The interpolation algorithm proposed in this chapter is manipulable enough to produce arbitrary combinations of the sizes accurately, as long as the templates are scattered fairly randomly throughout the whole possible ranges of variations of the inputs.

The procedure consists in making the robot produce random sizes while performing a loop, scanning it while taking each of the shapes, extracting the required measurements, and collecting all the acquired size templates in a database. The Vitronic 3D laser scanner used in this chapter consists of a cabin darkened by means of impermeable curtains covering all over its surroundings, with four laser sensors placed at the corners, such that while moving back and forth through the vertical sliders, they point at the center of the cabin. The aforementioned sensors act as both transmitters and receivers of the laser beam, thereby they calculate the distance between them and the body surface points.

The templates being considered for verifying the efficiency of the proposed inter-

polation algorithm have been prepared using the ANTHROSCAN full body 3D scanner²⁷ provided by Human Solutions GmbH²⁸ [16], which offers a strong level of accuracy, i.e. makes a maximum of 1 mm miscalculation while determining the body girths.

While calculating the input commands associated with each of the target distances or sizes, it is first investigated that between which consecutive size-dictionary templates it falls, where the related control values for producing them have been already saved in the database. Afterward, the appropriate input variables required for producing the target distance or size are determined as follows. The first step is obtaining the distance between the desired output parameter and the ones represented by the consecutive templates. Subsequently, the ratio of them to each other is computed. Finally, the control commands are calculated such that the foregoing ratio is maintained between them and the actuator positions associated with the aforementioned consecutive templates.

If the foregoing consecutive template values are included in multiple templates, the ones used in the interpolation process are selected based on the solution to an optimization problem which intends to minimize the Euclidean distance between the vectors containing all the sizes associated with the desired output and that of the templates, as described mathematically in the next section. Besides, if it happens that one of the desired output values is already included in one of the templates, the exact values will be directly used for mimicking that particular measurement. A detailed mathematical description of the foregoing interpolation process is provided in Section 3.4.

In the specific context of the case-study considered in this chapter, the actuator values associated with each of the important girth measurements are interpolated based on a pair of the already existing templates including the values between which the desired value falls. Through the optimization procedure, which will be mathematically described in what follows, the objective function consists of the Euclidean distance between the vectors containing all the three desired measurements and that of the templates.

The above attempt is motivated by the fact that the girth measurements are not independent from each other, i.e. as aforementioned, changing the control commands supposed to be associated with each particular body girth might affect the others as well. Therefore, while selecting templates being used for interpolation, two criteria should be taken into account. First, they have to include the values immediately lower and higher than the desired one. Second, they should lead to body-size vectors that are as close to the desired measurements vector as possible. While calculating the actuator positions suitable for producing a specific bust girth, waist and hip girths should be also selected by minimizing the Euclidean distance between the vectors containing the desired sizes and that of the tem-

plates.

Assume that D_b is representing a desired bust girth whose value is between two consecutive template values such as P_b and N_b , i.e. $P_b < D_b < N_b$ and M_{w_j} and M_{h_j} are denoting waist and hip girth values in the associated size-dictionary templates. The mathematical description of the interpolation procedure can be formulated as follows:

$$d_{b_i} = w_{b_p} p_{b_i} + w_{b_n} n_{b_i}, \quad (3.1)$$

$$i = 51, 52, \dots, 56, \quad \text{or} \quad i = 61, 62, \dots, 66,$$

$$d_{w_i} = w_{w_p} p_{w_i} + w_{w_n} n_{w_i}, \quad (3.2)$$

$$i = 31, 32, \dots, 36, \quad \text{or} \quad i = 41, 42, \dots, 46,$$

$$d_{h_i} = w_{h_p} p_{h_i} + w_{h_n} n_{h_i}, \quad (3.3)$$

$$i = 11, 12, \dots, 16, \quad \text{or} \quad i = 21, 22, \dots, 26,$$

in which

$$w_{b_p} = \frac{N_b - D_b}{N_b - P_b}, \quad w_{b_n} = \frac{D_b - P_b}{N_b - P_b}, \quad (3.4)$$

$$w_{w_p} = \frac{N_w - D_w}{N_w - P_w}, \quad w_{w_n} = \frac{D_w - P_w}{N_w - P_w}, \quad (3.5)$$

$$w_{h_p} = \frac{N_h - D_h}{N_h - P_h}, \quad w_{h_n} = \frac{D_h - P_h}{N_h - P_h}, \quad (3.6)$$

resulting in

$$d_{b_i} = p_{b_i} + (n_{b_i} - p_{b_i}) \frac{D_b - P_b}{N_b - P_b}, \quad (3.7)$$

$$i = 51, 52, \dots, 56, \quad \text{or} \quad i = 61, 62, \dots, 66,$$

$$d_{w_i} = p_{w_i} + (n_{w_i} - p_{w_i}) \frac{D_w - P_w}{N_w - P_w}, \quad (3.8)$$

$$i = 31, 32, \dots, 36, \quad \text{or} \quad i = 41, 42, \dots, 46,$$

$$d_{h_i} = p_{h_i} + (n_{h_i} - p_{h_i}) \frac{D_h - P_h}{N_h - P_h}, \quad (3.9)$$

$$i = 11, 12, \dots, 16, \quad \text{or} \quad i = 21, 22, \dots, 26,$$

regarding the actuators belonging to the layers influencing the bust, waist and hip girths, respectively, where d_{b_i} , d_{w_i} and d_{h_i} stand for the actuator positions required for producing the desired bust, waist and hip girth measurements, namely,

D_b , D_w and D_h , respectively. Similarly, the sets $(p_{b_i}, p_{w_i}, p_{h_i})$ and $(n_{b_i}, n_{w_i}, n_{h_i})$ denote that of the size-dictionary templates including the initial and final interpolation values, being represented by (P_b, P_w, P_h) and (N_b, N_w, N_h) , having been obtained through solving the following optimization problems:

$$\begin{aligned} & \min \left(\sqrt{(D_w - M_{w_j})^2 + (D_h - M_{h_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (P_b, M_{w_j}, M_{h_j}) \in M, \end{aligned} \quad (3.10)$$

$$\begin{aligned} & \min \left(\sqrt{(D_b - M_{b_j})^2 + (D_h - M_{h_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (M_{b_j}, P_w, M_{h_j}) \in M, \end{aligned} \quad (3.11)$$

$$\begin{aligned} & \min \left(\sqrt{(D_b - M_{b_j})^2 + (D_w - M_{w_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (M_{b_j}, M_{w_j}, P_h) \in M, \end{aligned} \quad (3.12)$$

and

$$\begin{aligned} & \min \left(\sqrt{(D_w - M_{w_j})^2 + (D_h - M_{h_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (N_b, M_{w_j}, M_{h_j}) \in M, \end{aligned} \quad (3.13)$$

$$\begin{aligned} & \min \left(\sqrt{(D_b - M_{b_j})^2 + (D_h - M_{h_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (M_{b_j}, N_w, M_{h_j}) \in M, \end{aligned} \quad (3.14)$$

$$\begin{aligned} & \min \left(\sqrt{(D_b - M_{b_j})^2 + (D_w - M_{w_j})^2} \right) \\ & \quad \text{over } j = 1, \dots, m, \\ & \text{Subject to } \exists j : (M_{b_j}, M_{w_j}, N_h) \in M, \end{aligned} \quad (3.15)$$

respectively, with $j = 1, \dots, m$ representing each template in the size-dictionary. The global set M is described as follows:

$$\begin{aligned} M = \{M_j | j = 1, \dots, m\} = \\ \{ (M_{b_j}, M_{w_j}, M_{h_j}) | j = 1, \dots, m \}. \end{aligned} \quad (3.16)$$

The calculated values are satisfying the following condition:

$$\begin{aligned}
& P_b < D_b < N_b, \\
& \nexists M_j \in M : P_b < M_{b_j} < D_b, \\
& \nexists M_j \in M : D_b < M_{b_j} < N_b,
\end{aligned} \tag{3.17}$$

$$\begin{aligned}
& P_w < D_w < N_w, \\
& \nexists M_j \in M : P_w < M_{w_j} < D_w, \\
& \nexists M_j \in M : D_w < M_{w_j} < N_w,
\end{aligned} \tag{3.18}$$

$$\begin{aligned}
& P_h < D_h < N_h, \\
& \nexists M_j \in M : P_h < M_{h_j} < D_h, \\
& \nexists M_j \in M : D_h < M_{h_j} < N_h,
\end{aligned} \tag{3.19}$$

where if any of the desired body shape values does not fall within the range imposed by the mechanical or electrical constraints, i.e. it is lower than the minimum or higher than the maximum, the closest value allowed in the foregoing interval will replace it.

3.4 The Experimental Results and Discussion

As aforementioned, a mannequin robot utilized for the purpose of photo-shooting in a virtual fitting room is considered as a use-case in order to validate the methodology introduced in this chapter. Typically, bust, waist and hip girths are taken into account as important measurements while mimicking specific body shapes for on-line try-on of clothes.

The experimental results demonstrate the acceptability of our proposed method compared to the alternative approaches. In order to finalize and test the size-dictionary interpolation algorithm, we created 20 random size-dictionary, and generated over 500 arbitrary body shapes according to them, where the maximum errors in cm can be represented as follows:

$$\begin{aligned}
& bg_m \in [bg_d - 0.5, \quad bg_d + 1.2], \\
& wg_m \in [wg_d - 1.3, \quad wg_d + 1.2], \\
& hg_m \in [hg_d - 0.6, \quad hg_d + 0],
\end{aligned} \tag{3.20}$$

where bg_d , wg_d and hg_d , respectively, stand for the desired bust, waist and hip girths. It should also be noted that bg_m , wg_m and hg_m denote the sizes generated

Table 3.1: Desired and produced measurements of the important body sizes (cm).

	Bust girth		Waist girth		Hip girth	
	Desired	Produced	Desired	Produced	Desired	Produced
1	103.2	104.4	83.8	84.4	100.3	99.7
2	110.9	110.9	99.6	98.3	110.8	110.7
3	113.9	114.1	88.4	89.1	106.5	106.4
4	113.6	113.1	100.6	100.3	111.0	111.0
5	114.6	114.7	88.2	89.2	106.4	106.3
6	115.1	115.1	100.1	99.6	111.0	111.0
Extremal error	-0.5↔1.2		-1.3↔1.0		-0.6↔0.0	

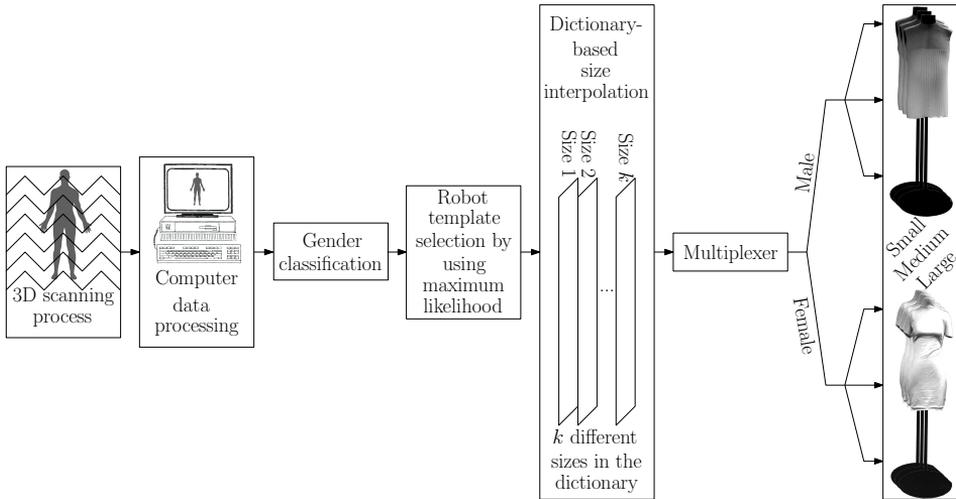


Figure 3.2: A flowchart showing the whole classification and automatic adjustment processes in a glance.

by the interpolation procedure for bust, waist and hip girths. The results of the foregoing calculations are provided in Table 3.1, where the sizes associated with the size-dictionary templates demonstrating extremum inaccuracies, along with the maximum tolerance, are included. The latter table verifies the reliability of the proposed interpolation process in the sense of accuracy. A schematic representation of the whole algorithm utilized for data processing, interpolation and, subsequently, producing desired body shapes via activating mannequin robots is shown in Fig. 3.2.

Utilizing the methodology proposed in this chapter, one would be able to produce size-dictionary templates using the automatic measurement calculation provided by the scanner, instead of manually taking all the measurements.

3.5 Conclusion

This chapter proposed a new approach for producing approximate shapes which do not exist in prepared size-dictionary templates, through linear interpolation. More precisely, in the proposed method, first the robot produces randomly determined body shapes, while a 3D laser scanner scans it, and automatically calculates the prescribed distances and girths, being recorded in a database, along with the corresponding input control commands. Afterward, in order to produce each new body size, the already existing template sizes surrounding it are considered as the initial and final values of a linear interpolation algorithm estimating the actuator positions required. Noticing the fact that the body sizes are not independent from each other, and changing each of them might affect the other ones as well, in the case multiple templates contain the same values required for interpolation, the one is selected that is the solution to an optimization problem aiming at minimizing the Euclidean distance between the vectors consisting of the desired sizes and those of the templates. The foregoing interpolation procedure tries to maintain the same ratio of the distances between the desired sizes and those of the templates while calculating the actuator positions based on the ones corresponding to the size-dictionary templates taken into account. As a use-case for testing the proposed method, mannequin robots being used for virtual fitting purposes were considered. The results of the implementation of the algorithm on the mannequin robot were showing an acceptable level of accuracy.

CHAPTER 4

A SURVEY OF 3D SCANNING

4.1 Introduction

In Photogrammetry and RS (PaRS) and computer vision, automatic interpretation of scenes plays an important role. Robotics, automatic control and manufacturing are some of the principal applications of 3D perception. Depth sensing, range imaging and 3D shape measurement are the most essential components of typical 3D reconstruction frameworks proposed in the literature, which mostly suffer from considerable pronenesses to inaccuracy, and lead to reconstructed point clouds which are not dense enough for a wide range of applications.

Moreover, the existing 3D scanning procedures which have been developed for making models of objects and scenes are limited to applications which do not require neither a real-time nor an automatic performance, although numerous advancements have been made through automating some of the underlying modules recently.

Furthermore, as a consequence of the lack of a sufficiently high number of relevant databases, developing robust 3D scanning algorithms is still considered a challenging task. Although ground-based Interferometric Synthetic-Aperture Radar (InSAR) and Total Station (TS) are generally relatively faster and more accurate, if the object of interest possesses a distinguishable surface, then the above tasks can be performed more efficiently and reliably using photogrammetric approaches.

3D scanning procedures have various applications in documenting, recording and monitoring heritage sites, landscapes and objects in multiple resolutions [262]. Progress reporting of construction sites is a fundamental application of PaRS, where based on the regular scans made and compared with each other, as well as the desired result, the rest of the plan could be decided [127]. ALS data were used for extracting roofs based on a hybrid generative statistical approach in [166]. Moreover, management programs and dune and beach behavior related

to anthropogenic and natural factors are dependent on coastal monitoring, which can be performed using TLS, satellite imaging or Light Detection And Ranging (LiDAR).

In [157], through an iterative process consisting in printing and editing a 3D model, it was evaluated with respect to a reference model, with a function measuring the accumulated error as the criterion. The main drawback of the proposed method is the time-consumption of the printing process, which prevents performing a high enough number of iterations. Moreover, it does not operate well on translucent or shiny surfaces.

In [342], MLS data of pavement cracks, road surfaces, manhole covers and road markings were acquired at driving speeds.

In [126], it was shown that as far as creating Digital Terrain Model (DTM)s is concerned, TS performs most efficiently, which, as a drawback, takes longer to collect the required data. LS, which performs similarly well, was reported to demand rather complex computational procedures. Aerial photogrammetry was thought to rely on rather expensive equipment for data collection. Finally, Real-Time Kinematic (RTK) Global Positioning System (GPS) led to the least robust performance.

In [58], a digital camera being traveled by means of a Unmanned Aerial Vehicle (UAV) took a task traditionally being done through LiDAR, in order to reduce the costs, which was tested on mapping the topography of a fault zone. The images were processed for reconstructing the ground using the Agisoft PhotoScan software [307]. It was also shown that SfM-based photogrammetry leads to a rather dense reconstruction.

In [286], in order to reduce the computational complexity of BA, an Reduced BA (RBA) algorithm was proposed, which is comparatively slightly less accurate, and corrects the camera poses based on their oblique and nadir parameters.

In [147], a flume under construction was monitored using a homemade scanner, which consisted of a camera and a projector, benefiting from a rather high precision and accuracy, given the compact setup offered.

According to [272], using high-speed video cameras [301, 244] or stroboscopic illumination [271] along with 3D image correlation photogrammetry may lead to an approach to measurements of full-field stress, strain, deformation and displacement. Ionic polymeric and dielectric elastomer artificial muscles [276, 44] can then benefit from more reliable measurements than the results of utilizing techniques such as moiré topography [289] or Electronic Speckle Pattern Interferometry (ESPI) [214].

In [272], a flywheel in a spin pit was monitored, where the maximum speed was 35,000 Revolutions per Minute (Rpm). Moreover, multiple scans were made as strain maps of a notched rubber dogbone sample pulled to failure in 38 ms.

Voxel- and surface-based approaches [55, 193] have been developed for deciding the visible points from a point cloud visualized at a certain pose. In [38], the foregoing issue has been investigated, covering topics such as Hidden Point Removal (HPR), gap detection and classification [170], camera network design [39, 148] and providing guidance using synthetic images [55].

According to [285], another application of photogrammetry is in recovering information based on geometrical analysis of images shared through social media, e.g. about important places and events.

Frequency-Modulated Continuous-Wave (FMCW)-driven frequency-scanning antennas were utilized in order to develop a compact 3D imaging radar in [142]. Essentially, it operates on the basis of two orthogonal images taken at a time. Under typical conditions, parallel and angular determination of the foregoing images suffices for acquiring the desired data. In the aforementioned study, 2D images were joined, in order to make pseudo-3D landscape models in real-time.

In [346], stereo pairs of Raster Electron Microscope (REM) images were utilized in order to generate 3D models of miniaturized objects. After calculating the initial grid using Random Sample Consensus (RANSAC), based on arbitrary magnification coefficients, point-of-view and parallel projection models were found. DEMs and orthophotography of surfaces larger than 3 km² were calculated at Very HR (VHR), using a SONY NEX-5N camera [303] and two delta kites of 4 and 10 km², being a compromise between cost, image quality and geometry and weight. Assuming the flight trajectory to be clear and open, the stability of the kite flight angle was evaluated through performing multiple flights, under various circumstances, and with different lengths.

In [167], the thickness of transmission lines was measured using photogrammetry, in order to prevent the potential resulting accidents. The system they developed consists of an inertial measurement unit, a laser range finder and an HR camera.

In [79], oil painting were reconstructed in 3D by acquiring orthophotomaps via UV inkjet printing [71]. Based on the condition equation, they represented the image details, and analyzed the reconstructive accuracy. In order to print 1-10-layer spot color swatches, according to spot color swatches in four colors for a UV printer, the hypsometric tints and contours of the model were found, followed by RGB image segmentation.

The aim of this chapter is to help obtain an overall understanding of different 3D scanning technologies, which is of paramount importance when it comes to doing the task related to the upcoming chapters, i.e. making 3D reconstructions of garments. Therefore, close-range, aerial, SfM and terrestrial photogrammetry, as well as mobile, terrestrial and airborne LS, along with ToF, structured-light and phase-comparison methods, will be discussed throughout the rest of the chapter, followed by a few comparative and combinational studies, and topics such as out-

lier detection and surface fitting. The articles which have been selected for the purpose of this chapter have been published during the last four years up to the year the thesis itself is being published, except for a few highest-cited classical ones.

4.2 Applications of Photogrammetry

A general use-case of photogrammetry is in target recognition. In [139], two algorithms were parallelized on Central Processing Unit (CPU) using OpenMP [24], taking advantage of Graphics Processing Unit (GPU) for target detection. They reported that high-resolution images affect the time-consumption of the algorithms which have been implemented on GPU more significantly than those on CPU.

Another application of photogrammetry is in monitoring protected natural areas. In [121], diapositive pairs were scanned, and then processed by means of Leica Photogrammetry Suite (LPS) 9, in order to monitor ecological degradation. Considering Geographic Information System (GIS) data, a survey of the aesthetic forest Kouri Almyrou Magnisias was carried out.

Photogrammetry can be utilized for measuring the relative pose of satellite-rocket during separations. In [312], using Orthogonal Iteration (OI) [219], i.e. LHM (Lu, Hager and Mjolsness), and Efficient Perspective-n-Point (EPnP) [331], the pose of the camera installed on the launcher was estimated based on the camera parameters and the cooperation signs, which must be more than six.

4.3 CRP

CRP is one of the main types of photogrammetry. Its main use-cases will be reviewed in this section.

4.3.1 Calibrating Cameras

Calibration of cameras according to the results of sensor analysis is a requirement for acquiring CRP data, which may be performed either in a manual or automatic form. In [222], the foregoing topic was investigated aiming at target-less scenes and targeted planar arrays.

Digital Single-Lens Reflex (DSLR), consumer-grade and high-end cameras [182] have to be calibrated independently from the scene of interest, which can be achieved using coded targets [136]. Color imaging, lens distortions, image scales, focal plane distortions, chromatic calibration, interior orientation stability, image point distribution and 3D object space control are the main issues which have attracted attentions of the relevant research communities during the last few

decades. It should be noted that the presence of noise makes the task of camera calibration challenging.

4.3.2 Detecting and Matching Lines

Performing line matching in industrial environments was discussed in [215], which involve varying illumination conditions and low-textured objects. A continuity criteria and phase congruency were taken into account in order to detect line segments, being then matched based on their photometric characteristics and local appearance. Fragmentations and overlaps are among undesired phenomena which need to be avoided.

4.3.3 Monitoring Forests

Monitoring forests is a main application of CRP. In [67], woody debris and understory vegetation were measured in 3D, for predicting and distinguishing plant diversity and understory fuels. Fuelbed depth measurements were obtained, in order to detect plant types and species and fuel types, at a scale of 10 cm, on the Air Force Base (AFB) in northwestern Florida as a case-study, using the photogrammetry software PhotoModeler Scanner [27], which resulted in accuracy rates of 44%, 39% and 44%, respectively. The “Smart Match” feature-based approach was utilized in order to find pixels with matching textures and colors across pair of images. Point-intercept and photogrammetric measurements were reported to be generally similar.

Similarly, in [227], eroded areas were detected using image classification with automatic thresholding, based on the Excess Green Vegetation Index. The spectral properties of the segments were described using the raster-based feature ExG, leading to accuracy rates between 90.7% and 95.5%.

4.3.4 SRT Simulation

In [69], self-load deformations of the SRT primary reflector where measured with respect to its desired shape, with a real-time performance. A BA process was developed to assess the accuracy according to Root-Mean-Square Error (RMSE). MATLAB [21], MicMac [22] and AICON 3D Studio [5] were utilized for simulation. The locations of interest were calculated based on the radial distance from the reflector vertex.

4.3.5 River Surface Topography

Hydro-morphodynamic studies rely, to a certain extent, on surface water topography, which provides useful information about the mass balance, momentum and

energy. In [153], HR DEMs were created using wax powder as seeds. 3DM Calibcam [4] and 3DM Analyst [3] software were used for analyzing wax powder stereo images.

4.3.6 Modeling Cities

In [282], city models were obtained using image sequences providing a 360° coverage. The camera poses were estimated for a set of key-frames, based on the fundamental matrix. The data were first converted to point clouds through SfM, using Agisoft Photoscan Pro [291]. Then 3D models were generated by means of Autodesk® 3ds Max® [261].

4.4 Aerial Photogrammetry

Using low-cost cameras along with modern robotic and computer-vision technologies enables UAVs to provide high-precision 3D measurements [87], e.g. for detecting landslide cuts which need to be repaired. In [124], multiple cameras with overlaps amounting to 60% of the Field of View (FoV) of adjacent cameras were utilized, which resulted in a relatively high accuracy. Other example applications of the foregoing type of technologies will be discussed in this section.

4.4.1 The ISPRS Benchmark

The ISPRS Benchmark on 3D Semantic Labeling [265] has been used relatively rarely in the existing literature, because of the fact that many classes demonstrate similar geometrical behaviors, which makes the task of classification difficult. For estimating image orientation and dense image matching, oblique airborne images were acquired in [237], along four nadir and cardinal directions, together with convergent and redundant terrestrial images, with two areas as case-studies. In [59], cylindrical and spherical neighborhoods were considered for the purpose of semantic labeling, employing a multi-scale representation of the areas around the points, which was applied to the ISPRS database.

4.4.2 Extracting and Matching Features

In [287], segment matching and tree key exchange were utilized along with the red-black tree structure [161], in order to alleviate the heavy computational load of Scale-Invariant Feature Transform (SIFT), with near-zero pitch/roll angles possibly making the results partially unrealistic.

In [92], the centers of optical targets and their distances to the UAV mapping elevations were found, being slightly affected by exposure-induced bias.

4.4.3 Historical Heritage Monitoring

In [221], the “Las Higueras” architectural structure was reconstructed in 3D, using a Canon EOS 5D Mark II camera, which has a resolution of 21.1 Megapixels [10]. Two third of the 964 images they took were used in the experiments conducted through MicMac [22], at a resolution of 4 pixels per mm.

Cartography is another application of UAV photogrammetry, which is necessary in cases where the target cannot be monitored using terrestrial devices. In [232], the foregoing task was performed based on DJI Phantom 3 [13] Professional UAVs and Sensefly Albris [29], with two historical buildings in Strasbourg, namely, the St-Pierre-le-Jeune Catholic church and the Rohan Palace facade, along with Josephine Pavilion, as case-studies.

4.4.4 Ground-level Modeling

Lightweight Unmanned Aerial System (UAS)s can be used for monitoring or modeling territory boundaries, which may involve steep surfaces such as river basins, landslides, quarries or mountains. In [76], a large sandpit with semi-circular, radial and linear configurations was modeled based on 300 control points, using topographic TS.

Multi-view aerial stereo images result in noisy and incomplete reconstructions. In [201], the resulting point clouds were converted to simplified meshes, treating with this task as a labeling problem.

In [75], the Microdrones md4-20041 UAV was utilized for modeling landslides through the Photomodeler Scanner V2012 [68] software. Mosaic images and DEMs underwent orthorectification, consisting in reprojecting the mosaic according to the morphology of the DEM, and then filling the holes in Digital Orthoimage (DOM)s using interpolation.

4.4.5 Monitoring Topography

3D models of cliff faces can be created for kinematic stability analysis using UAVs, more easily than TLS. In [146], two relatively sensitive areas of the Portuguese northwest coast were considered for the foregoing purpose, which was followed up in [269], with the Ravenna beach dune system considered for the test. The results of such studies are of principal importance in management and economy, as well as in biology, for observing behaviors of habitats and vegetation.

4.4.6 Fault Plane Detection

According to [42], correctly adjusting control points and considering non-metric and oblique camera combinations increase the precision and accuracy of aerial

photogrammetry, producing LiDAR or DEM data, where through interpreting extreme heights and detecting cross-sections, fault planes can be detected.

In [337], the Tanlu fault zone throughout China East was analyzed using aerial photogrammetry. The Inpho photogrammetry software [17] was used for a morphotectonic survey of vertical and horizontal deformations, resulting in gradient and landscape maps.

4.5 SfM-based Photogrammetry

TLS technologies, including GPS, can be costly and inconvenient, due to inaccessibility of some of the required control points. In [317], the SfM algorithm was proposed for alleviating the foregoing problems using redundant BA processes [300]. In principle, it operates based on matching feature points across collections of images acquired using a computer-level digital camera. It can be flexibly utilized for obtaining accuracies of decimeter order on land levels with various scales and topographies, which was tested on a glacially-sculpted bedrock ridge, a breached moraine-dam complex and an exposed rocky coastal cliff.

An application of SfM photogrammetry is in ecology, e.g. for monitoring dry-land vegetations, constituting 40% of the whole terrestrial surface. An example study of the foregoing type has been reported in [96], using a UAS, on a semi-arid ecosystem. 3D height models of landscapes with cm-level spatial resolution can be obtained for above-ground biomass prediction from canopy volume. Relatively high coefficients of determination, with r^2 within the range 0.64 to 0.95, were achieved in the aforementioned study, which were sensitive to vegetation structures. The latter is beneficial for monitoring temporal or spatial environmental changes over areas with discontinuous canopy covers.

4.6 Terrestrial Photogrammetry

Landslide monitoring can be performed using terrestrial photogrammetry as well. In [264], a local computer and two reflex cameras were used to generate Digital Surface Model (DSM)s, with the Mont de la Saxe landslide as a case-study.

In [268], the C-Pro system was proposed for coastal projector monitoring as a terrestrial photogrammetry framework, which projects photographs to a georeference plane. The roll and pitch angles were estimated according to the horizon curve.

4.7 Typical Applications of LS

With LS, topography or heritage monitoring can be performed more accurately than with traditional RS. Protection of the heritage and the costs of equipment shipping are considerable challenges in the foregoing context. In [200], large statues were digitally modeled. Historic evidence can be incorporated as well, so that according to the trends observed, the original dimensions and shapes of the sites could be estimated [220].

As regards mining, monitoring variations of underground areas, subsidences, headframes or landslides, coal gangue dump surveying, land reclamation regulatory, measurement and 3D reconstruction of open pits are among applications of LS.

In [152], two roadway modeling and tunnel engineering methods were proposed based on cylinder projection and Poisson equation, where the latter led to smoother results, although it was computationally heavier as well.

In [122], a spatial cable setup was described based on elliptic functions, according to cylindrical shape, following the Saint-Venant guidelines [296]. The model was then evaluated using LS.

4.8 MLS

The main applications of MLS are in extracting road information and street object detection. However, the applications developed in the foregoing context may be computationally expensive, and require manual interference.

Autonomous vehicle navigation using Binary Kernel Descriptor (BKD)s including Gaussian kernel estimation and binarization [327] is a major application of MLS. The widths and lengths of the road lanes can be found using classifiers such as Random Forest (RF), according to the markings and curbs.

In [326], using filtering methods was proposed for removing non-ground points. The curbs were categorized into three different classes, namely, slope change, point density and jump. Boundaries touching asphalt areas to one side and grass, vegetation or soil to the other one were reported to be challenging to deal with.

In [335], the non-ground points were detected based on a grid dividing the area into voxels, along the vertical direction. The curbs were detected using an energy minimization function.

In [313], the Significant Voxels (SigVox) method was proposed based on eigenvectors, and recursive subdivision of point clusters through the octree method [228]. PCA was utilized in order to find the strongest eigenvectors in each voxel, which were then mapped to the relevant triangles from a icosahedron approximated using a sphere, in different scales, depending on the complexities of the target shapes. On a 4 km test road, they achieves an accuracy of 94%.

In [334], street lighting poles were detected from the data returned by a RIEGL VMX-450 MLS system [151], clustering the data and performing segmentation through voxel-based Normalized cut (Ncut) [280], followed by statistical analysis of the geometric features.

4.9 TLS

The main applications of TLS are in making documentations of archaeological sites, indoor scenes and urban environments, which will be discussed in this section.

In [332], the virtues of earthbound 3D LS, including reduction of the manual labor required, were demonstrated. In [182], the PIXEL-PHOTO software was used to compare the performance of photogrammetric stereo matching techniques with DSLR cameras, using the Byzantine Land Walls of Istanbul as a case-study.

In [255], a RIEGL VZ-400i [2] was utilized, in combination with a Nikon D700 possessing a 14-mm lens as the color camera, in order to make a model of the Temple of the Sacred Tooth, through the RiSCAN PRO [11] software. Moreover, a structured-light scanner, namely, Artec MHT [6], was employed for capturing high-level details. The scanning software Artec Studio [7] was then used for triangulation, followed by denoising the results through the Geomagic Studio [15] software. Finally, AutoCAD [207] and the Add-on Pointsense Heritage from Faro/Kubit [14] were used to obtain drawings based on the refined models.

In [162], sample geoarchaeologically important sites of Spain, Egypt and Jordan were documented, by means of an RTK-GPS [290] and the LMS-Z420i from Riegl [273], followed by georeferencing the point clouds [156]. The task of registration was performed using local surveying networks [78]. Path tracing was considered for simulation of lighting, estimating ceiling thickness of the caves and volumetric determination, where the minimum measuring distance, the time-consumption and the weight of the device were reported to be the main limitations. The results of such an approach can be imported into 3D GIS [189] as well.

In [199], the SketchUp 3D modelling software [320], the Leica Cyclone 7.3 point cloud processing software [18], being used for denoising and aligning multiple point clouds, and the Leica ScanStation C10 [32], being a medium-range LS device operating at 0.1~300 m distances, were compared in terms of accuracy.

In [48], LiDAR laser scanners were utilized for detecting segmenting out and classifying objects. Building facades were modeled using binary range imaging, followed by dividing the remaining amount of data into voxels and super-voxels. Classes of objects such as car, pedestrian and tree were recognized according to features, using trained, boosted decision trees. The facades were reconstructed by means of the ShadVis algorithm [125], where the point clouds represented the rest

of the objects were fitted to fixed 3D mesh models. The NAVTEQ database [343] was considered for testing the proposed method.

4.10 ALS

ALS has been widely utilized for mapping vegetated areas and forests, weakly-textured surfaces, narrow, long features and small objects such as power lines, roof-top heights for communication antennas, urban planning, creating DSMs of urban areas and fast-response applications.

In [130], rockslides and landslides were analyzed based on breakline tracking and image correlation, with the latter being reported to be more accurate for areas with high changes of surface patterns, on the basis of displacement vectors.

In [306], forest canopies were reconstructed with a resolution of $400\text{-}900\text{ pixels.m}^{-2}$, and a density of $0.6\text{-}0.8\text{ points.m}^{-2}$, using numerical optimization [98], computational geometry and topological connectivity, resulting in ordered sets of simplices. After filtering out canopy voids, they divided the point space into weighted simplicial complexes, and then quantified it.

In [306], through observing the persistent homology, tetrahedral volumes were calculated with degrees of determination (R^2) ranging from 0.65 to 0.78, depending on the predictors taken into account. Field measurements were thought to improve the filtration and the homology persistence of stem volume point clouds. Leaf Area Index (LAI) which is an indicator of the energy exchanged between the atmosphere and the land surface, was calculated in [225]. Despite Beer's law-based approach [293], the foregoing methodology generates shape estimates as well.

4.11 Phase Comparison, Structured Light and ToF

3D scanning technologies taking advantage of phase comparison, structured light or ToF are advantageous in terms of costs and ease of use. Methodologies making use of turntables mostly require calibration processes employing markers or similar tools, which is necessary for maintaining the rotation axis across different configurations. As another consideration, the accuracy of inter-beam points may affect the resulting precision significantly.

Phase-shifting was used in [316], by means of stereo and active illumination. The system they proposed is capable of handling multiple moving objects, at a frame-rate of 17 Frames per Second (FpS), using correlation-based calculations. The motion error was analyzed as well, in order to eliminate artifact motions.

In order to alleviate the typically limited lines of sight, and consequently, the limited coverage, possessed by structured-light 3D scanners, in [36], it was proposed

to use a stereo pair of cameras, along with a non-calibrated projector, which allows to resolve the foregoing issue rather effortlessly.

In [95], a probabilistic multi-scan alignment method was proposed in combination with 3D super-resolution, being applied to the data captured from different viewpoints around an object, using a ToF sensor, possibly suffering from systematic bias and random noise.

In [37], a near-InfraRed (IR) structured-light 3D scanner was proposed for objects smaller than 1 m^3 , with a 18th-century doorway in Benedettini monumental complex in Catania (Sicily) as a case-study. In order to alleviate the lack of precision, the scanning was performed in 23 segments, with 30% overlaps between adjacent pairs of frames. The result was also compared with a reference model made by an HDS 3000 by Leica Geosystem ToF laser scanner [263], which showed discrepancies within the range 5 to 9.6 mm.

Kinect sensors [12] have been utilized as affordable means of 3D scanning during the last few years. In [120], the problem of making 3D reconstructions of deformable objects was investigated, aiming at obviating the necessity of having access to prior knowledge, e.g. initial 3D models, or making the object stand still. Dense nonrigid BA was employed for correcting the drifts and evenly distributed errors calculated based on the loop closures detected. As its disadvantages, the proposed method might not be able to handle complex changing topologies, and is time-consuming: 9~10 hours is required for processing 400 frames on a 8-core 3.0G Hz Intel Xeon CPU, and with 64 GigaByte (GB) of memory. In [284], 3D miniatures of persons were created through filming them using a Kinect sensor, while being rotated on a swivel chair, and then printing the resulting models. A real-time performance was achieved employing SDF [70], where the holes caused by occlusion were automatically detected and filled. A hollow, watertight shell was utilized in order to reduce the 3D printing costs, where the quality of the result depends on various factors, including, the rotation speed, deformations, self-occlusions and the camera pose.

Making 3D reconstructions of translucent objects using phase shifting and polarization, especially if comprising of different segments of non-homogeneous materials or colors, is principally challenging, since while trying to do so, due to subsurface scattering, the intensity peak is shifted beneath the surface, which results in errors [81].

A wideband 45° polarized electrical scanning antenna array was proposed in [154], for metallic 3D printing, using a ridged horn, resulting in a loss of less than 0.3 dB. In [46], a 3D digitizer was proposed for metallic objects, based on Scanning from Heating (SfH), using an IR camera and laser. The proposed method was shown to be less accurate than deflectometry and polarization imaging, but applicable to larger sets of images. Moreover, it was proven to be less sensitive to the roughness

of the surface, compared to systems developed in the visible domain.

4.12 Comparison of 3D Scanning Techniques

A wide range of 3D scanning techniques and devices exists, elements of which may be categorized or compared from different perspectives. Some of the foregoing evaluation methodologies will be reviewed in this section.

4.12.1 Comparison of 3D Scanning Devices

In [60], 3D scanning technologies based on phase comparison and ToF were categorized into three classes, namely, with ranges represented by the intervals 0.1 to 1, 1 to 10 and 10 to 100 m. Point acquisition speed was considered as a criterion for evaluating the practical usefulness of 3D scanning devices. FoV, which is usually limited to 40° by 40° , constitutes another major characteristic of a 3D scanning device, where a wider FoV will lead to obtaining more information from a single shot. High-quality cameras can be incorporated into the scanning setup, which brings the advantage of obtaining a photo-realistic view of the scene or object of interest. Additionally, light weight and portability can be fundamental advantages of a 3D scanner, which may preferably come along a portable power generator, as opposed to a power line supply. Moreover, the software associated with a 3D scanner can greatly affect the ease of use, which should be kept simple and fast, and make it possible to compare the time-consumption incurred by scanning different scenes and objects.

4.12.2 Comparison of ALS and Photogrammetry

In [51], the fundamental differences between ALS and photogrammetry were investigated, whose results have been summarized in Table 4.1.

4.12.3 Comparison of Photogrammetric 3D Modeling Software

In [40], different photogrammetric 3D modeling software, including VisualSfM [321], Bentley ContextCapture [8], MicMac [22] and Agisoft Photoscan [307], were compared with each other, as well as with LS, in terms of making models of urban areas and buildings, according to the Cloud to Cloud (C2C) distance. Based on the 150 photos they took and examined, it was inferred that ContextCapture and Photoscan perform similarly to TLS, with Photoscan being more reliable in the sense of the soundness of the resulting geometrical interpretations. MicMac was reported to be geometrically accurate as well, with the drawback of providing

Table 4.1: Comparison of ALS and photogrammetry. The Table has been adopted from [100].

ALS	Photogrammetry
Active	Passive
Polar geometry brought by point sensors	Perspective geometry brought by frame or linear sensors
Point-based sampling	Area coverage
Direct acquisition	Indirect acquisition
Low-quality monochromatic images or no imaging	High-quality multispectral images (radiometrical and geometrical)
Height of flight over 6000	Height of flight over 1000
Less flexibility and variability	High variability: <ul style="list-style-type: none"> • Multiple-line, linear and 2D approaches and point detectors • Spectral, radiometric or geometric resolution • Non-metric, semi-metric or metric accuracy assessment
Lower flying speed	Higher flying speed
In principle, fully automatic	Demanding manual interference

possibly incomplete or noisy point clouds. VisualSfM was observed to provide the noisiest and geometrically least meaningful results.

4.12.4 Analyzing Vegetated Areas and Forests

In [150], UAV was shown to be more reliable than TLS for photogrammetric purposes, which was realized through detecting land relieves based on high-precision data representing mappings of anthropogenic and natural subsidences and uplifts, where the areas whose heights had been extremely affected by vegetation were appropriately detected and handled. A rod-mounted reflector and a tacheometer were utilized for capturing dense data.

4.12.5 Comparison of DEM Methods

Geomorphology depends on DEM for detecting topographical changes. In [117], global- and regional-scale and LiDAR DEMs were evaluated and compared, using a task of analyzing rugged topography over coalfields of southern West Virginia,

USA. It was then inferred that Global DEM (GDEM)s and Shuttle Radar Topographic Mission (SRTM) DEMs are more informative and reliable than LiDAR DEMs, in terms of modeling surface topography.

In [256], DEMs provided by TSX add-on for Digital Elevation Measurement (TanDEM-X) were inspected, with respect to their counterparts produced through LiDAR. It is TSX's twin, which is a German satellite for Earth observation, utilizing Synthetic Aperture Radar (SAR) technology, acquiring interferometric information in bistatic mode. Overall an approximate RMSE of 5 m was observed between the aforementioned two.

In [144], DTMs standing for ground elevation, which had been obtained through aerial LiDAR [308] and restitution of altimetric features [297], were compared, based on vegetation covers of the island of Tenerife, Canary Islands, Spain, at a nadir density of $0.8 \text{ points.m}^{-2}$ and a pixel size of 5 m. Using nonparametric TS, they acquired and investigated three elevation profiles, where the accuracy was between 0.22 m and 3.20 m and between 0.54 m and 24.26 m for the LiDAR-based and photogrammetric method, respectively, at the 95th percentile. Overall, the foregoing quantities show that compared to photogrammetry, LiDAR leads to a more accurate model representing the elevation, especially where the surface is not visible to the photogrammetric device, e.g. in the case of Canarian pine forests.

4.13 Combinational systems

Due to the fact that each of the 3D scanning methods described in the previous sections brings certain advantages and disadvantages, combinational systems have been proposed as well, which try to find a compromise between the benefits and limitations of the underlying components. In this section, some of the most prominent frameworks of the foregoing type will be reviewed.

In [178], Site Index (SI) was measured as an indication of site productivity, which is required for forest management. SI is calculated for each tree species, according to the height and age of the largest tree. They used RS information resulted from combining airborne hyper-spectral data and ALS, on the basis of Individual Tree Crown (ITC) delineation. Additionally to the species, age and height, Diameter at Breast Height (DBH) was estimated at ITC level, per 400 m^2 .

In [163], spaceborne SAR and digital stereo imagery data were combined for obtaining measures of forest inventory attributes, according to a forest area in Finland. DTM- and LiDAR-based ALS data were compared as well. The manner of selecting an option from the foregoing list or a user-ready utility providing DSM data was shown to depend on the user's preferences, being affected by the costs of acquiring and preprocessing the data.

In [129], the performances of Panoramic Spherical Photogrammetry (PSP), SfM and LS were tested and compared, based on the Cham towers in Vietnam. LS data were acquired by means of Z+F IMAGER 5006h [229], and then refined through the Cyclone software [138]. The SfM procedure was employed through the Agisoft Photoscan software [307]. Moreover, PSP models were obtained utilizing two Cannon cameras with resolutions of 12 and 14 Mega-pixels. The PT-Gui 9 [28] software was then used for stitching the images, followed by performing bundle block adjustment [179]. It was reported that the LS-based process is more accurate, but less convenient to use, than the one based on PSP.

In [329], ToF TLS, stereo digital photometry, monoscopic multi-image evaluation, TLS and digital photogrammetry were used for making documentations of historical building facades. A Nikon D100 DSLR camera [23] was utilized to acquire images. The Computer Aided Design (CAD) utilities embedded into the MicroStation software [345] were employed for generating 3D photogrammetric line drawings. For digital photogrammetry, the Z/I Phodis ST30 software was applied to the Fatih Mosque facades. Moreover, Muayede (Ceremonial Hall) of the Dolmabahce Palace was surveyed using ToF TLS through the LMS-Z420i from RIEGL [273], with a Nikon D70s calibrated digital camera [145] mounted on it. Last but not least, DOMs were made by means of the RiSCAN PRO software [11]. In [240], a system was created based on a combination of TLS and UAV, which was tested on the Roman theater in Ventimiglia, Italy. The drawings were made at a Ground Sample Distance (GSD) of less than 4 mm, and a scale of 1:20. The UAV was used to create both oblique and vertical images. Separately, terrestrial images of vertical structures were acquired, which were then converted to dense point clouds, being repeated based on different ground controls.

In [254], MLS data were acquired, refined and classified, in order to create point clouds representing streets. At a later stage, Mobile Mapping System (MMS) and terrestrial images were obtained, registered onto the aforementioned point clouds, and then projected onto the images, through weighted window-based z-buffering. The geometrical consistency between the stereo images and point clouds was checked through re-projection and rectification. The temporal changes of the point clouds were finally detected according to color, depth and class information, using over-segmentation-based graph cut optimization [208]. Illumination changes and co-registration errors were appropriately handled as well.

In [274], Relative Vertical Accuracy (RVA) analysis was used for assessing the performance of ALS DEM. VHR ALS Earth models were also evaluated based on TLS reference models and visual representations. In building, grass and open areas, the Absolute Vertical Accuracy (AVA)s of VHR ALS DSMs were observed to be distributed with normalized median absolute deviations between 3 and 5 cm, being 35 cm in high-vegetation ones. The AVA of ALS DEMs repre-

senting canopy areas was reported to qualify for a 1/1000-scale topographic map. In [88], stereoscopic and Fourier fringe analyses [63] were combined for more accurate and denser point clouds, with a reasonable load of computations.

4.14 Post-processing

Removing outliers and smoothing the surface are essential post-processing procedures which need to be applied to point clouds resulted from 3D scanning techniques.

In [243], it was suggested that a plane can be fitted to the inliers falling within a neighborhood of the point under study, followed by determining outliers based on the majority of acceptable points within the aforementioned neighborhood. Reference Z-score (RZ-score), as a distance-based measure, was calculated for detecting outliers, as follows:

$$Rz_i = \frac{p_i - \text{median}(p_i)}{MAD}, \quad (4.1)$$

in which MAD denotes Median Absolute Deviation (MAD). Outliers were then determined based on finding out observations z_i associated with Rz-scores greater than or equal to 2.5.

As an alternative, they calculated Mahalanobis Distance (MD), as follows:

$$MD = \sqrt{(p_i - \bar{p})^T \sum^{-1} (p_i - \bar{p})} \quad (4.2)$$

where \sum and \bar{p} represent the covariance matrix and mean of the sample, respectively. Outliers were then detected based on MD being greater than 3.075.

Local surface fitting and normal estimation were performed in [241], using DR PCA (DRPCA), which takes advantage of robust statistical techniques. As the first step, Robust Distance (RD) is applied in order to make the data more homogeneous, through removing candidate outliers. Next, more outliers are detected, and a local surface plane is fitted, by means of Robust PCA (RPCA), outperforming RANSAC, M-estimator Sample And Consensus (MSAC), Demixed-PCA and classical PCA, based on the tests they performed on simulated and actual data. Similarly, in [242], statistically robust methods, RPCA and the Deterministic Minimum Covariance Determinant estimator [169] were utilized for smoothing the surface.

4.15 Conclusion

In order to provide information required for deciding on the steps needed to be taken for making 3D garment reconstructions, this chapter presented a review of the 3D scanning methods and technologies proposed in the existing literature, which include airborne, terrestrial and mobile LS, terrestrial, SfM, aerial and close-range photogrammetry, and phase-comparison, structured-light and ToF approaches. Some of the comparative studies and combinational systems were reviewed as well, along with post-processing tasks such as removing outliers and smoothing surfaces.

CHAPTER 5

3D SIZE-ESTIMATION BASED ON THE GEODESIC DISTANCE MEASURED BY PHOTOGRAMMETRIC DEVICES

5.1 Introduction

The remaining steps of the thesis consist, mainly, of 3D modelling and visualizing human body for the purpose of rendering human body models, instead of the actual, physical mannequin robots, while wearing the clothing item, and if necessary, try to represent them in different configurations, e.g. with various translations and rotations.

As an example application aimed at the above purpose, the software MakeHuman [19] can be mentioned, which, apart from the source code being written in Python, provides a Graphical User Interface (GUI), enabling manipulating the appearance of the body through changing the shape and size of numerous parts of the body. One of the virtues of the latter software is that it provides the opportunity of making the initial avatar, and further modifying it in Blender [9], i.e., there are tools for linking the two software, and deciding which parts or specifications to modify in each. Besides, it is possible to animate the body in blender.

The existing 3D body models might not provide realistic representations while given a certain number of measurements, i.e., although the required sizes and girths are successfully produced, the proportions of the neighboring parts, influencing the overall appearance and shape, do not still correspond to the reality, which necessitates either incorporating more measurements, which is tantamount to considering more details, or changing the geometric structure of the avatar. Besides, one of the important steps would be trying to change the construction of the avatar itself if necessary. More clearly, there are a lot of details affecting the appearance and quality of the avatar, for example, the number of the edges

and vertices, as well as their locations inside the avatar, and the bones utilized for modifying the shape and size, each of which has to be optimally determined in order to obtain the best possible representation of the human body in terms of being realistic, robust and flexible.

The problem of approximating 3D sizes and geodesic distances has been investigated from numerous perspectives by researchers. For instance, some of the strategies from the latter list are developed on the basis of binocular stereo vision [195]. Some others are concerned with devising indices representing anthropomorphic biometric characteristics via home-based imaging approaches [209], which requires taking images of the body with different orientations of the camera, and subsequently, reconstructing the whole body by means of image processing techniques with the consistency between the different segments of the silhouettes as the objective of the optimization algorithm. It is worth noticing that the depth information could be obtained through stereo-vision or structured-lightening techniques [231] even if the existing camera is not able to provide it readily.

In the existing literature, numerous studies have investigated the concept of reliable vision-based 3D measurement estimation, each of which discusses the related technical and theoretical complexities from its own point of view, based on the associated experimental setup, i.e., the approaches proposed heretofore would apply to the specific practical settings considered in the context of the corresponding study, but not necessarily to others.

For devising measurement algorithms functioning properly under the conditions sought from the outset, due attention should be paid to various physical aspects and requirements of the particular devices supposed to be utilized. In what follows, a summary of the details, along with the outstanding advantages and disadvantages, of the most recent studies carried out in the foregoing field is provided, in order to illustrate the current state of the art, as well as to clarify the prospective contributions of the strategy proposed in this chapter to the ones that are either directly connected with it or could well benefit from appropriately extending it.

One of the studies that has tried to develop a methodology aiming at solving the range-finding problem is [195], which concentrates, particularly, on taking the specifications of mobile robots into account, by means of DSP, through utilizing the so-called “binocular stereo vision”. The algorithm is devised such that first paired images of the target objects are captured, which are processed using Gaussian filter and improved Sobel kernels. Afterwards, employing feature-based local stereo matching, the location of the object is determined.

Confidence and left-right consistency filters are further applied to the result of the foregoing procedure, in order to reduce the possibility of the existence of mismatches, which will enhance the quality of the output. The whole algorithm is implemented using the DSP/BIOS operating system, which will enable real-time

control over the position detection process. The point-to-point distance calculation procedure, in ranges up to 120 cm, has, reportedly, demonstrated an accuracy of 99%, while taking 39 ms, in the worst case, for performing the task, which could prove suitable in the associated context.

In [298], poor perception of body shape, i.e., invalid thoughts about the proportions, sizes and distances, has been attributed to disorders such as anorexia and bulimia nervosa. Nevertheless, such misconceptions are reported to be difficult to formulate and represent numerically. In the foregoing study, it has been stated that the previous ones had not provided sufficient analysis of the trend of the changes in the body shapes, e.g. in the parts where the level of fat changes considerably. More clearly, the shortcoming lies in the fact that, despite the simplistic nature of most of the approaches proposed so far, the changes of the body shape, normally, do not represent trends that could be resembled by consistently stretching the surface in either of the directions, as proposed by the “Distorting Video Technique (DVT)”. In other words, the body shape shows different trends of changes in all the parts, which differ in terms of amount, type or complexity, and cannot be simulated using simple linear scaling.

The above study tries to tackle the aforementioned problems, and achieve a more realistic way of representing the body shapes, through considering the available biometric information and performing separate analyses on each body part, which enables to adjust the shape and size of each segment as desired. The results of the foregoing study, which include Perimeter-Area Ratio (PAR), and subsequently, Body Mass Index (BMI), will provide the opportunity to verify the authenticity of the body shape representation perceived by the user by means of comparison. As an outstanding outcome of this study, it has been concluded that the majority of the users tend to overestimate their sizes, where the actual value of BMI is around 15 in the case of anorexics, which is the boundary between underweight and emaciation, but is deemed 20 by the instances.

Anthropometric body measurements have been investigated from another perspective in [209], along with the associated ratios, which influence the perception of the body shape and fitness delivered by the user. The main goal of the foregoing study is to devise a home-based imaging system capable of extracting anthropometric body measurements, which consists in capturing multiple images of the same object in front of the camera, or *vice versa*. Subsequently, the entire body can be reconstructed through applying image processing techniques on various silhouette segments. Finally, another processing step is implemented, which improves the consistency between the different parts, taking an energy function into account as the criterion.

Another solution to the problem of estimating 3D distances is proposed in [231], which makes use of stereoscopy in order to create a visual representation on the

basis of two images taken from slightly different positions. The latter can be realized through either employing two separate cameras, each of which possessing a single lens, or utilizing a stereo camera with a pair of juxtaposed lenses. The main application of the foregoing method would be to obtain the approximate distance between the camera and the object of interest. The latter might involve numerous physical parameters and constants, such as the focal length of the camera or the distance between the two cameras, in the case of using two separate cameras. The procedure is such that first, the distance of the object from one of the cameras is calculated, and then the value associated with the other one is estimated through minimizing the dissimilarities that inevitably exist between the two aforementioned values.

As another example of the studies addressing the problem of 3D size estimation, [188] can be mentioned, which aims at developing a strategy for automatic, computer-based extraction of body sizes and measurements, again according to the anthropometric data. The resulting utility is supposed to be applicable to many industrial contexts, such as security, medicine, sport and clothing. Reportedly, it has been meant not only to minimize the annoying characteristics of the measurement system to the subject being measured, but also to maximize the accessibility, ease of use and portability of the service to the operator in terms of efficiency and reliability, for the specific case-study of garment industry.

It should be noted that the applications of the problem of 3D size estimation are not limited to extracting human body sizes and shapes, i.e. there are numerous fields of industry and research that benefit from the resulting utilities. For example, structural analysis of masonry arches, which are ubiquitous in the sites including architectural heritage, can be mentioned. Such structures could be still analyzed using 3D laser scanners, which are rather cumbersome and expensive. One of the important issues that has been addressed in the foregoing study is that real geometry, rather than the idealized one, has to be used, in order to enable the proper use of 3D surveying techniques, which will lead to more realistic results. The proposed methodology is based on TLS, which, as aforementioned, tries to come up with realistic and reliable 3D measurement estimations.

Another field of study and technology that could significantly benefit from the development of 3D measurement techniques is underwater quantification and measurement. For example, the methodology proposed in [187] can be mentioned, which deals with a laser/video system consisting of a frame, a roll/pitch motion reference sensor, a video camera and three micro-lasers. In the context of the foregoing study, the orientations of the micro-lasers are taken into account for the purpose of triangulation, which results in the calculation of the position of specific objects within the field of view of the camera. Besides, on the basis of the foregoing information, as well as the one reported by the motion sensors, the

magnification is determined throughout the whole field of view. Subsequently, the necessary parameters, such as perspective overlays, the distance from a particular point or location, the scale in each part of the image and measurements of area can be estimated through employing a prescribed bunch of image processing techniques.

One of the important properties of the above study is that any characteristic of the whole system, including “laser wavelength and power, camera sensitivity and resolution, dynamic range and mounting geometry”, could be adjusted as desired, in order to accommodate the specifications and particular requirements of the application at hand, as well as the level of accuracy sought from the outset. The system has been reported to be flexible enough to be used in any sub-sea application, device or platform, since it performs at a comparatively low cost, in an automatic manner, for the general purpose of acquiring quantitative data from the standard underwater video streams. The associated software, moreover, has the virtue that the video information could be post-processed in real-time.

The first specific application of the utility resulted from the above study is estimating the number of ground-fish species within a “transect area”, which are either commercially or recreationally used. The benefit brought about by employing such technique is that the density of particular fish species throughout a “high-relief area” can be estimated directly, but non-intrusively, which is not affordable by the existing typical techniques.

In [224], the depth information obtained by the sensors is analyzed for extracting biometric soft indicators, such as lengths and girths. Supervised training is considered while making use of multi-part pose clusters. Afterward, the process tries to iteratively match the 3D body shape descriptor model, which outperforms RF [172] in the sense of demanding less amount of training data.

From another perspective, [190] deals with the problem of detecting hands based on RGB-D information of the upper part of the body for Human-Computer Interaction (HCI) purposes, which is affected by the changes in illumination and viewpoint, and involves overcoming variousness of the possible appearance of the hand due to the high flexibility of the hand, especially, at the wrist part. The underlying assumption, which has been verified through conducting relevant experiments, is that the hand landmarks are always located in certain positions with constant geodesic distances from the an anatomic reference point, which is also automatically detected. The methodology consists in segmenting the human body on the basis of the depth information, and afterward, obtaining a graph representation including the geodesic paths originating from the aforementioned reference point, $G^t = (V^t, E^t)$, where $V^t = \mathbf{B}^t$ denotes the vertices, and $E^t \subseteq V^t \times V^t$

stands for the edges, such that:

$$E^t = \left\{ (\mathbf{p}_{ijk}, \mathbf{p}_{i'j'k'}) \in V^t \times V^t : \|(i, j, k)^T - (i', j', k')^T\|_\infty < 1 \right\}. \quad (5.1)$$

It should be noted that in the above formulation, $\|\cdot\|$ returns the infinity norm of its vector component, and $(i, j, k)^T$ and $(i', j', k')^T$ denote the 3D coordinates of the points \mathbf{p}_{ijk} and $\mathbf{p}_{i'j'k'}$, respectively, in \mathbf{B}^t , which, if being neighbors, are connected with the edge $e = (\mathbf{p}, \mathbf{p}') \in E^t$, corresponding to the weight $w(e) = \|\mathbf{p} - \mathbf{p}'\|_2$, on which basis, the geodesic distance between the two points is defined as follows:

$$d_G(\mathbf{p}, \mathbf{p}') = \sum_{e \in EP(\mathbf{p}, \mathbf{p}')} w(e), \quad (5.2)$$

which includes the edges along the shortest path between the points, being determined based on the min-path Dijkstra's algorithm [56]. Moreover, in order to achieve more tangible and consistent geodesic paths and avoiding inauthentic edges of the neighboring body parts, the dense optical flow vectors are utilized. Again, the computational complexity and costs are rather satisfactory, and the geodesic distances have proven helpful for the goal sought from the outset.

The rest of this chapter is organized as follows. In the next section, the notion of the proposed method is described, both literally and mathematically. Afterward, the results of the implementation of the foregoing methodology are visually presented in various scenarios, along with discussion on the reliability of the method, and the authenticity and accuracy of the numerical values. Finally, the chapter concludes through summarizing the main points discussed and presented throughout the chapter, and providing a set of hints as to how the approach proposed in the chapter can be extended and followed up in the ongoing studies.

5.2 The Proposed Size Estimation Algorithm

As the first stage of the 3D measurement method proposed in this chapter, the coordinates should be mapped from the screen to the real-world ones. For the latter purpose, a variety of approaches are available in the literature. For example, optical depth cameras can be used, most of which are capable of performing the task automatically. Besides, 3D reconstruction may be considered, as mentioned in Section 5.1, through stereo-vision or structured light.

Assuming that the above goal is accomplished, a function is present in the form $f : \mathbb{N}^2 \rightarrow \mathbb{R}^3$, which handles the mapping module, as follows:

$$f(x, y) = (X, Y, Z), \quad (5.3)$$

in which $x, y \in \mathbb{N}$ and $X, Y, Z \in \mathbb{R}$ stand for the screen and real-world coordinates, respectively.

Every measurement requires a set of starting and end points as P_s and P_e , respectively, which are specified based on the user input. A naive solution for calculating the distance between the foregoing two is to make use of the Euclidean one, which takes the real-world coordinates corresponding to them into account, as follows:

$$d_E(P_s, P_e) = \|f(P_s) - f(P_e)\|. \quad (5.4)$$

The above framework, however, does not lead to results correlating with the real geometry, in the most cases, i.e. the value reported by such a formula is correct only where a straight line lies on the surface that connects the two points. To overcome the latter shortcoming, a more realistic path could be drawn between the points, being obtained while considering the nature of the surface, including the ripples, convexities and concavities, etc., which can be formulated as follows:

$$d(P_s, P_e) = \int_C h(X, Y) ds, \quad (5.5)$$

through integrating numerous portions of the whole path, which should be as small as possible, and connect the beginning and end points step by step. In the above relationship, C denotes a path from $f(P_s)$ to $f(P_e)$, being resulted from projecting the straight line $P_s P_e$ on the screen onto the real-world coordinates standing for the surface points. Moreover, $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ represents a function responsible for outputting the Z coordinate of each projected surface point, which takes the pair (X, Y) as the input.

It is worth noticing that, at least in the context of this chapter, the smallest possible components of the path being integrated are determined based on the differences between the real-world coordinates of the points represented by juxtaposing the pixels on the screen, which is performed by the aforementioned map.

Geometrically, a weighted graph is considered, in which every pixel is connected to four nodes resulted from the foregoing projection. The mapping could be thought of as a weighted graph, such that each pixel, being considered a node, is connected to its four neighbors, which are all mapped from the screen coordinates to that of the real world. As a major difference between the methodology proposed in this chapter and the approaches suggested in the previous studies, in order to achieve physically sound results, while calculating the weights, the contribution of the Z -coordinate value is discarded, i.e. only the X - and Y -coordinates are taken into account, which is also the main difference between the foregoing definition of distance and that of the Euclidean distance.

For the purpose of determining the shortest possible path connecting the beginning and end points, the Dijkstra Algorithm is employed, followed by piecewise

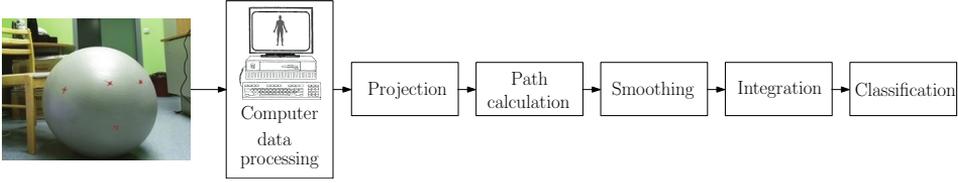


Figure 5.1: A sample measurement scenario, with the red marks representing the beginning and end points, accompanied with a schematic diagram summarizing the main modules of the proposed method.

calculation of the distance based on the foregoing path. Assuming that p_n stands for the n^{th} point visited while iterating through the path, the overall distance, D , could be reliably estimated as follows:

$$D = \sum_{n=2}^m \|p_{n-1} - p_n\|, \quad (5.6)$$

where m is the total number of the points visited.

The underlying fact is that the aforementioned path is determined in such a way that it would resemble the shortest real 3D line lying on the surface, despite other studies reported in the literature that mostly consider the Euclidean distance between the points, which, due to the convexities and concavities that normally appear on the faces of almost all objects, does not necessarily lead to the actual measurement supposed to be calculated.

As a relevant point, the body surface does not usually present sudden change in its depth, being denoted by the Z -coordinates. Otherwise, it is considered wrong information, i.e. noise, which might have arisen from various grounds, including electrical or mechanical fault. To be more clear, since extreme fluctuations of the body surface are not expected to occur, in the case they appear, they should be properly denoised, which would, otherwise, lead to undesired overestimations of the sizes.

In order to reduce the chances of the above negative impact, averaging of the path is considered, which maintains the general trend of the changes in the body surface, but ignores ripples smaller than a certain threshold, which are, almost certainly, caused by noise, malfunction or miscalculation. Fig. 5.1 shows a sample scenario, along with a schematic diagram summarizing the fundamental elements of the proposed 3D measurement procedure. The accuracy accomplished, on the basis of the exhaustive experiments conducted, is lower than 0.5 cm, while using Kinect 2 as the optical depth camera.

The main contribution of the proposed method is making it feasible to grasp a reliable approximation of the essential body sizes and 3D geodesic distances,

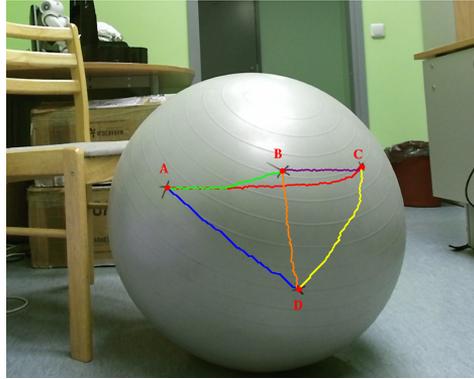


Figure 5.2: A ball with markers used for measurements.

Table 5.1: Results of the Measurements

Points	Proposed Method (cm)	Manual Measurement (cm)
AB	20.87979	20.7
AC	37.34988	37.2
AD	28.36199	28.0
BC	16.29063	16.2
BD	19.71188	19.3
CD	24.80985	24.6

for example, bust, waist and hip girths, via optical cameras, rather than 3D scanners, which typically, impose considerably more restrictive physical limitations in practical and commercial contexts.

5.3 The Preliminary Tests

The proposed method was tested in several scenarios, one of which was a large ball with marks on it, as shown in Fig 5.2, where some others are also illustrated in Fig. 5.3. In the former, there are four points marked on the ball and measurements were performed between all of them. We refer to these four points by A , B , C , and D . The results of the automatic measurements using the proposed algorithm are compared with measurements obtained by manual measurements using a measuring tape. These results are shown in Table 5.1.

It can be seen that the measurement results of the proposed method are very accurate and are close to the results that were obtained manually. The maximum difference was only 0.41cm, while the average difference was 0.23 cm. On average, the accuracy of the proposed automatic measurement for different scenarios



Figure 5.3: Illustration of some of the experimentation scenarios.

is 0.25 cm. Hence it is concluded that the proposed method can be used to measure distances along surfaces with high accuracy.

5.4 The Experiments, Results and Discussion

The experimental results required taking RGB-D images using Kinect 2, analyzing the visual data, extracting the measurements, and classifying the body shapes based on the size, according to the known gender. The images were taken through a simple setup with a white background from the people who were chosen such that they would be as diverse as possible in terms of the foregoing criteria, so that they could suffice for representing all the possible classes. It is worth mentioning that the images should be taken under circumstances where the depth information is not vanished, which would otherwise mean that the parts of the image undergone depth loss would appear as totally black.

One of the advantages of the approach suggested in this chapter is that it improves the ease of use and convenience significantly, as well as reduces the costs and hardware requirements, in that the images are obtained using a single camera,

while taking a single shot is enough. Most of the similar methods proposed in the literature so far demand making use of either a 3D scanner or multiple cameras, and in the most cases, multiple shots are required for gathering and processing the information needed for extracting the 3D distances and sizes.

Although the methodology introduced in this chapter is capable of calculating any arbitrarily specified 3D measurement, the experiments showed that the cross shoulder and torso width at waist values were sufficient for the aim and scope of this thesis. More clearly, the main goal of this chapter is to validate the 3D measurement approach in the sense of providing reliable information for classification of human bodies into the classes of small, medium and large, though it has been devised as a general-purpose method, which may have various applications of numerous sorts, some of which were discussed in Section 5.1.

Many researchers have been engaged in trying to fulfill the above task, which is vital for the aim of utilities such as virtual fitting rooms, and have proposed solutions approaching the problem from different perspectives. In [101], for example, it has been concluded that the most important measurements that could help make a mannequin robot resemble a body shape realistically are bust, waist and hip circumferences.

For the aim of this chapter, on the other hand, the problematic constraint was that the method was supposed to lead to the required classification information only through a single shot, which means that it is not possible to readily calculate, or even reliably estimate, the circumferences, since there is no access to the rear side of the body, i.e. there is only a frontal image. As a consequence, the choices of the sizes were limited to the ones that lied completely in the frontal part of the body, meaning that only widths and heights could be taken into consideration.

As shown in Fig. 5.4, the way the initial sets of widths and heights were selected was such that the ones which, according to our previous experiments and studies, had proven helpful were tried, most of which were located on the parts of the body that separate two or more juxtaposing segments, i.e. they stand for the most extreme changes in the shape and size of the cross-section contours, which are crucial for deciding whether the cloth, at least, would not be simply too large or too small when it comes to satisfying the minimum requirements of fitting the body.

Further exploration of the usefulness and informativeness of the aforementioned measurements led to the conclusion that, among the ones shown in Fig. 5.4, the cross shoulder and torso width at waist band were enough for deciding on the class to which the body size belongs.

The above process included, mainly, trying out as many combinations of the conditions on the sizes as possible, along with different thresholds, and checking whether the results of classification on such a basis leads to acceptable results.



Figure 5.4: A sample image showing the placement of the imaginary markers for calculating the seven important body measurements on one of the instances, which include, among others, shoulder and waist widths.

Heretofore, this has been implemented rather manually, through considering picking up the conditions randomly, and adjusting the thresholds by means of trial-and-error after starting with an initial guess being determined intuitively. However, a more systematic way of dealing with this problem would be to find out the best possible solution through an optimization procedure or, if computationally possible, searching for it exhaustively, which is beyond the scope of this thesis. The proposed 3D measurement algorithm was applied to 156 instances, and seven important body sizes were calculated for all of them, where in 98% of the cases, the classification result corresponded to the one sought from the outset, meaning that both the geodesic 3D measurements and the classification procedures are reliable and almost accurate. As aforementioned, only shoulder and waist widths were taken into account for the classification purpose, which resulted in a reasonable performance. For the sake of illustration, the seven important measurements of a couple of sample body sizes, along with the associated size classes decided by the classification algorithm are provided in Table 5.2.

5.5 Conclusion

This chapter proposed a novel strategy for almost accurate estimation of body measurements and sizes based on the depth information provided by photogrammetric optical scanning devices. The procedure starts with constructing a map

Table 5.2: A set of sample body measurements, as well as the body size classes determined by means of the proposed classification technique. All the values are in mm.

Sizes⇒ Sample↓	S	s_1	s_2	s_3	s_4	s_5	s_6	s_7	Class
1	M	462	394	380	406	412	169	480	Medium
2	M	477	373	377	445	410	239	520	Large
3	M	437	361	368	363	392	209	448	Medium
4	M	424	351	353	359	370	174	493	Small
5	W	381	302	292	259	348	189	420	Medium

Legend

S: Sex

M: Man

W: Woman

s_1 : Cross shoulder

s_2 : Trunk width at bust

s_3 : Trunk width under the bust

s_4 : Torso width at waist band

s_5 : Trunk width at hip

s_6 : Chest height

s_7 : Upper-body height

relating the screen coordinates to that of the real world, and subsequently, integrating the geodesic distance gradients through iterating the connecting points on the path obtained via projecting the straight line connecting the beginning and end points on the screen onto the real line lying on the physical surface. Moreover, in order to overcome the negative effect of the high-frequency noise present in the determined path, which emerges in form of extreme fluctuations of the surface depth, a smoothing procedure is applied, which averages the path connecting the points throughout the line, and removes the undesired ripples. The methodology was implemented on an industrial application, namely, classification of human bodies according to size, as a case-study, which is utilized in the context of virtual fitting rooms employed for online try-on of clothes. The experimental results certified the credibility and reliability of the suggested approach, where the accuracy of the distance estimation was bounded to less than 0.5 cm using Kinect 2 as the optical camera.

CHAPTER 6

3D OBJECT AND SCENE RECONSTRUCTION

6.1 Introduction

Reconstruction of 3D objects and scenes has various applications in research and industry contexts, examples of which include Simultaneous Localization And Mapping (SLAM) [49] by autonomous systems [281]. It has undergone significant development during the last few decades [133, 341, 99, 336, 35]. Research communities from numerous fields, such as medical sciences and robotics, aim at replacing real, physical materials and processes by virtual alternatives, for the sake of exploiting both the practical and economic benefits, as well as the ease of use, brought about upon doing so [292, 251]. In other words, simulation and visualization of real-world objects and operations lead to representations and understandings of the associated phenomena which obviate the need to perform costly tasks required for creating, trying out, manipulating and maintaining them [102]. Although it may initially sound expensive, time-consuming and challenging, producing 3D models of materials and simulating the interactions taking place between them, in the most cases, prove to be able to pay off fruitfully in not a very long term, and overall, should be appropriate substitutes, given a fundamental condition, namely, that their results have to be close enough to the reality, so that they could be relied upon for analyzing or predicting the notions sought from the outset, having been based on empirical experiences.

One of the grounds on whose basis researchers choose a device to be used for collecting databases is their capability of filming large sequences quickly. The foregoing goal can be accomplished through employing, for example, the Asus Xtion PRO LIVE, which has a similar functionality to the Kinect, and is used in [324] for creating the Database of 3D Spaces, consisting of rooms and apartments, where the objects included in some of the frames are annotated after cap-

turing them. In [94], on the other hand, the aim is to create a database of 10000 RGB-D images of indoor spaces that would be fully annotated, by means of four different sensors, namely, Intel RealSense, Asus Xtion, Kinect 1 and Kinect 2.

A major application of 3D imaging is in investigation and verification of possible changes made to surfaces. In [128], for instance, a new method is proposed for capturing 3D data to be used for the forensic detection of bite mark injuries on the human skin. Reportedly, the latter strategy is more robust against operator mistakes and angular distortions which usually take place in the case of relying on 2D images. The latter improvement is brought by avoiding the reduction of the 3D data into a 2D coordinate system.

As another approach to 3D modeling, reconstruction of point clouds standing for objects or scenes on the basis of sequences of image taken from different view-points, in particular, has experienced considerable progress during the last few decades [133, 341, 99], especially because of the constant increase in the available processing powers and, more recently, the emergence of low-cost multi-purpose sensors, such as Kinect, which help obviate the necessity of focusing on compressing and coding the depth maps for storing, transmitting and visualizing purposes, especially when it comes to large amounts of data, e.g. 3D videos [340].

For example, the methodology proposed in [324] is meant to reconstruct a 3D scene based on RGB-D images taken by Kinect. First, the matching key-points between the consecutive frames are detected using SIFT [211], and the outliers are rejected using RANSAC [132]. Moreover, the loop-closings are detected according to a Bag of Words [131], which is followed by approximation and optimization of the poses via BA [217]. Then through SfM, the camera poses are estimated, and key-point descriptors are matched. They make sure that the frames that are essentially the same, or are more similar to each other than a certain threshold, will be matched and unified, followed by scene completion. Consecutive frames are matched, and the required transformation mapping each frame to the next one is calculated by means of RANSAC. Only the stronger matches are kept. Initial guesses for matching are chosen through Gaussian smoothing [319], non-maximum suppression [235], and dilation. The initial pose estimate is obtained through multiplying the matrices in the order corresponding to that of the associated consecutive transformations. Cross-bilateral filtering is also utilized for the purpose of refining the raw depth information, and voxelized using the Truncated SDF (TSDF) [171]. What remain are object annotation and BA. In this step, the errors are compensated for, and the final reconstructed 3D point cloud is improved, through labeling in an interactive object annotation environment. The objects are named, and their boundaries are more precisely determined by placing and adjusting the required number of control points, which are then propagated throughout all the frames. Afterward, possible occurrences of the same object

with the same pose in more than one frame are handled. In general, this method is supposed to fit the requirements of reconstructing large scenes, which can still be improved through making use of annotations adjustable by means of an online annotator. The point cloud will then be refined and used in order to build a 3D mesh as a model, which is followed by incorporating color, lighting and texture properties from the original 2D images into the model.

Determining the matching parts of the consecutive images necessitates using an algorithm such as ICP in order to find out the similarities between the depth images, and then creating a point cloud resulting from merging them. Ever since its introduction in [57], ICP has been one of the most common tools for 3D shape registration. It finds the rigid transformation between frames by minimizing the global error metric based on point-to-point distances. The most challenging part is finding the point-correspondences, which is done iteratively until the error metric cannot be improved anymore. Given overlapping 3D frames and a reasonable initial transformation guess, the algorithm usually leads to high-quality results.

Having been tried for the first time on medical scanning data [216], in accordance with the widespread of devices capable of extracting the depth information from the environment independently, the interest in using the latter information for construction of full 3D object or scene surface representations based on 2D slice scans of depth intensity has proportionally grown, since employing it enhances the chances of extracting the shape reliably and computationally efficiently.

The basic ICP algorithm was introduced in [82], and further enhanced in [57]. It takes the distance between the overlapping parts of the views into account as the objective function, i.e. tries to minimize it, which would geometrically be tantamount to finding the optimal transformation taking a frame to the other one, while, desirably, superimposing the common points of the former on their counterparts in the latter. A variant of the ICP algorithm that implements it with enhancements made to the manners through which the control points are selected and the aligned views are integrated is presented in [119], where, as a means to avoid noisy results, the captured data is first refined via smoothing, followed by the verification of the points, and for integrating the overlapping points of the views, a weighted average is utilized.

The original approach, i.e. the one proposed in [82], uses the *point-to-plane* metric as the criterion, which is, in fact, the summation of the distances between the data points and the tangent planes where the matched model points lie. Alternatively, in [57], the *point-to-point* metric has been employed, which aggregates the squared distances from the data points to those of the model, with a, reportedly, comparatively slower convergence, but a lower computational cost [266]. The ICP algorithm using the point-to-plane error metric in its original form may be too slow to handle practical use-cases. If the angular distance between each pair

of consecutive views is small enough, then an optimization process can be applied so as to alleviate the foregoing problem [218]. A thorough explanation of all the steps involved in the ICP algorithm is provided in [266], along with possible parameters to be quickly optimized and configured for most efficiently utilizing it under a set of prescribed experimental conditions.

Comparison and analyses of different aspects of ICP were provided in [266]. They also proposed a combined version of different approaches, which was optimized for speed. Recently, with the introduction of RGB-D video cameras, such as Kinect 2, ICP has been used for tracking the camera pose [171]. Since the movement between consecutive frames is typically small, ICP fits very well for the foregoing purpose. ICP can very well be used in combination with RGB features such as SIFT or Features from Accelerated Segment Test (FAST) [160]. In the case of doing so, the initial transformation estimation is performed using RGB features and RANSAC, and then fine-tuned by means of ICP. The approach can be utilized for constructing objects and scenes [194], where another important step is segmenting the object out from the background, if applicable, which can be improved by combining the RGB and depth information [194].

In [283], color images and depth data of outdoor urban scenes, e.g. houses and streets, taken from multiple viewpoints, were used to construct 3D models with surface textures. First, they extract all the planar surfaces from the depth frames. They start with small clusters, and then minimize the surfaces. If it turns out that the small surfaces are close enough to each other, they will be merged. Second, they find the intersecting lines of the surfaces, and filter the lines belonging to the non-neighboring planes out. Afterward, only the parts of the lines that are of interest, e.g. the length of the building corner, but not the entire infinite line, are chosen. Next, they find transformation matrices between the depth images, using the intersecting lines. Subsequently, the edges are found from the 2D color pictures. Finally, the user manually matches the lines from the 3D and 2D data, and the colors are mapped onto the surfaces.

In addition to outside scenes, using the color images and 3D data taken from inside scenes has been researched. For example, in [267], applications for household robots were discussed. They extract objects using partial views, since the robot cannot observe the environment with a turntable. They use the RGB information to increase the accuracy of the ICP algorithm through an extra layer of verification, along with mapping the environment [267].

In [245], revolving conic surfaces have been utilized to create a novel model for elastic articulated objects. The model is based on 3D object skeletons and deformable surfaces imitating deformations of human bodies, where each limb is adjusted using a certain 3D deformation parameter. The underlying notion enabling to do so is to obtain a sequence of 2D stereo images, and then to calculate

the required information from the corresponding points.

Nevertheless, through the course of the above process, each individual transformation usually entails a certain level of error, which when accumulated throughout a large sequence, will cause a noticeable misalignment between the to ends. The problem of correctly closing the loop for achieving global consistency in 3D reconstruction has been investigated and approached through different techniques in the existing literature. In the most cases, especially in the presence of large sequences of frames, the latter is necessary.

Incrementally tracking the motion by accumulating the drift throughout the frame transformations [53] is one of the earliest examples that has been utilized along with SfM in [302]. The offline optimization procedure proposed in [133] is another early example from the foregoing list.

One of the first algorithms accomplishing a real-time performance in creating globally consistent 3D representations of objects or scenes based on sequences of frames taken by a hand-held camera was devised on the basis of probabilistic analysis of feature position approximations [97], which was not capable of dealing with sequences larger than a certain amount. More clearly, due to the high computational cost involved, the latter method will fail to demonstrate a real-time performance if a large object or scene is going to be reconstructed, which demands creating great feature vectors and incorporating them into the calculations. Even with smaller objects or scenes, the amount of data to be handled is larger than what could be sustained along with a dense filter map, which incurs having to ignore some of the features, incurring a loss of accuracy.

Detecting the loop apart, the majority of loop-closure correction methods proposed in the literature heretofore rely on complex and time-consuming statistical and mathematical algorithms and operations, including computationally expensive optimization procedures, which often require intense manual inputs and interferences from the user as well, practically preventing a real-time functionality, even if other elements of the pipeline comply with it.

In [159], the loop-closure correction problem was handled using a pose-graph optimization algorithm based on the features extracted from the RGB data. The foregoing approach is an example of the long list of methods possibly leading to impressive global alignments, but in the case of being exposed to large databases, either delaying the whole reconstruction process or loosing the potential precision and accuracy due to implementing a global fusion algorithm, which is responsible for reducing the frames to representations that are based on dense patches.

In [135], the loop-closure was estimated based on key-frames, where each new key-frame was matched against all the previous ones. It was detected based on the entropy ratio, where small errors in transformations coincide with high entropy values.

To reduce the computational cost in frame feature matching, an efficient mechanism detecting loop-closures via landmarks was presented in [295]. They were compared across the input images. The loop-closure was then detected based on the correspondences.

In [288], a method was proposed for aligning very large sets of 3D point clouds. From an initial estimate of the sensor paths, a 3D graph was constructed, and the alignment problem was decomposed into smaller ones based on the loop-closures that exist in the graph. The data were aligned with Simultaneous Generalized ICP (S-GICP), which exploits the loop-closure property to produce highly accurate intra-loop registration results. The individual loops were then combined into a single, consistent point cloud via an inter-loop alignment step which reconnects the graph of the loops according to a least-squares-based optimization process.

The loop-closure problem is most widely explored in the context of SLAM, where the absence of appropriate detection and error correction frameworks can cause large errors. In SLAM problems, usually, a pose-graph is built, and then corrected using the loop-closure constraint. A suitable approach to doing so was proposed in [318, 176], which is based on fast incremental matrix factorization for the correction of transformation matrices. With the help of QR decomposition of the matrix, only the values that change were updated, resulting in a fast performance. The information matrix was also used for error estimation. Similarly, in [175], it was proposed to use Bayesian trees, i.e. data structures that provide a better understanding of the matrix factorization in terms of probability densities. It was shown how the fairly abstract updates to a matrix factorization could be translated to a simple editing of the Bayes tree and its conditional densities. As a result, the system was rather fast and accurate.

In [314], in order to minimize the conflict between the sequential constraints and the loop-closure constraints, gradient descend-based error minimization [149] was utilized for the constraint networks, which was run each time a loop-closure was detected, using the output of the previous run as the initial guess.

It is worth noticing that the above solutions are, nevertheless, mostly overly complicated for simpler and more controlled environments, where building a location graph would be unnecessary.

In order to construct the prerequisite building blocks of the 3D garment reconstruction platform to be expatiated in Chapter 7, this chapter introduces an efficient and reliable system resulted from an exhaustive investigation of the specifications of the ICP algorithm, and attempts to enhance it for the sake of coming up with a practically useful variant, such that the resulting point cloud be realistic enough to represent the associated object or scene as precisely as possible, which is accomplished by means of the contributions resulted from certain modifications made to it. Moreover, for the sake of devising a fast and reliable capturing pro-

cess, an innovative hardware setup is proposed, which benefits from an automatic turntable-based circuitry, and ensures a quick approach for obtaining consistent and reliable scans. Besides, the proposed method brings about the advantage that, in order to have the point cloud resulting from the reconstruction process already resemble the object with the same texture as the original object, the color data is transformed throughout the merging procedure, accompanying the 3D coordinates, and mapped to the final point cloud.

Furthermore, a loop-closure correction algorithm with a negligible computational load is proposed, which is referred to as Proportional Error Back-propagation (PEB), and aims at applications where the sequence always possesses similar first and last frames, whose examples, among others, include scanning a scene or object while stopping at a pose similar to that of the starting frame. The main virtue of the PEB is its unparalleled fastness, i.e. it usually takes a fraction of a second for it to correct the transformations throughout the whole sequence. Such an algorithm can be used for producing realistic models to be used in virtual fitting rooms or virtual reality applications [322, 103, 185, 223, 192].

The remainder of the chapter is organized as follows. The proposed method will be introduced in the next section. Afterward, the experimental results will be presented and discussed. Finally, the Chapter is concluded.

6.2 The Hardware Setup

In the case of reconstructing objects, in contrast to doing so on scenes, in order to alleviate the shortcoming arisen from the fact that the existing 3D reconstructions solutions suffer from the lack of a fast and reliable hardware setup, a novel configuration is proposed in this chapter, which will be illustrated and discussed in terms of its advantages over the available approaches throughout this section.

First and foremost, the capturing procedure needs to be fast, to which end, a configuration has been devised which takes advantage of a closed-circuit turntable-based setup, aiming at automating the whole capturing procedure, such that the computer communicates with the RGB-D sensor in a bidirectional manner, as well as is able to operate the turntable on top of which the object of interest has been mounted. The process consists of the computer making the turntable start rotating when required, and having the sensor capture the object simultaneously, where the data are transferred to the computer at the frame-rate designated, and finally, stopping the turntable when the sufficient number of frames have been captured and stored. The proposed setup has been schematically illustrated in Fig. 6.1, along with the related circuitry wiring.

It should be noted that not only does the proposed hardware solution lead to a fast and convenient framework for capturing the data, it significantly improves the

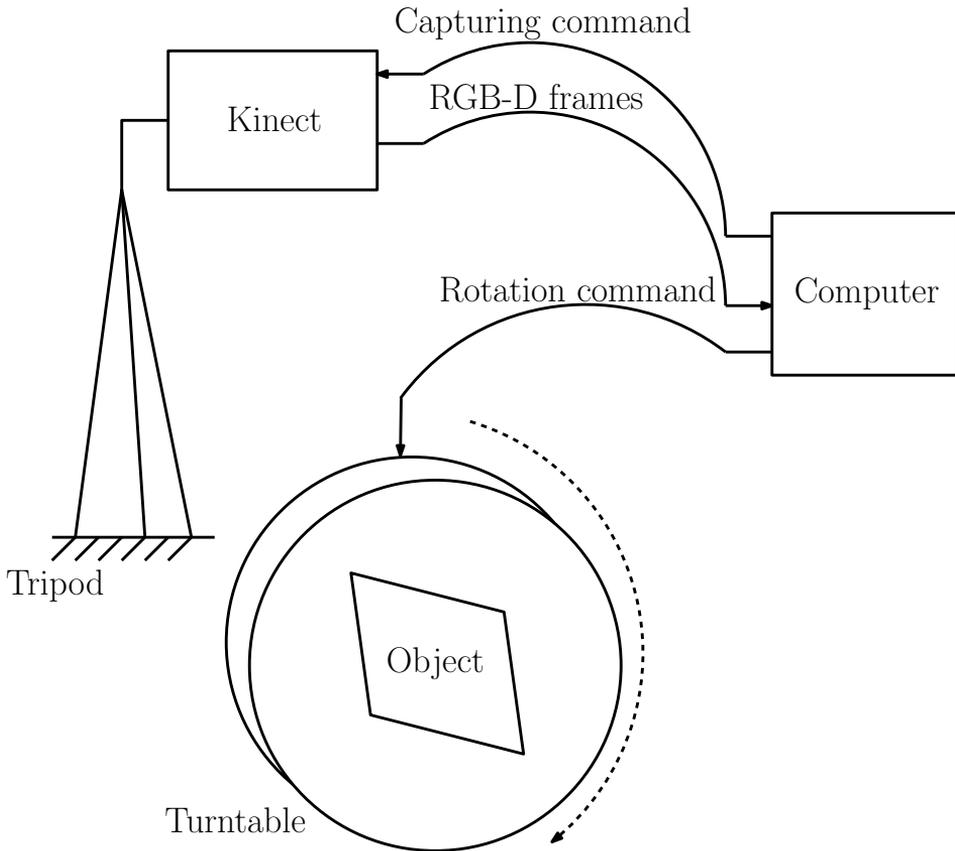


Figure 6.1: Schematic representation of the proposed automatic capturing hardware setup and the associated circuitry wiring.

quality of the individual depth maps returned by the sensor, which is due to the fact that the camera is steadily mounted using a tripod, which prevents it from shaking or undesirably moving, thereby reducing the amount of noise. Moreover, as each pair of consecutive depth maps is captured from closely similar viewpoints, whose differences are approximately the same across the whole sequence, having been brought by the automatic rotation of the object by means of the turntable, estimating the relative transformations can be performed in a considerably more robust and accurate fashion compared to the case where the camera or the object has been manually held by a human being.

The algorithm should be made take snapshots more clearly, so that the angular difference between the consecutive frames is not more than a certain amount. Otherwise, i.e. if the frames are far different from each other, while matching them by the ICP, the chances are high that the results would end up being trapped

in a local minimum. The reason, in fact, is that when the points are being matched, if the frame orientations differ considerably from each other, since the ICP tries to assign each point of a frame to the closest point in the other one, it is less likely that the process will successfully find the correct matches for them, at least as fast as it would happen if the consecutive frames were close enough to each other.

In a worse case, a local minimum would cause the matching procedure to get stuck where mathematically, the least possible sum of square distances between the points from the two frames has been achieved, but the reality is that there is no alignment of any sort. Matching the depth frames requires, first of all, removing the background, such that the object of interest is distinguished from the rest of the scene captured by the image. Therefore, we consider a consistent solid-colored background, using a white paper, which can help facilitate the task of removing the background.

6.3 The Proposed Method

As aforementioned, in this chapter, an algorithm for obtaining 3D point clouds of entire objects or scenes using frame sequences acquired by the Microsoft Kinect 2 camera, based on the real-world coordinates extracted according to the depth information, is proposed, which readily maps the original texture pattern to the reconstructed point cloud. The RGB-D frames are recorded while it is placed on a rotating turntable, by a single Kinect 2 camera. In the case of scenes, the camera is, however, held and moved throughout the required range manually, due to the fact that filming a scene using a Kinect mounted on top of a turntable would make it impossible to keep the sensor close enough to the targeted surfaces for them to always be within the range which could be captured under the existing hardware-related limitations. For preparing the frames as inputs to the proposed algorithm, they should be preprocessed, whose components will be discussed in what follows.

Since the initial point cloud from the Kinect displays the entire frame, if needed, i.e. in the case of reconstructing objects, a bounding box is manually set along the x , y and z axes, the points lying outside of which are disregarded. Besides, the depth information at the far edges of objects should be excluded, as they might not be authentic. Moreover, the obvious outliers that can be caused by lighting conditions or other unpredictable circumstances are detected, and removed, using a low-pass filter mask, which is strengthened by applying a bilateral filter being responsible for removing the remaining outliers and smoothing the surface representing the positions of the points within the point cloud. More clearly, the average distance from each point to four neighboring points is calculated, which is then compared to a threshold, namely, one standard deviation, for making the foregoing decision.

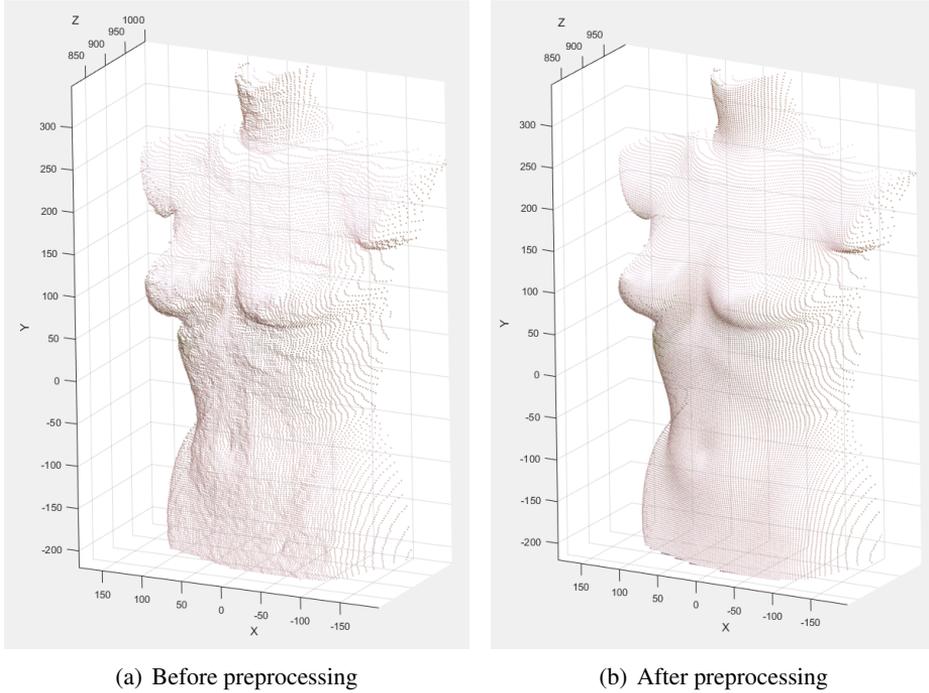


Figure 6.2: A representation of the effect of applying the proposed preprocessing framework on a sample RGB-D frame.

Fig. 6.2 represents the effect of applying the proposed preprocessing framework using a sample RGB-D frame. Afterward, in order to reduce the computational cost by deleting sets of points that are too close to each other, the point cloud is down-sampled by means of a box-grid filter, which divides the space into 1 mm^3 cubes, and keeps a single point with average coordinates of all the points included in each. For the same purpose, the grid size used to represent the Kinect depth data can be increased.

The ICP algorithm is then adopted so as to register the frames and merge them into a 3D point cloud of the whole object or scene, as briefly reviewed in what follows. The ideal rigid transformation T resulted from the ICP algorithm transforms the point cloud P_{moving} to the point cloud P_{fixed} , such that:

$$\forall \mathbf{p}_i \in P_{moving} \implies \exists \mathbf{q}_j \in P_{fixed}, \|\mathbf{T}\mathbf{p}_i - \mathbf{q}_j\| = 0. \quad (6.1)$$

If the correspondences were known from the outset, the solution would be trivial, the practical lack of which necessitates employing an iterative process, where, in the context of the proposed algorithm, in order to avoid the noisy point coordinates reported by the Kinect, rather than the point-to-point metric, the point-to-plane

one is utilized as the objective function, which is calculated as follows:

$$\sum_{i=1}^N \left\| T\mathbf{p}_i - \mathbf{q}'_j \right\|^2 \quad (6.2)$$

where N denoted the number of the points in each point cloud, and \mathbf{q}'_j is the member of S_j that minimizes $\|T\mathbf{p}_i - \mathbf{q}\|$, in which S_j is the tangent plane of P_{moving} at \mathbf{q}_j .

It is worth mentioning that due to the closeness of the poses of the consecutive point clouds used in our experiments to each other, the chances of getting stuck in locally optimal solutions is negligible. Based on the same notion, the identity transformation is considered the initial guess, which brings about the advantage of robustness against unintentional motions of the camera, but entails the drawback of degenerating when dealing with totally symmetrical objects. In order to overcome the latter problem, an approximate rotation should be provided by the user.

For a sequence containing n frames, the first frame is loaded, preprocessed and copied into the variable P_{fixed} . Then the coordinate system through which its points are represented is taken into account as reference, transforming the points from all other frames to which is, in fact, the main goal of the algorithm. Next, an absolute transformation matrix T_{accum} , which is initially set to identity, is created, being updated to perform the foregoing task on the i^{th} frame, $i = 2, 3, \dots, n$, during the corresponding iteration of the following routine: The i^{th} depth frame is loaded, and a generic Kinect-specific transformation is applied to obtain the associated point cloud, which is then preprocessed to extract the point cloud P_{moving} ; The ICP algorithm is implemented to obtain an approximated transformation matrix T that transforms P_{moving} to P_{fixed} ; T_{accum} is replaced by TT_{accum} ; T_{accum} is applied to P_{moving} , and the result is merged with $P_{current}$, while maintaining the color data obtained from the RGB frames representing both P_{moving} and $P_{current}$, followed by checking whether more than one point exists within each $1 \times 1 \times 1$ mm cube, in which case, in order to avoid unnecessarily high density, the n points present in the same cube, \mathbf{p}_i , $i = 1, \dots, n$, will be replaced by a single point, \mathbf{p} , possessing the mean 3D coordinates and RGB color values of all of them, as follows:

$$\mathbf{p} = \frac{1}{n} \sum_{i=1}^n \mathbf{p}_i, \quad (6.3)$$

$$r = \frac{1}{n} \sum_{i=1}^n r_i, \quad g = \frac{1}{n} \sum_{i=1}^n g_i, \quad b = \frac{1}{n} \sum_{i=1}^n b_i, \quad (6.4)$$

with the sets (r, g, b) and (r_i, g_i, b_i) denoting the RGB color channel values associated with \mathbf{p} and \mathbf{p}_i , respectively; P_{fixed} is then rewritten by P_{moving} . Finally, the

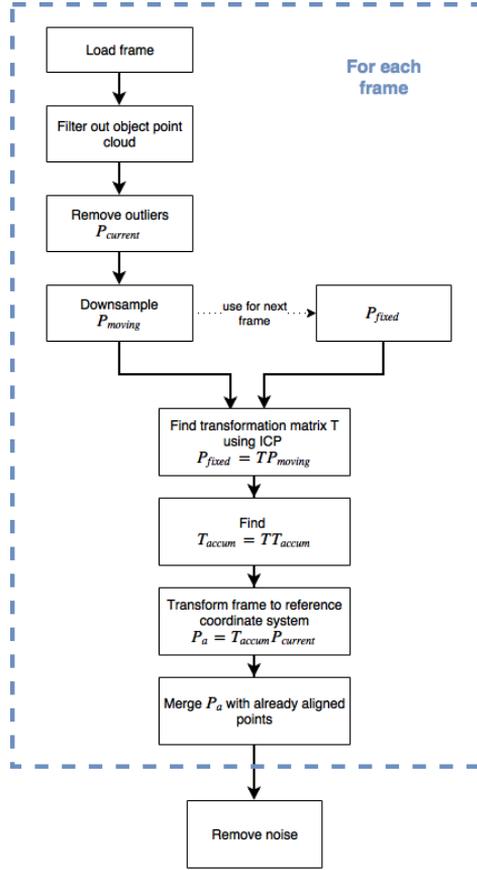


Figure 6.3: A flowchart summarizing the proposed 3D point cloud reconstruction algorithm.

resulting point cloud, i.e. $P_{current}$, is refined by removing the outliers, similarly to that of preprocessing. Fig. 6.3 presents a flowchart summarizing the proposed point cloud reconstruction method.

If retexturing the resulting point cloud using a different pattern is required, it can be performed using the procedure suggested in [47]. In the context of the latter algorithm, the pixel to take the color from for every point on the point cloud is determined based on the normalized depth value. First, the camera coordinates are mapped to the real-world ones, as follows:

$$\omega : (x, y) \rightarrow (X, Y, Z), \quad (6.5)$$

where $x, y \in \mathbb{N}$ stand for the camera coordinates, and $X, Y, Z \in \mathbb{R}$ denote the real-world ones corresponding to them. Besides, the camera coordinates are

mapped to the texture domain, using the functions f_u and f_v , as follows:

$$u = f_u(x, y) = W \frac{\omega_x(x, y) - X_{min}}{X_{max} - X_{min}}, \quad (6.6)$$

$$v = f_v(x, y) = H \frac{\omega_y(x, y) - Y_{min}}{Y_{max} - Y_{min}}, \quad (6.7)$$

where u and v denote the texture coordinates, $\omega_x(x, y)$ and $\omega_y(x, y)$ stand for the X and Y coordinates, respectively, which are determined from the map $\omega(x, y)$, and X_{min} , Y_{min} , X_{max} and Y_{max} are the smallest and largest X and Y coordinates, respectively. Moreover, W and H denote the width and height of the texture image, respectively. It is worth noticing that the texture coordinates are not dependent on the actual Z -coordinates, i.e. the map is such that the perspective effect is avoided.

The remainder of the pipeline consists in refining the transformations returned by the ICP algorithm, in order to fix the usually existing loop-closure errors, i.e. misalignments, by means of the algorithm proposed in this thesis for the foregoing aim, namely, the PEB. The input to the PEB is a sequence of depth frames, containing the 3D coordinates of the corresponding points in the associated systems, and a set of homogeneous transformations each of which supposedly maps every point in a frame to the system of coordinates through which the points from the one preceding it within the sequence are represented. More clearly, the foregoing transformations have been calculated by a registration algorithm which is expected to find the camera poses for all the frames, based on which it obtains transformations that approximately map every point in a frame to the one matching it in the previous frame.

However, the above transformations are usually not totally accurate, and the negligible error entailed by each of them still contributes to a considerable overall error, which causes an overall inconsistency, and prevents the reconstruction loop from closing properly. More clearly, all the points in each frame are supposed to be mapped to their locations in the reference coordinate system, which is tantamount to that of the first frame, by means of a homogeneous transformation resulted from accumulating a sequence of transformations each of which maps them one step backward, i.e. from the coordinate system associated with a frame to that of the one preceding it, where although no outstanding misalignment might show up at every step, the aggregated error may be considerable. The latter errors may have been caused or compounded by a variety of factors, including vibration of the camera or its movement in directions, or under axes, other than the intended ones and measurement noise.

The purpose of the PEB is to overcome the above error, and correct the loop-closure, taking the following principle into account as the criterion: If the first

and the last frames are exactly the same, the cumulative homogeneous transformation taking the latter to the former must be equal to the identity transformation. If the aforementioned condition is met, i.e. if the first and the last frames are the same, the overall error is equivalent to the existing cumulative transformation supposedly mapping the last frame to the first one. By the PEB, to correct the foregoing error, it is back-propagated throughout the chain of transformations based on their proportional contribution to the overall transformation.

In order to do so, the rotation and translation components of the transformations are modified separately and respectively, where an extra module mediating between them compensates for the effect of the modification of the rotation on the translation. It should be noted that the condition that the first and the last frames input into the PEB must be the same necessitates making a copy of the first frame and inserting it to the end of the sequence before performing registration, where the difference between the poses of the camera between the first frame and the original last one needs not to be larger than the threshold that could be tolerated by the registration algorithm when finding the transformation mapping one to the other, meaning that the scanning process should finish at a pose close enough to its starting one.

The underlying methodology of the PEB will be described in mathematical terms in what follows. Assuming that n distinct frames, being each stood for by a set F_i , $i = 1, 2, \dots, n$, exist in the whole sequence, the points in the i^{th} one, i.e. F_i , are represented through a Cartesian coordinate system \mathcal{F}_i which is defined by the origin O_i and the axes X_i , Y_i and Z_i , as follows:

$$\forall j, j \in \{1, 2, \dots, n_i\} \implies \left[\mathbf{p}_{i_j(3 \times 1)} \right]_{\mathcal{F}_i} \in F_i, \quad (6.8)$$

where \mathbf{p}_{i_j} represents the position vector of the j^{th} point in F_i , namely, P_{i_j} , and n_i is the total number of the points in F_i . It should be noted that i and j are dummy variables to be changed throughout the chapter. Then, having in mind that a copy of the first frame, $F_{n+1} = F_1$, has been added to the end of the sequence, i.e. there are now $n + 1$ frames in the sequence, upon constructing the homogeneous coordinates of P_{i_j} , namely, $\{\mathbf{p}_{i_j}\}_{\mathcal{F}_i}$, as follows:

$$\forall i \forall j, i \in \{1, 2, \dots, n + 1\} \wedge j \in \{1, 2, \dots, n_i\} \implies \left\{ \mathbf{p}_{i_j} \right\}_{\mathcal{F}_i} = \left[\left[\mathbf{p}_{i_j} \right]_{\mathcal{F}_i}^{\text{T}} \quad 1 \right]^{\text{T}}, \quad (6.9)$$

the homogeneous transformation matrix $\mathbf{T}_{i(4 \times 4)}$, which has been obtained by the registration algorithm, maps the homogeneous coordinates of every point in the $(i + 1)^{\text{th}}$ frame, $\{\mathbf{p}_{i+1_j}\}_{\mathcal{F}_{i+1}}$, from its native coordinate system, being \mathcal{F}_{i+1} , to that of a point $\left\{ \mathbf{p}_{i+1_j(3 \times 1)}^* \right\}_{\mathcal{F}_i}$ supposed to match it in the preceding one, namely,

\mathcal{F}_i , meaning that:

$$\forall i \forall j, i \in \{1, 2, \dots, n\} \wedge j \in \{1, 2, \dots, n_{i+1}\} \implies \left\{ \mathbf{p}_{i+1_j}^* \right\}_{\mathcal{F}_i} = \mathbf{T}_i \left\{ \mathbf{p}_{i+1_j} \right\}_{\mathcal{F}_{i+1}}, \quad (6.10)$$

where:

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{Q}_{i(3 \times 3)} & \mathbf{t}_{i(3 \times 1)} \\ \mathbf{0}_{(1 \times 3)} & 1 \end{bmatrix}, \quad (6.11)$$

in which \mathbf{Q}_i and \mathbf{t}_i stand for a rotation matrix and a translation vector, respectively, and $\mathbf{0}$ denotes a vector of all-zeros. If the camera poses have been calculated flawlessly, the latter transformation will map the homogeneous coordinates of every point from the corresponding coordinate system to its own representation in the coordinate system associated with the previous frame, i.e. ideally, $\left\{ \mathbf{p}_{i+1_j}^* \right\}_{\mathcal{F}_i}$ should be equivalent to $\mathbf{T}_i \left\{ \mathbf{p}_{i+1_j} \right\}_{\mathcal{F}_{i+1}} = \left\{ \mathbf{p}_{i+1_j} \right\}_{\mathcal{F}_i}$, which is usually not the case, due to the errors having, as aforementioned, arisen because of a variety of reasons. The foregoing inconsistency, in fact, explains the cause of the loop-closure error, i.e. the accumulation of the error throughout the transformations prevents the to ends of a closed loop of the frames from coinciding with each other at the pose they are supposed to do.

In order to define measures describing the overall error, which is tantamount to the loop-closure error and should be back-propagated so as to correct the loop-closure by modifying the transformations, one could find the overall transformation supposedly mapping the points from the coordinate system associated with the newly inserted last frame to that of the first frame, which are in fact the same, and upon noticing that ideally it has to become an identity homogeneous transformation, derive the loop-closure pose error from it. In other words, the accumulation of the first to the n^{th} homogeneous transformations, namely, $\mathbf{T}_{T(4 \times 4)}$, which can be found as follows:

$$\mathbf{T}_T = \prod_{i=1}^n \mathbf{T}_i, \quad (6.12)$$

can be considered as a homogeneous transformation constructed on the basis of the parameters standing for the loop-closure error, such that with the following representation:

$$\mathbf{T}_T = \begin{bmatrix} \mathbf{Q}_{T(3 \times 3)} & \mathbf{t}_{T(3 \times 1)} \\ \mathbf{0} & 1 \end{bmatrix}, \quad (6.13)$$

\mathbf{Q}_T and \mathbf{t}_T denote a cumulative rotation matrix and a cumulative translation vector, respectively, which could be utilized to extract the rotation and translation

loop-closure errors.

For extracting the loop-closure error correction terms based on the overall homogeneous transformation, i.e. \mathbf{T}_T , first, Eq. (6.12) can be expanded through substituting each individual homogeneous transformation by the expression describing it from Eq. (6.11) in order to find \mathbf{Q}_T and \mathbf{t}_T in Eq. (6.13), as follows:

$$\mathbf{Q}_T = \prod_{i=1}^n \mathbf{Q}_i, \quad \mathbf{t}_T = \sum_{i=1}^n \left(\prod_{j=0}^{i-1} \mathbf{Q}_j \right) \mathbf{t}_i, \quad (6.14)$$

where $\mathbf{Q}_{0(3 \times 3)} = \mathbf{I}_3$ is an identity matrix.

In fact, the goal of the PEB is to find modified rotation matrices $\hat{\mathbf{Q}}_{i(3 \times 3)}$ and translation vectors $\hat{\mathbf{t}}_{i(3 \times 1)}$, according to the original ones \mathbf{Q}_i and \mathbf{t}_i , respectively, $i = 1, 2, \dots, n$, based on their proportional contributions to the overall rotation matrix \mathbf{Q}_T and the overall translation vector \mathbf{t}_T , respectively, such that the resulting homogeneous transformation matrices $\hat{\mathbf{T}}_{i(4 \times 4)}$ constructed as follows:

$$\hat{\mathbf{T}}_i = \begin{bmatrix} \hat{\mathbf{Q}}_i & \hat{\mathbf{t}}_i \\ \mathbf{0} & 1 \end{bmatrix}, \quad (6.15)$$

would overall accumulate to the identity homogeneous transformation, represented by the 4×4 identity matrix \mathbf{I}_4 , meaning that:

$$\hat{\mathbf{T}}_T = \begin{bmatrix} \hat{\mathbf{Q}}_T & \hat{\mathbf{t}}_T \\ \mathbf{0} & 1 \end{bmatrix} = \prod_{i=1}^n \hat{\mathbf{T}}_i = \mathbf{I}_4, \quad (6.16)$$

where $\hat{\mathbf{Q}}_T$ and $\hat{\mathbf{t}}_T$ denote the corrected cumulative rotation matrix and the corrected cumulative translation vector, respectively.

In the context of the transformation correction procedure of the PEB, first, the rotation matrices are corrected. The purpose is, in fact, to modify each rotation matrix \mathbf{Q}_i , $i = 1, 2, \dots, n$, such that the corrected overall rotation would become an identity rotation. In order to do so, a set of rotation correction matrices $\mathbf{Q}_{e_i(3 \times 3)}$, $i = 1, 2, \dots, n$, should be calculated to be incorporated into the construction of the corresponding corrected rotation matrices. In what follows, the proposed mathematical framework for achieving the latter goal is explained, where the virtue of the fact that the inverse of every rotation matrix is equal to its own transpose has been resorted to for the sake of reducing the consequent computational cost.

In order to format the structure of the rotation correction procedure, an expression for each corrected rotation matrix, $\hat{\mathbf{Q}}_i$, $i = 1, 2, \dots, n$, in terms of the rotation and rotation correction matrices is first derived in a way that would enable the algorithm to manipulate the corrected overall rotation, which must become identity,

by adjusting the rotation correction matrices. To this end, the cumulative rotation matrices $\mathbf{Q}_{c_i(3 \times 3)}$, $i = 1, 2, \dots, n$, and their corrected counterparts $\hat{\mathbf{Q}}_{c_i(3 \times 3)}$ are defined as follows:

$$\mathbf{Q}_{c_i} = \prod_{j=1}^i \mathbf{Q}_j, \quad \hat{\mathbf{Q}}_{c_i} = \prod_{j=1}^i \hat{\mathbf{Q}}_j. \quad (6.17)$$

Then by noticing that each corrected cumulative rotation matrix should assimilate all the associated rotation correction matrices, meaning that:

$$\forall i, i \in \{0, 1, \dots, n\} \implies \hat{\mathbf{Q}}_{c_i} = \left(\prod_{j=0}^i \mathbf{Q}_{e_j} \right) \mathbf{Q}_{c_i}, \quad (6.18)$$

where $\mathbf{Q}_{e_0} = \mathbf{I}_3$, a closed-form expression can be found for the corrected rotation matrices $\hat{\mathbf{Q}}_i$, $i = 1, 2, \dots, n$, as follows:

$$\begin{aligned} \forall i, i \in \{1, 2, \dots, n\} \implies \hat{\mathbf{Q}}_{c_i} &= \hat{\mathbf{Q}}_{c_{i-1}} \hat{\mathbf{Q}}_i \implies \hat{\mathbf{Q}}_i = \hat{\mathbf{Q}}_{c_{i-1}}^T \hat{\mathbf{Q}}_{c_i} = \\ & \left(\left(\prod_{j=0}^{i-1} \mathbf{Q}_{e_j} \right) \mathbf{Q}_{c_{i-1}} \right)^T \left(\prod_{j=0}^i \mathbf{Q}_{e_j} \right) \mathbf{Q}_{c_i} = \\ \mathbf{Q}_{c_{i-1}}^T \left(\prod_{j=0}^{i-1} \mathbf{Q}_{e_j} \right)^T & \left(\prod_{j=0}^{i-1} \mathbf{Q}_{e_j} \right) \mathbf{Q}_{e_i} \mathbf{Q}_{c_i} = \mathbf{Q}_{c_{i-1}}^T \mathbf{Q}_{e_i} \mathbf{Q}_{c_i}, \end{aligned} \quad (6.19)$$

where $\mathbf{Q}_{c_0} = \hat{\mathbf{Q}}_{c_0} = \mathbf{I}_3$.

Assuming that the accumulation of the above rotation correction matrices is represented as $\mathbf{Q}_{e_T(3 \times 3)}$, which is calculated as follows:

$$\mathbf{Q}_{e_T} = \prod_{i=1}^n \mathbf{Q}_{e_i}, \quad (6.20)$$

one has:

$$\hat{\mathbf{Q}}_T = \left(\prod_{i=1}^n \mathbf{Q}_{e_i} \right) \mathbf{Q}_T = \mathbf{I}_3 \implies \prod_{i=1}^n \mathbf{Q}_{e_i} = \mathbf{Q}_T^T, \quad (6.21)$$

meaning that \mathbf{Q}_{e_T} shall become the inverse of the overall rotation matrix, i.e. \mathbf{Q}_T , meaning that the individual rotation correction matrices can be constructed such that they stand for rotations around the same axis as that of \mathbf{Q}_T , but lead to rotation angles which accumulate to the negation of that of \mathbf{Q}_T . Therefore, assuming that \mathbf{Q}_T is represented by e_T and ϕ_T as the unit vector along the rotation axis and

the rotation angle, respectively, using the same rotation axis and the following rotation angles:

$$\forall i, i \in \{1, 2, \dots, n\} \implies \phi_{e_i} = -\frac{|\phi_i|}{\sum_{i=1}^n |\phi_i|} \phi_T, \quad (6.22)$$

meaning that:

$$\sum_{i=1}^n \phi_{e_i} = -\sum_{i=1}^n \frac{|\phi_i|}{\sum_{i=1}^n |\phi_i|} \phi_T = -\phi_T, \quad (6.23)$$

the corresponding rotation correction matrices, i.e. \mathbf{Q}_{e_i} , can be obtained, where ϕ_i stands for the rotation angle associated with \mathbf{Q}_i , the ratio $\frac{|\phi_i|}{\sum_{i=1}^n |\phi_i|}$ being meant to make each rotation correction matrix proportional to the contribution of the corresponding original rotation matrix to the overall one.

However, the correction of the rotation matrices affects the translation vectors as well. More clearly, while the rotations are being fixed, further drift will be introduced into the alignments, which appears as a higher level of error in the translations. Therefore, the translations are first revised such that the effect of the changes in the rotation would be minimized. In order to do so, it is assumed that the average position of the points from a given frame must be affected in the same manner before and after revising the corresponding rotation, which can be mathematically represented as follows:

$$\begin{aligned} \forall i, i \in \{1, 2, \dots, n\} \implies \mathbf{Q}_i \frac{\sum_{j=1}^{n_{i+1}} \mathbf{p}_{i+1,j}}{n_{i+1}} + \mathbf{t}_i = \\ \hat{\mathbf{Q}}_i \frac{\sum_{j=1}^{n_{i+1}} \mathbf{p}_{i+1,j}}{n_{i+1}} + \mathbf{u}_i \implies \mathbf{u}_i = \mathbf{t}_i + \left(\mathbf{Q}_i - \hat{\mathbf{Q}}_i \right) \frac{\sum_{j=1}^{n_{i+1}} \mathbf{p}_{i+1,j}}{n_{i+1}}, \end{aligned} \quad (6.24)$$

where $\mathbf{u}_i, i = 1, \dots, n$, is the revised translation vector. Thus the overall translation based on the newly obtained rotation matrices and translation vectors could be found based on Eq. (6.14), as follows:

$$\mathbf{v}_T = \sum_{i=1}^n \left(\prod_{j=0}^{i-1} \hat{\mathbf{Q}}_j \right) \mathbf{u}_i = \hat{\mathbf{Q}}_{c_{j-1}} \mathbf{u}_i, \quad (6.25)$$

Similarly to what preceded regarding correcting the rotations, when it comes to doing so on the translations, the above vector, i.e. \mathbf{v}_T , can be considered as the

new overall translation error, since given the fact that the sequences consisting of $n + 1$ frames stands for a fully closed loop with identical first and last frames, if the transformations had been calculated perfectly, then it would need to become zero.

Thus the task of correcting the translations will consist in distributing the aforementioned overall translation error to the individual translation vectors, \mathbf{u}_i , proportionally to their contributions. The latter are represented as follows:

$$\forall i, i \in \{1, 2, \dots, n\} \implies \mathbf{v}_i = [v_{i1} \quad v_{i2} \quad v_{i3}]^T = \hat{\mathbf{Q}}_{c_{j-1}} \mathbf{u}_i. \quad (6.26)$$

Thus the translation correction vectors can be constructed as follows:

$$\begin{aligned} \forall i, i \in \{1, 2, \dots, n\} \implies \mathbf{t}_{e_i} = \\ - [|v_{i1}| \quad |v_{i2}| \quad |v_{i3}|]^T \oslash \sum_{j=1}^n \left([|v_{j1}| \quad |v_{j2}| \quad |v_{j3}|]^T \right) \odot \mathbf{v}_T, \end{aligned} \quad (6.27)$$

which are proportional to the corresponding contributions \mathbf{v}_i to the overall error, i.e. \mathbf{v}_T , and their cumulative value is its negation, being realized as follows:

$$\begin{aligned} \sum_{i=1}^n \mathbf{t}_{e_i} = \\ - \sum_{i=1}^n \left([|v_{i1}| \quad |v_{i2}| \quad |v_{i3}|]^T \oslash \sum_{j=1}^n \left([|v_{j1}| \quad |v_{j2}| \quad |v_{j3}|]^T \right) \odot \mathbf{v}_T \right) = \\ - \sum_{i=1}^n \left([|v_{i1}| \quad |v_{i2}| \quad |v_{i3}|]^T \right) \oslash \sum_{j=1}^n \left([|v_{j1}| \quad |v_{j2}| \quad |v_{j3}|]^T \right) \odot \mathbf{v}_T = \\ - \mathbf{v}_T. \end{aligned} \quad (6.28)$$

Subsequently, the revised contributions to the overall translation are obtained as follows:

$$\forall i, i \in \{1, 2, \dots, n\} \implies \hat{\mathbf{v}}_i = \mathbf{v}_i + \mathbf{t}_{e_i} = \hat{\mathbf{Q}}_{c_{j-1}} \mathbf{u}_i + \mathbf{t}_{e_i}. \quad (6.29)$$

Lastly, in order to incorporate the above conclusion into the calculation of the corrected translation vectors, Eq. (6.26) is recalled, and the relationship between the translation contribution vectors found through Eq. (6.29) and the corrected translation vectors is established as follows:

$$\forall i, i \in \{1, 2, \dots, n\} \implies \hat{\mathbf{v}}_i = \hat{\mathbf{Q}}_{c_{j-1}} \hat{\mathbf{t}}_i, \quad (6.30)$$

based on which, utilizing Eq. (6.29), the corrected translation vectors can be obtained as follows:

$$\begin{aligned} \forall i, i \in \{1, 2, \dots, n\} \implies \hat{\mathbf{t}}_i = \hat{\mathbf{Q}}_{c_{j-1}}^T \hat{\mathbf{v}}_i = \\ \hat{\mathbf{Q}}_{c_{j-1}}^T \left(\hat{\mathbf{Q}}_{c_{j-1}} \mathbf{u}_i + \mathbf{t}_{e_i} \right) = \mathbf{u}_i + \hat{\mathbf{Q}}_{c_{j-1}}^T \mathbf{t}_{e_i}. \end{aligned} \quad (6.31)$$

6.4 The Experimental Results and Discussion

In order to validate the proposed 3D reconstruction algorithm, and verify its efficiency and applicability, it is applied to a database of objects and scenes with different characteristics. Although Kinect 2 is capable of taking maximally 30 FpS, 10 is deemed enough for the purpose of this thesis.

In the case of objects, since it takes for the turntable approximately 30 seconds to make a full rotation, each sequence contains around 300 frames, meaning that the angular distance between each pair of consecutive frames is estimated to be 1.2° . At the implementation level, taking every fifth frame into account, i.e. considering the ones captured every 6° , proved to be capable of making a reasonable compromise between accuracy and speed. The processing time with 30 iterations applied to 60 frames, using an Intel i7 3.3 GHz processor, is between 30 to 90 seconds, depending on the size of the point clouds, ranging from less than 1000 to around 16000.

As an important point, the effective range where the Kinect 2 registers the depth information starts from 0.5 meters, which, given the fact that relatively small objects, having sizes of up to $20 \times 20 \times 20$ cm, have been used for creating our database, causes the object to occupy a limited portion, only around 5 percent, of the whole 200k points involved in the frame. It is also worth noticing that Kinect tends to return true depth values for surfaces facing it directly, but not the ones deviating from it by large angles, which is perhaps due to not receiving sufficient amount of reflected light back in such cases, possibly, causing distortions in the represented shape, an example of which is shown in Fig.6.4. It can be seen from the latter figure that the points at the edges of the cup have clearly incorrect depth values, which have to be filtered out.

Besides, for more complicated objects, determining the threshold specifying the filtering bounding box might also be challenging. Nevertheless, the rotation axis of the object is important for a reasonable depth filtering threshold. This is also why we suggest to position the camera perpendicularly to the rotation axis.

As shown in Fig. 6.5 for illustration, the output point cloud might suffer from considerable levels of distortion, which is alleviated by taking as many neighboring points into account as possible when calculating the average distances. However, the latter number should be determined carefully, because in the case of surfaces with low point densities, it may cause an undesirable loss of valid information.

On the other hand, as aforementioned, the PEB has been devised such that given the assumption that the sequence showing a scene or object possesses first and last frames which have been taken from similar poses, the raw transformations which have been found using an alignment method could be revised, thereby removing the apparent overall misalignment from the point cloud resulting from merging the individual depth maps.

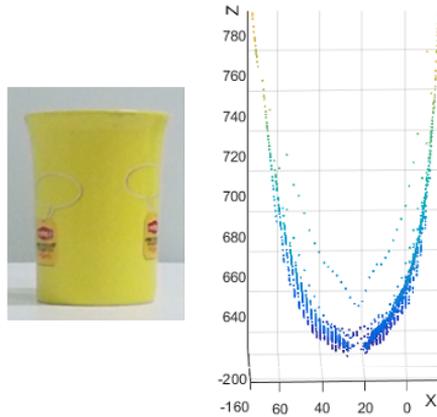


Figure 6.4: The depth values of a sample object presenting wrong scales around the edges.



Figure 6.5: Sample point clouds before and after noise removal, shown on the left and right sides, respectively.

A typical task to be left upon the PEB could be to modify the preliminary outcome of a standard 3D reconstruction pipeline consisting in filming an object, e.g. a piece of garment, while being rotated on top of a turntable, using an RGB-D sensor. In such a scenario, although the individual transformations may appear to be reasonable, the slight misalignments present in them usually accumulate, and appear as a noticeable diversion between the parts of the reconstructed point cloud corresponding to the initial and final images from the sequence. The effect of applying the PEB to a sample reconstructed model of a piece of garment has been shown in Fig. 6.6, while the manner through which it can be obtained will be discussed in more detail in Chapter 7.

The aim of the PEB is to distribute the overall misalignment to the individual transformations, so that the structure of the object would be maintained. Similarly, if a scene, e.g. a rectangular room, has been filmed instead, the PEB can be employed to modify an initial reconstructed point cloud. However, filming scenes, as opposed to objects, usually takes a higher number of frames, which is due to

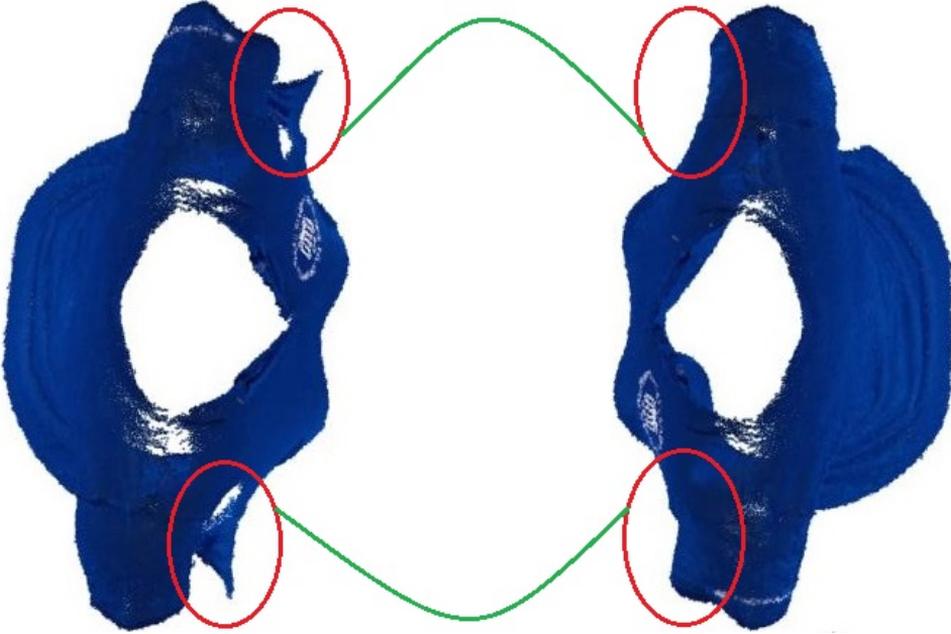


Figure 6.6: The effect of applying the PEB to the initial reconstructed model of a sample piece of garment. The images to the left and right show the model before and after applying the PEB, respectively.

the fact that every pair of consecutive frames fed into an alignment algorithm need to have been taken such that the pose of the camera in the second frame relatively to that of the first one would lead to a reasonable difference, in order for the optimization routine to converge with a tolerable level of error. More clearly, if the pairs of consecutive frames are too different from each other, then the transformation returned by the alignment algorithm may be wrong enough for the PEB to perform weakly in terms of compensating for the present misalignments.

Therefore, in order to evaluate the performance of the PEB under more extreme conditions, and examine its robustness, various trajectories have been considered for the motions of the camera while filming scenes, which include wavy patterns and movements of the camera in the opposite direction of the general trend. The latter is necessary for realizing whether the PEB could handle cases where the experimental setup requires the user to perform indisciplined movements, e.g. due to the restrictions caused by the lengths of the cables connecting the camera and the computer to each other, as well as to the electricity outlet.

The scene sequences considered for the purpose of evaluating the PEB consist of series of RGB-D frames taken while the user moves throughout the room, and holds the camera such that it is facing a part of one of the walls at all of the times-

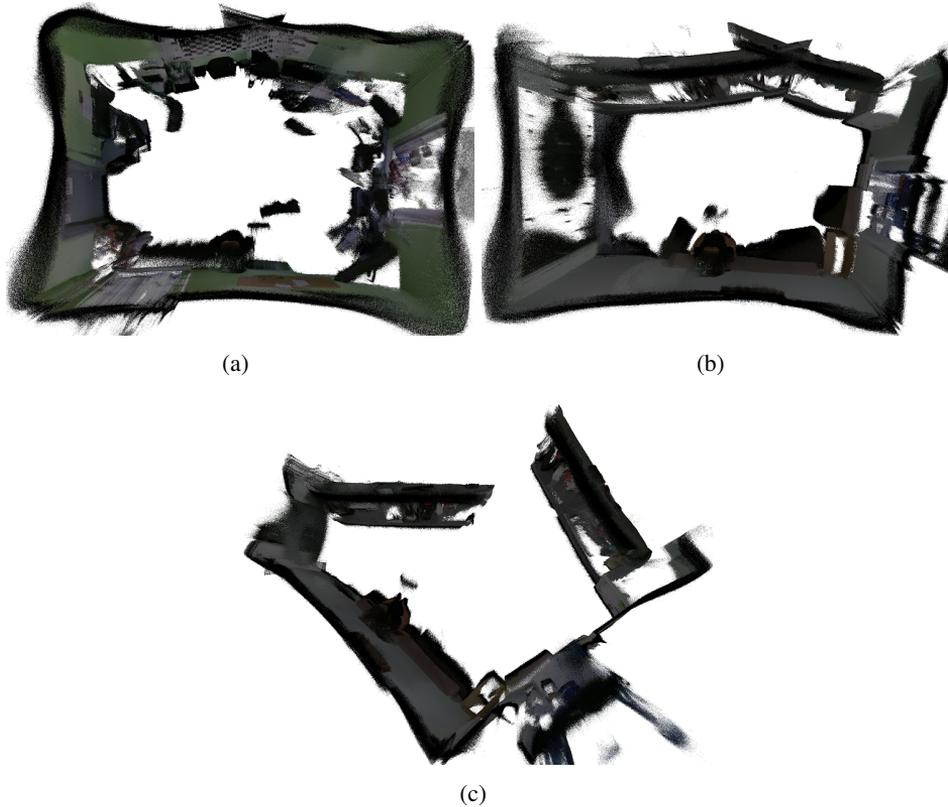


Figure 6.7: The initial scene reconstruction results.

tamps, which leads to around 500 frames for a 3×4 m rectangular room, using a frame-rate of 30 Hz. The foregoing frames have all been intentionally kept and fed into the reconstruction pipeline, i.e. downsampling has been avoided, for the sake of introducing a strong amount of misalignment, thereby verifying the robustness of the PEB.

The initial results of reconstructing sample scene sequences are shown in Fig. 6.7, whose counterparts which have been improved through applying the PEB can be found in Fig. 6.8.

As it could be seen from the results shown in the aforementioned figures, although the PEB incurs a negligible computational load, it provides a reliable platform for revising the transformations returned by a typical alignment algorithm such as ICP, which leads to smoothly distributing the overall error to the relative poses, thereby obtaining a visually appealing representation of the scene that believably corresponds to the expected 3D structure.

Moreover, for the sake of evaluating the performance of the proposed method on

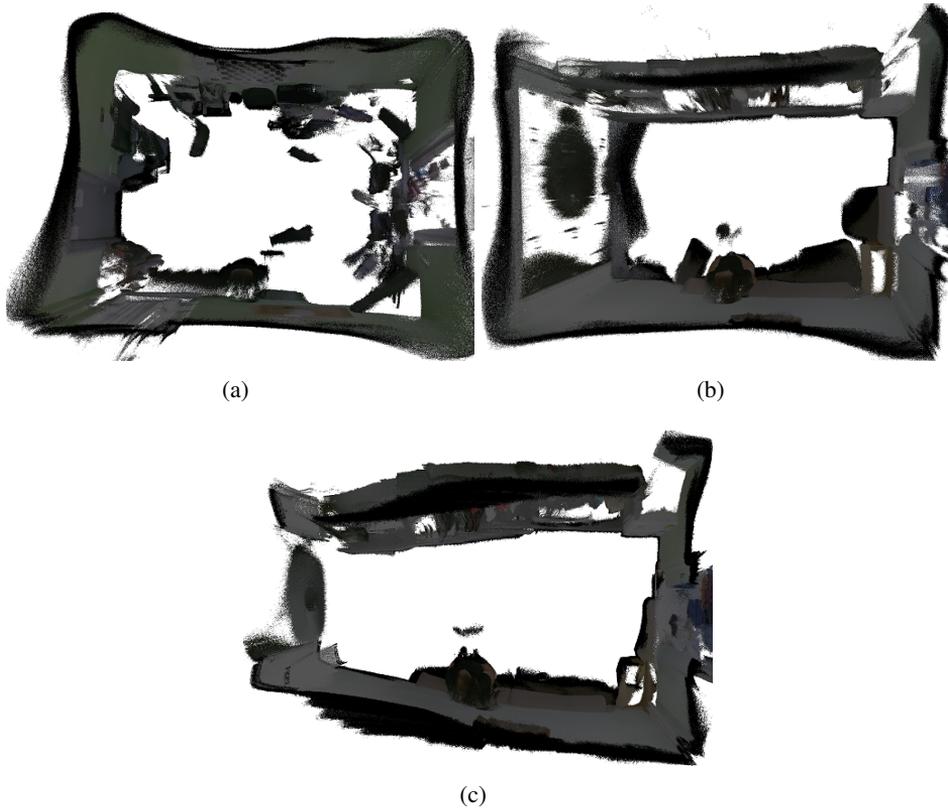


Figure 6.8: The counterpart of the reconstruction results shown in Fig. 6.7, which have been corrected using the PEB.

arbitrary objects, it has been applied to numerous sample sequences, where for illustration, RGB images and four different views of the 3D point clouds resulted from successfully applying the proposed method to some of the sequences can be seen in Fig. 6.9. Both object shape and surface characteristics affect the outcome, where the best result is achieved with objects that have smooth surfaces and a distinctive, but simple, shape, such as the human bust and the yellow teapot. Reasonable results are also obtained from symmetrical objects which have recognizable surface textures or other features that change from frame to frame, e.g. in the case of the bowl, the cup and the red hat. Objects with more complicated shapes, on the other hand, usually result in more noise in the results, overcoming which, as aforementioned, necessitates missing essential information, such as the hands of the Achaemenian soldier and the edges of the PVC object.

Fig. 6.10 shows the results of the implementation of the proposed method, while incorporating and mapping the color information, on a set of sample objects,

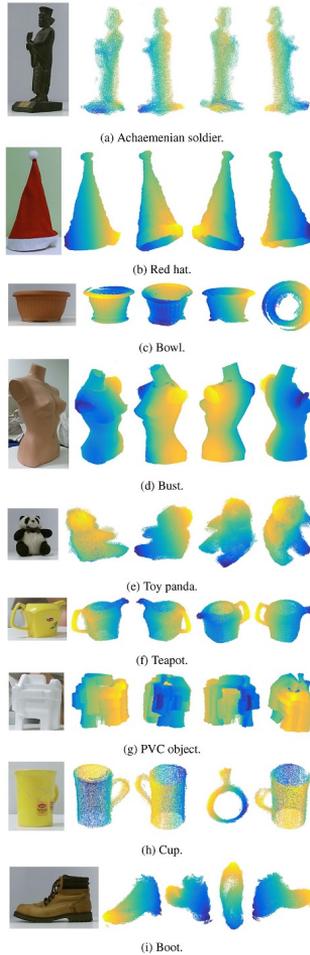


Figure 6.9: Successful results of the implementation of the proposed method.

which demonstrate its robustness and reliability in terms of successfully handling the 3D reconstruction and texturing of objects with different sizes, shapes and color-pattern types, under various experimental settings, including illumination conditions.

Despite the advantages of the proposed method, as illustrated in Fig. 6.11 it does not necessarily operate successfully under all possible experimentation scenarios. Transparent surfaces such as the drinking glass, for example, could not be handled, since they do not reflect enough light back to the sensor while being filmed. Completely black objects, e.g. the black box, moreover, absorb the light, and cause a similar problem, especially for the surfaces having sharp angles with the sensor.

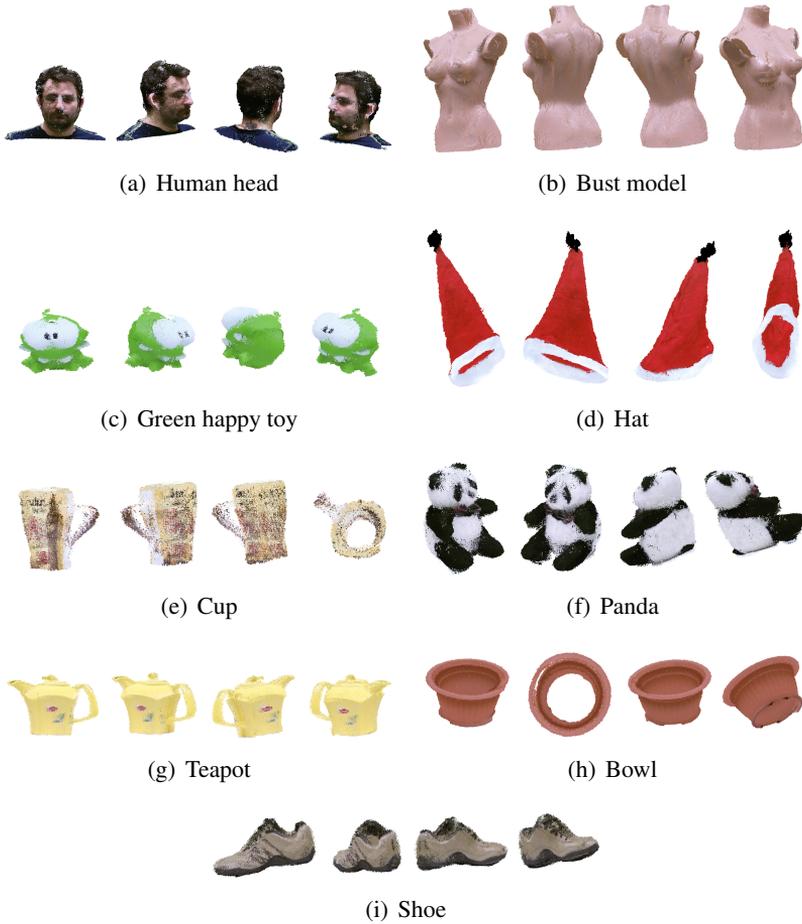


Figure 6.10: The results of applying the proposed method on sample objects, along with color mapping, each being viewed from four different orientations.

Besides, if the object has too many edges, especially if they are reflective, the noisiness would be higher. In mathematical terms, if the slope of the surface with respect to the normal plane of the camera is more than a certain threshold, which may be caused either by the pose at which the object is placed or the natural orientation of different segments of the surface with respect to each other, the noise will affect the depth frame more noticeably.

Furthermore, the method does not handle fully symmetrical objects. If the depth data is identical in a pair of consecutive frames, then the algorithm readily converges with the initial transformation guess, which, as aforementioned, is set to identity. It can be fixed by taking the color information into account or placing

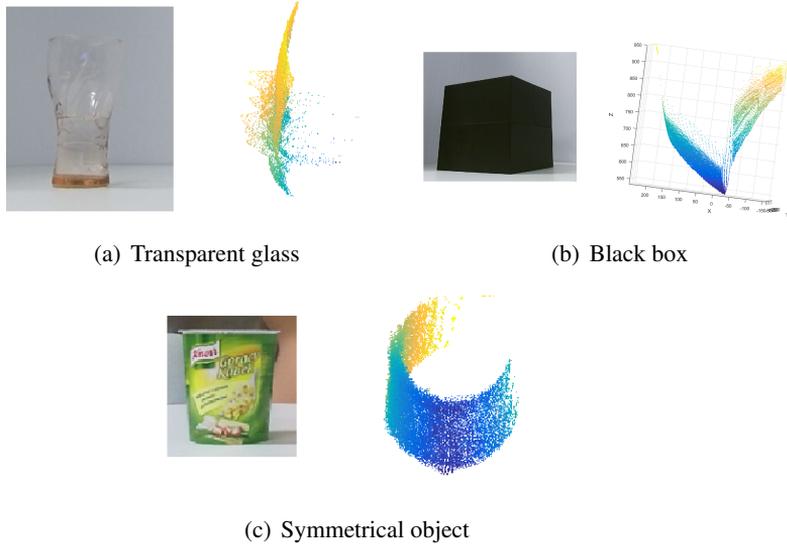


Figure 6.11: RGB images and the corresponding point clouds standing for unsuccessful experiments.

markers on the turntable, being beyond the scope of this thesis. Finally, yet importantly, as another step to follow, post-processing of the created 3D point cloud and applying surface integration [77] should be considered, which will be discussed in Chapter 7.

For the sake of presenting a measure of the computational cost associated with the proposed method, the numbers of frames and points utilized for reconstructing some of the 3D point clouds, along with the corresponding running times, as well as the number of neighbors considered for the denoising process, are listed in Table 6.1. It is worth noticing that the algorithm has been implemented at a speed of 4.3 GHz on an Intel i7-5820k processor.

6.5 Conclusion

This chapter introduced a 3D reconstruction algorithm using RGB-D sequences filmed by Kinect 2. The experimental setup was explained, along with a merging process, being, mainly, based on a certain variant of the ICP procedure. As a contributions of the proposed method, an automatic hardware setup was proposed for fast and consistent scanning using a turntable-based circuitry, and the algorithm was devised such that the configurations of the merging procedure were adjusted aiming at a relatively fast and reliable performance, as well as the fact that the re-

Table 6.1: A list of the numbers of points and frames and the running times for some of the reconstructed 3D point clouds, along with the numbers of neighbors considered for denoising. All time quantities are in seconds.

Sequence	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_o	n_f	n_p	n_{p_d}	n_n
Human head	0.9	5.6	1.3	9.5	0.0	1.3	10.6	29.2	58	285,156	265,876	400
Bust model	0.9	5.2	2.7	6.8	0.0	3.5	10.2	29.4	61	759,713	726,089	100
Green happy toy	0.8	5.1	0.5	6.1	0.0	0.4	0.1	13.1	61	43,049	39,708	10
Hat	1.0	6.8	1.6	4.7	0.0	1.7	1.1	16.8	61	348,131	335,430	10
Cup	0.8	5.2	0.0	5.8	0.0	0.4	0.4	12.6	61	54,987	50,317	40
Panda	1.0	6.8	0.7	6.1	0.0	0.7	2.7	18.0	62	136,005	132,555	200
Teapot	0.8	6.2	0.6	6.0	0.0	0.5	1.4	15.6	61	90,838	85,934	150
Bowl	0.8	5.1	0.0	7.7	0.0	0.7	1.8	16.2	61	142,921	128,394	100
Shoe	0.9	6.2	0.7	6.4	0.0	0.7	5.9	20.8	62	132,868	127,692	450

- t_1 : The running time for calculating the correction of the loop-closure error
 t_2 : The running time for the segmentation of the object of interest from the input frames
 t_3 : The running time for denoising the frames
 t_4 : The running time of the ICP algorithm
 t_5 : The running time for correcting the loop-closure error
 t_6 : The running time for merging the frames
 t_7 : The running time for the denoising the reconstructed point cloud
 t_o : The total running time
 n_f : The number of Frames
 n_p : The number of points in the reconstructed point cloud before denoising
 n_{p_d} : The number of points in the reconstructed point cloud after denoising
 n_n : The number of neighbors considered for denoising

sulting point cloud already had the color texture of the original object mapped to its surface, based on the idea that when transforming the frames, the color values were maintained along with the depth coordinates. Moreover, the noise was removed both from the RGB-D frames and the reconstructed results through applying a low-pass mask filter, where each frame was further denoised and smoothed using a bilateral filter. Furthermore, as the most essential contribution of the chapter, a fast loop-closure correction algorithm, namely, PEB, was proposed, which performs the task in a fraction of a second on a sequence of frames meant to reconstruct a 3D representation of an object or scene, where the overall transformation error is distributed to the individual relative poses proportionally to their share of the cumulative transformation. The underlying assumption was that the initial and final frames from the sequence need to have been taken at similar poses of the camera, where the camera films its surroundings or the object of interest trying to stop at a similar pose as it has started. Due to the settings suggested throughout

the chapter, the proposed method, according to the experimental results, proved to be fast and efficient. Although it might require manual inputs from the user in some cases, high-quality 3D representations were achieved for objects with comparatively simple shapes and non-reflective, non-transparent surfaces, as well as arbitrarily decorated rooms.

CHAPTER 7

3D GARMENT RECONSTRUCTION

7.1 Introduction

A fundamental stage in the course of development of a virtual fitting room would be to create 3D models of garments to be tried on the users' body. More clearly, since a garment has to be tried on different bodies, it will significantly reduce the costs, and increase the flexibility and ease of the whole online fitting utility if one could create a 3D model, and change its properties as necessary. If a 3D model of a cloth exists, it may not only be resized and reshaped, but also change in the sense of more detailed characteristics, such as color and texture, in order to adjust the virtual prospective appearance based on the specifications particular to both the customer and the garment itself.

The above will also bring about the advantage that the cloth prototype could even be used for fashion or regular design purposes, which provides the designer with the possibility of testing the cloth to be physically tailored, from any desired point of view, in order to make appropriate arrangements for changing it, if needed, before finalizing the design, minimizing the risks arisen from the unpredictable behavior of the cloth while put on different bodies of the users targeted.

Creating the 3D model of a garment, on the other hand, encourages to use a simpler device, for example, an optical camera, for the model generating purpose, which will facilitate the whole process considerably, at least, with respect to 3D laser scanners. More clearly, a laser scanner is not necessarily available to all the users, because it requires a significant amount of free space and high costs.

Nevertheless, since optical cameras possess limited FoVs, they cannot provide a 3D model based on only one shot, i.e. several 2D images of the same object will have to be taken from different orientations covering the whole surrounding, and then combined. In other words, the cloth is kept fixed while the camera rotates around it, and takes images on a regular basis, at a rate which is normally considered 30 FpS, or the camera is fixed, looking at the garment, which rotates while

being photo-shot.

In the above settings, the number of images taken is vital for creating a comprehensive and representative 3D models, i.e. taking more 2D images will result in a higher resolution, accuracy and flexibility of the resulting 3D model. After obtaining the required number of 2D images, they have to be combined for building a 3D model. As per the classical approach, first, the images have to be standardized and rendered compatible. More clearly, since the images are taken from different angles, and perhaps with different distances and positions with respect to the object, the 2D images might bear incompatible representations of the object in terms of scale, position, orientation and lightening. The latter might have arisen from different shades of color and light visible from different parts of the scene being pictured, due to the light moving in different directions and casting from different surfaces, which consist of various components of ambient, diffuse and specular light.

On the other hand, the camera parameters, such as focal length, might need to be calibrated, which is necessary for obtaining the depth information, i.e. the distance each point in the scene has from the camera. In the case of this study, as Kinect 2 is used, the operator will be exempted from a certain part of the foregoing concerns. In other words, the camera readily reports the depth data as a part of the results.

The initial steps consist in capturing Kinect frames displaying the garment from different viewpoints, where possibly more than a single camera can be used so as to obtain a more comprehensive visual description, segmenting the garment out, and following a reconstruction procedure similar to the one discussed in Chapter 6.

However, reconstructed point clouds usually suffer from noise, which is resulted from both intrinsic errors in the depth data reported by the Kinect and misalignments in registering the frames. Therefore, the remaining modules will be meant to post-process the resulting point cloud, which mostly involves denoising and then converting it to a colored polygonal mesh, constituting the main building blocks of the current chapter.

In the rest of the chapter, first, the experimental setup will be illustrated, which is followed by discussing the manner of segmenting the garment out from a raw RGB-D frame. Afterward, the ICP algorithm is utilized for the sake of merging the preprocessed frames through finding transformations mapping them to the base frame, where a cumulative transformation matrix is calculated by multiplying the basic transformations taking every frame to the one preceding it, similarly to the procedure presented in Chapter 6.

Next, various strategies will be provided for post-processing and refining the reconstructed point cloud. Finally, the surfaces represented in terms of point clouds

will be converted to triangulated meshes, which are deemed the ultimate outcomes of the proposed 3D modeling system.

7.2 The Experimental Setup

According to empirical data, given the highest angular distance between a pair of consecutive frames for a successful alignment, under typical experimental conditions, assuming a fixed relative pose between the cameras and the garment, near 20 cameras are required, as otherwise, the ICP algorithm might end up reporting completely wrong transformations. As having more than a few Kinects working at the same time does not comply with the present hardware limitations, rotating the garment on the mannequin robot, using a turntable, is necessary, which may be accompanied by using more than one horizontal layer of cameras, in order to obtain further levels of details from the upper and lower parts of the garment, which might not be visible to the central one. In the specific case of a garment worn by a mannequin robot, often the shoulder part does not provide enough number of points that would be required for a pleasant reconstruction result. Therefore, we place three Kinects at different heights, while the mannequin robot is being rotated by the turntable.

Due to the electrical structure of the process Kinect considers for obtaining raw data and calculating the depth information accordingly, it might be under the influence of the physical and mechanical properties of the experimental setup, which should be paid due attention for minimizing miscalculations of the depth information and wrong scales, as well as missing or extra points possibly being caused by various factors either related to the way the garment is being filmed or the environmental conditions under which the experiment is conducted.

Regarding the former, if possible, using a turntable for scanning the garment usually produces better results in the sense that moving the camera while filming, which inevitably involves shakiness and movement along some or all of the five DoFs other than the one sought from the outset, worsens the capability of the camera in finding out the depth information properly, resulting in a miscalculation of the depth, removal of existing points from the point cloud or emergence of non-existing points.

Having in mind that the final point cloud is the result of merging multiple depth frames, if every individual depth frame already suffers from a considerable level of noisiness, that of the result will be beyond tolerance. However, while using the turntable, the object will move only along a certain DoF, which is determined by the line perpendicular to the horizontal plane. Moreover, it should be ensured that the lighting is such that the surface whose distance from the camera is being measured is not shadowed or lightened inconsistently, as otherwise, the noise

introduced into the depth frame will be much more noticeable compared to the case where the lighting is ideal. The latter can be achieved by means of a photo-shooting cabin.

The Kinects should not be in an unstable position while filming, because practically, it may be unhealthy for the reliability of the depth data, and result in too many noisy points, as well as blinking of the whole frame. The cameras are usually placed around 80 cm away from the garment, which is ideal for the quality of the depth map. The foregoing value has been found based on trial-and-error. It should be noted that the hardware and software limitations do not allow to operate, and communicate with, more than three Kinects simultaneously.

7.3 Segmentation

After taking the raw RGB-D frames, the garment needs to be detected and segmented out, which will be described in this section. The algorithm works based on the following rules: Only the closest object to the center of the image is segmented; There must be a clear separation between the object and the surrounding environment, for example, more than 30 cm.

The inputs to the algorithm are the point cloud locations, the IR image, the erosion mask size and another parameter specifying whether the largest object should be chosen. Therefore, pieces of a certain reflective material are used under the garment, in order to facilitate the segmentation task, as shown in Fig. 7.1. It is worth noticing that adjusting the erosion parameter to erode artifacts near the reflective areas is necessary.

By default, the accepted distance between the points is 15 cm or less. If the setup allows, the foregoing value can be increased to allow bigger distances between the points. IR thresholding is considered as well, where the minimum and maximum IR values can be specified by the user, taking advantage of the fact that the parts associated with the reflective material appear in a significantly darker tone on the IR image.

In the case the camera is looking at certain parts of the garment sideways, the IR information may not be reliable for detecting the reflective material, since the surface does not have a large angle with the camera normal, and as a consequence, the IR ray is not cast back towards the camera as expected, which makes it difficult to recognize the reflective material. In such a scenario, one can use colored covers instead, as described in Section 7.4.3.



(a) Color image.

(b) IR image.

Figure 7.1: Pieces of the gray reflective material inserted under a garment for facilitating IR-based segmentation.

7.4 Point Cloud Post-processing

A point cloud resulted from applying the ICP algorithm usually suffers from various problems, including holes, jaggedness and extra points, each of which will be discussed, along with the proposed solutions, in what follows.

7.4.1 Hole Filling based on Statistical Analysis

Usually, a triangulation algorithm looks for the closest triples of the points, so that the triangles would be created while considering them as the vertices. If the points in the point cloud are distributed and positioned evenly and smoothly, and appear to be parts of a regular hyperplane, then the mesh would be such that the

triangles have normal and surface orientations that are changing throughout the whole surface smoothly and seamlessly, i.e. given that the point cloud is dense enough, rather than a combination of the triangles that are juxtaposed with each other, it would appear to be a unified and consistent surface.

However, not only is the surface of a raw point cloud resulted from performing the ICP algorithm often noisy, but also suffers from jaggedness and is froth with holes. From the point of view of 3D reconstruction, the latter problem is considered a major issue, since the step following creating the mesh is to retexture and shade the surface, which considerably depends on, and is affected by, the quality of the surface resulted from the mesh creation algorithm. In other words, if the surface of the mesh is not smooth and consistent enough, then the retexturing and shading modules would also be misled by the wrong orientations of the triangles and their inauthentic relationships with the neighboring ones. More clearly, if the triangles are not oriented properly, the normal vectors and other properties of the triangles would be accordingly incorrect, since they directly affect the calculations required for finding out the shades of colors under certain illumination and lighting conditions.

Thus it is of paramount importance to post-process the point cloud resulted from a 3D reconstruction procedure, such that it would be dense enough, and contain surfaces that, from smoothness point of view, are as close as possible to the points appearing on the surface of the physical object. In the initial stage, filling the holes should be considered, so that the surface would look as unified and consistent as possible. For doing so, first, it should be found out where the holes exist. A preliminary idea for the latter purpose would be to calculate the distance that every point has with its closest neighbor, using the geodesic distance, and based on that, decide whether the point is located in an area the density of the points in which is low enough to qualify as a hole. More clearly, if the distance between the point and its closest neighbor is smaller than a certain threshold, it could be inferred that the area around it is dense enough, and would not be considered a hole. The latter threshold could be specified through a variety of manners, including considering mean or median of the distances all the points in the point cloud have with their neighbors.

However, the above process might involve numerous complications, depending on the experimental setup and the type and shape of the object to be modeled. For example, parts of the object that should naturally be empty, i.e. do not contain any material, might be wrongly classified as holes. For alleviating the latter problem, another threshold may be specified, which would guide the iterations such that low-density areas larger than a certain extent are not considered holes anymore.

In either case, for objects with complex shapes, occlusions, self-occlusions, etc., manual interaction with, and contribution from, the user might be required. An-

other solution to finding the holes is to look for the closest point to every point separately on each of the coordinate axes, on both of the possible directions, followed by the same action as aforementioned, i.e. considering the points being too far from their closest neighbors parts of holes, which is especially helpful in the cases where the hole is relatively large and completely empty, but surrounded by dense areas all around. Otherwise, in such cases, since all the points have neighbors close enough to them, it would be sought by the variant of the algorithm taking the geodesic distance into account that there is no hole.

However, checking the sole condition of having closest points that are too far is not enough, and might be misleading. For example, when checking the points at the edges of the surfaces, their distances with the closest points on the axis pointing outward the object will perhaps be infinity, which would normally lead to the false notion that a whole exists around such a point.

Thus an extra layer of verification should be adopted that could check for the occurrence of the above scenario. The latter could be handled through introducing another threshold that checks whether the distance to the closest neighbor is less than a maximum. In other words, the condition is that if the distance with the closest point is unreasonably large, probably it means that the empty area is, in fact, supposed to be left empty.

For the sake of automating the above procedure, both of the aforementioned minimum and maximum thresholds could be determined based on a function of statistical invariants of the point cloud, such as median or standard deviation. After spotting the holes out, they should be filled in, for which numerous approaches exist. In the case the points were to appear with regular distances and positions on the independent axes, the value associated with the dependent one could be interpolated. However, since, in practice, the latter does not hold, the task involves implementing interpolation on scattered sets of points, which is more challenging. Thus a straightforward solution is to perform linear interpolation between the points at furthest edge points of the holes, and then insert a prescribed number of points, with equal distances, which could, again, be approximately the same as the median of the distances the points have to their closest neighbors.

As another solution, which is considered for the purpose of our experiments, the x and y axis coordinates can be linearly interpolated, and after determining the intermediary x and y values, they would be used to find the z value as a function of them throughout a fitting hyperplane, by means of spatial interpolation methods, such as cubic [137], nearest-neighbor [252], natural-neighbor [61] and spline [116]. The latter strategy performs relatively well in the sense of feeding enough number of the points at appropriate x and y coordinates, and filling all the holes and gaps. Nevertheless, a serious drawback is that the points fitted, in many cases, might not possess correct values that would contribute to the smoothness of

the resulting surface, i.e. they might be further away from the desired locations, either outward or inward the surface.

One approach to overcome the above problem would be to interpolate the x , y and z values separately, each time based on the linearly interpolated values of the other two, which, again, is insufficient for removing the noise, as the surface is still too jagged and uneven, which lies in the fact that the object does not necessarily have only a single piece of surface associated with every combination of the two coordinate axes being linearly interpolated, potentially leaving the interpolation algorithm no alternative other than averaging all the instances of the associated combinations appearing on different surfaces, not being reasonable, since the resulting interpolated values might be completely irrelevant.

Besides, it might be that the points having been linearly calculated along the two axes are being extrapolated on the third axis, which might lead the interpolated point to lie outside the real object. To resolve the foregoing problem, instead of refining the point cloud resulted from the reconstruction process, one could do it on every single depth frame, in the preprocessing stage, where the z value is again considered a function of the x and y values, whence the interpolated z value would be less prone to miscalculations and wrong assumptions caused by the multi-layered nature of typical objects while having been projected on a surface lying on a plane spanned by the associated two axes. The latter idea also has the advantage that the independent point coordinates, i.e. x and y values, are distributed completely evenly and regularly, which, as aforementioned, renders the interpolation task much more straightforward. The reason lies in the fact that the screen coordinates associated with every point is readily available based on the pixel information, where each point in the real-world space is associated with a unique counterpart in the screen coordinates.

Moreover, since it is ideally expected to have a point representing the surface of the object at every pixel belonging to it, looking for the closest point to every point, and subsequently deciding whether it belongs to a hole would be deterministic. More clearly, the approach to finding the holes would be to check at which places on the object frame in the screen coordinates, for a value falling within a prescribed interval, on either of the x or y axes, not enough number of points exist in a row. Furthermore, checking the foregoing condition, in contrast to the case where the interpolation is performed on the reconstructed point cloud, does not have to be performed on all the six possible combinations of the axes and directions, i.e. it suffices to check the condition by going through each of the x and y axes only once, in an arbitrary direction. More clearly, the third axis, i.e. z , is no longer taken into account while calculating the distances between the points, meaning that a Euclidean distance is considered on the XY plane.

The above notion is based on the fact that the points are not supposed to appear

at arbitrary places, i.e. are not scattered, whence the verification has to be performed only on a limited number of points associated with the pixels in the screen coordinate space. The foregoing scenario is in contrast to the case of applying the interpolation on the reconstructed point cloud, where not only do the points not have a known or manageable number, but also are distributed without any consistency or regularity throughout the space, which causes a considerably higher computational complexity and time-consumption while looking for the closest points to every point on the point cloud.

Applying denoising is of essential importance to smoothen the surface of the reconstructed and refined 3D object point cloud, since in the most cases, despite all the efforts intended to interpolate within the holes as accurately as possible, there will be points not appearing at the desired locations. As a matter of fact, the reconstruction and interpolation procedures might even worsen the point cloud in the sense of noisiness, since they may amplify the errors and coordinate misrepresentations the frames originally suffer from. Thus the denoising process might need to be implemented more than once, e.g. before and after merging the frames, etc., so as to avoid further amplifications in the upcoming stages.

7.4.2 Point Cloud Smoothing and Surface Approximation

Smoothing the point cloud is one of the paramount steps in post-processing, which will be discussed in what follows, subsequently to a brief review of the relevant literature.

In [330], a shape denoising method has been proposed, which deals with surfaces constructed based on triangle meshes and soups, along with an algorithm for comparing mesh denoising schemes. Their method is developed on the basis of non-local image denoising and bilateral mesh smoothing approaches. New positions of the noisy points are calculated using a weighted average of a set of points determined according to their similarity to the neighborhood of the original point.

In [234], a point cloud denoising technique has been proposed, which preserves the sharp features, and overcomes the shrinkage caused by the linear local fitting within the PCA. For this purpose, a weighted variant of the PCA has been suggested, which, for the sake of alleviating the computational cost and making the algorithm robust against the outliers, instead of weight factors inversely proportional to the Euclidean distance to the mean, makes use of an inversely proportional repartition of sum of the distances to the mean.

[165] concentrates on an anisotropic approach for denoising and smoothing point clouds. The image denoising technique mean shift filtering is extended to 3D surfaces. In the context of their method, vertex normals and curvatures constitute the range component, where the vertex position is considered as the spatial compo-

ment. A kd-tree data structure is used to determine the local mode of the vertices based on the local neighbors. Segmentation is performed on the basis of the similarities between the local modes. For reducing the noise while maintaining the geometric structure of the point cloud, the positions and the directions of the normals are adjusted through trilateral filtering.

In [65], an algorithm for processing TLS data has been proposed, which handles elevated noise and outliers.

[66] proposes the usage of discrete regularization for image and point cloud filtering, which minimizes a weighted average of a regularization term based on the p -Laplace operator and an approximation term. The latter results in a nonlinear filter with the smoothness level p , as well as the graph weight functions as other properties. The foregoing filter would be a graph-based version of the types such as bilateral filter, TV digital filter or nonlocal mean filter.

[279] proposes an adaptive simplification method for reducing the point cloud density. An automatic recursive subdivision procedure is introduced for choosing representative sets of points and leaving the redundant ones out, which clusters the points using the k-means algorithm [155], and takes the maximum normal vector deviation into account as the criterion for partitioning the clusters into smaller sub-clusters, where the boundaries are preserved through a separate process preceding the latter stage. Besides, a refinement algorithm is utilized to prevent the points from becoming unbalanced. The overall approach is aimed at coming up with a point cloud that possesses higher density in the curvature regions than the flat areas. It should be noted that by increasing the strength of smoothing, parts of the garment may be removed undesirably, and by increasing the minimum cluster distance, more than necessary connections will be created throughout the point cloud.

In [134], Moving Least-Squares is utilized as a robust statistical approach to create the neighborhoods required for generating piecewise smooth surfaces from point clouds hypothetically suffering from noise, and possessing sharp edges. The forward search paradigm is employed for detecting and removing the outliers. The notion of projecting the points on locally, rather than globally, smooth surfaces has been reported to be beneficial in terms of improving the stability of the foregoing function.

In [168], point clouds involving noise and outliers, as well as interfering surface sheets, are consolidated by means of robust normal estimation. A weighted locally optimal projection function is used to refine the point cloud by removing the noise and outliers and evenly distributing the points, which helps the PCA algorithm by providing it with better initial estimates of the surface normals, which is handled through a priority-driven iterative process. The priorities are set based on front-stopping at thin features, along with normal flipping for reliably handling

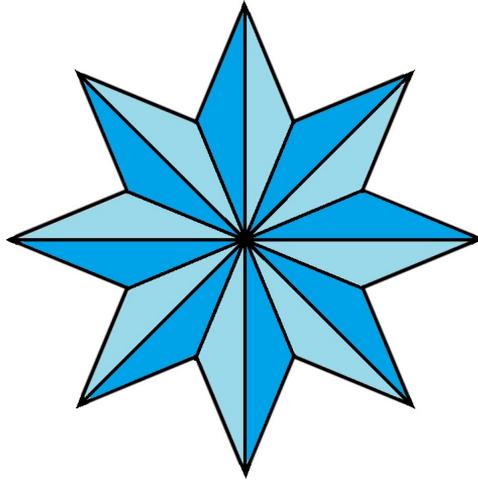


Figure 7.2: The octagram from which opening kernels are created by resampling, in order to smoothen Kinect RGB images. The image has been taken from [31].

the surface interferences.

In [338], point cloud denoising and parametrization are approached via minimizing an energy function iteratively. The 2D parametrized surface is triangulated to create the connectivity, which has been reported to be able to maintain the local distances from the original point cloud.

Normal estimation for unorganized point clouds has been investigated in [202] as well. Robust statistics is used for estimating the best tangent plane at every point. One of the main advantages of their approach is that it preserves the sharp features without demanding a manual input from the user. A noise scale estimator and an objective function based on Kernel Density Estimation (KDE) are the fundamental elements of their algorithm. The system is reportedly robust against sharp features and outliers.

In the context of this thesis, in order to avoid the noise preliminarily as possible, we use the highest-precision source of color data provided by Kinect, which is the Full HD (FHD) RGB image. It should be noted that the images originally returned by the Kinect naturally contain irrelevant information which will affect the results of automatic segmentation of the garment from the color image. Trying to eliminate those by opening the images using typical kernels, one might miss important curvature or corner information. Therefore, a new mask is created by sampling from an octagram, which is shown in Fig. 7.2, and has been taken from [31], in order to ensure that corners of different orientations would be preserved.

It should be noted that kernels of different size will result in different outcomes. Usually, using larger kernels leads to more jaggedness, but on the other hand,

more overall smoothness. Consequently, finding a compromise depends on the configurations of the experimental scenario, i.e. it should result in both an acceptable overall smoothness and corners' sharpness.

Next, the edges have to be further smoothed. There are numerous approaches proposed in the existing literature that could assist with the foregoing purpose. However, most of them cause missing the sharp corner information. To the end of alleviating the latter issue, one could think of first detecting the corners and using a gentler level of smoothing in the areas encompassing them, which is realized through detecting the active contours subsequent to segmentation of the garment part from the color image. More precisely, we find the outer contours of the garment from the depth frame, and then leave out the outermost parts to achieve the final model. Examples of the results applying the proposed preprocessing algorithm to the initial garment masks are shown in Fig. 7.3.

The noisy points appearing in the reconstructed point clouds are mainly because of the fact that in every frame, the points close to the edges, especially, outer boundaries, of the objects within its FoV usually either are irrelevant, or suffer from wrong scale in positioning, miscalculated color, etc., which although do not count to a high number compared to the whole frame, but when aggregated and all transformed to the reference frame, cause obviously visible masses of noisy points, which are difficult to distinguish and remove.

For creating a relatively smooth garment surface from the processed point cloud, we project the data points to integer values. The same process is repeated in order to eliminate distances that are less than 1 mm. Then we slice the shape horizontally. After obtaining a structured point cloud, each slice is treated separately. The aim is to convert a 2D point cloud to a line, which is achieved by creating a binary mask, and to extract the skeleton from it. Next, we create the shape, where the garment becomes a surface which is represented by an organized point cloud, as shown in Fig. 7.4.

Regarding the task of detecting and filling the missing areas of the reconstructed garments, the underlying notion is based on fitting a piecewise-continuous Spline onto all the horizontal cross-sections of the garment, and inserting new points at the missing ones. More clearly, the algorithm will detect both of the sleeves and the trunk as separate segments of the cross-section curve, and fit a Spline on each of them. Nevertheless, as the definition of such a function requires the curve to return a unique value for each individual input, we first divide the cross-section into two parts abiding by the latter requirements, being followed by optimizing the parameters of the Spline models.

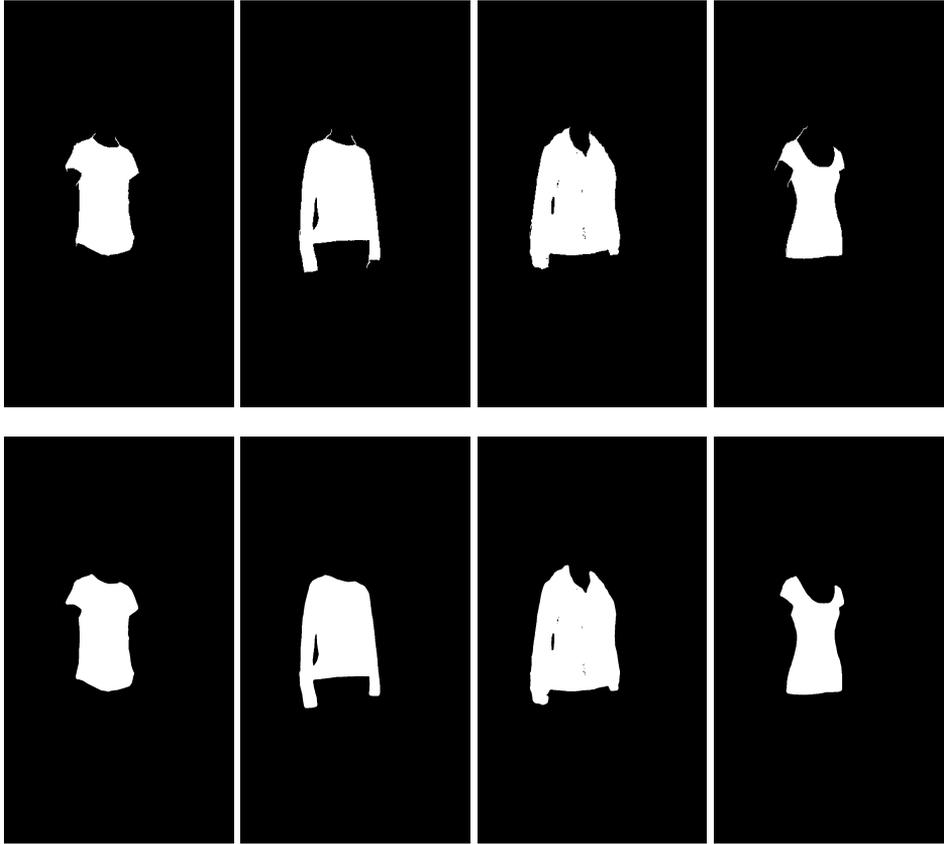
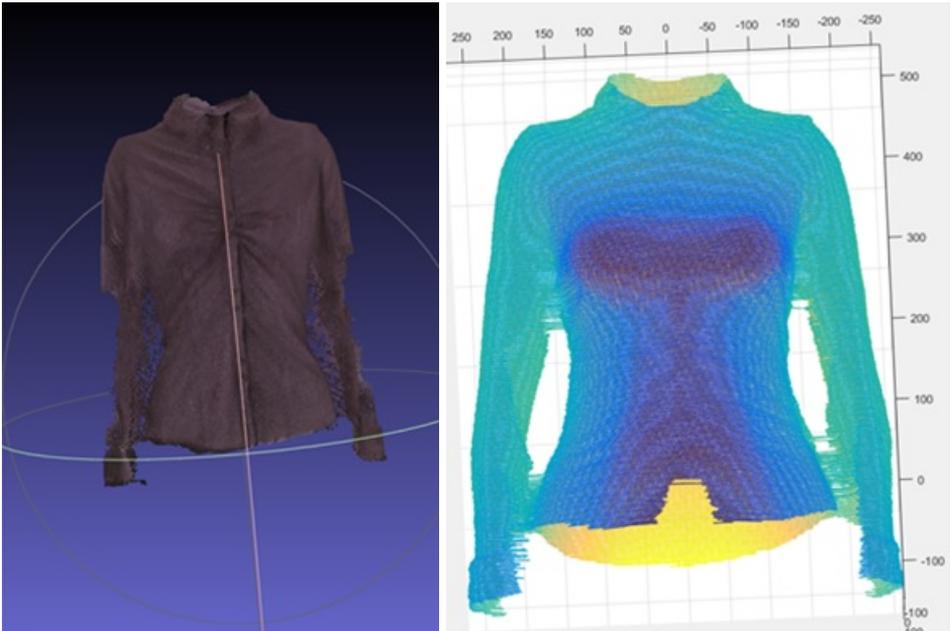


Figure 7.3: Preprocessing the initial garment masks. In every column, the images to the top and bottom show the garment mask before and after preprocessing, respectively.

7.4.3 Removing Extra Points

Obviously, a higher number of points usually leads to preserving more details. However, the latter quantity depends on many factors, and can be adjusted according to the quality requirements and computational limitations, amongst which a reasonable compromise needs to be found. First of all, while merging the depth maps, we keep a single point from every cube in the space. Therefore, the number of points present in the initial reconstructed point cloud depends on the size of the cube. Second, before creating the mesh, the unstructured point cloud is converted to a structured one, which is accomplished by mapping every point to its closest representative point from a mesh grid. Consequently, the number of points depends on the precision of the grid as well. The points kept in the foregoing step will constitute the vertices, which will affect the number of faces as well.



(a) The input point cloud.

(b) The resulting surface.

Figure 7.4: Converting a garment point cloud to a surface.

An inspection of the noisy points reveals that the colors representing them tend to fall within the ranges associated with either bright or dark gray colors. By virtue of the latter concept, we utilize a denoising module that, on top of the typical procedure described in Chapter 6, removes the noisy points based on their colors. An advantage of the foregoing strategy is that it will readily remove the points corresponding to the cover, rather than the cloth itself, and leads to a point cloud containing the garment itself only.

We use a condition which checks for the value associated with at least one of the color channels to be higher than a certain threshold, and another condition, for at least one to be lower than another threshold. There are also other conditions that check for the color not to be too similar to the typical colors of the noisy points, and for the extreme values on the x axis. One of the limitations is that if the main color of the garment falls within the aforementioned interval, it will not lead to an acceptable result.

The K-means clustering algorithm is utilized in order to find out the prominent color clusters. However, by removing minor color clusters, the details required for representing the garment with a sufficient level of shading will be lost, which lowers the visual quality.

If the points showing up in a cluster, which is supposed to resemble them with the highest level of concentration compared to others, do not have enough density, then it means that they do not appear where they should, i.e. on the part of the garment they belong to, which might have been caused by a functional Kinect error, such as the ones taking place at the edges or on reflective surfaces, or they are so trivial and unimportant that removing them would not have a significant effect on the quality and informativeness of the point cloud data.

With the above assumption, if we classify the colors, and find their different clusters, then the parts of the clusters that do not appear with enough density can be safely removed, without losing important information. The foregoing goal can be achieved through applying spatial denoising to each of the point clouds obtained from classifying the original one based on color.

After detecting different color clusters using K-means, only the ones that do not typically represent undesired colors are kept. In order to detect these colors, the centroids of the color clusters that usually represent undesired colors are picked up, and added to a color bank. Next, after clustering the new samples, the Euclidean distance between the centroid of each cluster to each of the colors present in the aforementioned bank is found, followed by removing the clusters that lead to small enough distances, which means that they either stand for noise or an undesired part of the scene.

Practically, it is concluded that using 13 color clusters is typically sufficient for garment modeling purposes. However, if the garment is such that it contains colors that are close to the noisy points or the mannequin cover by nature, a higher number of clusters should be considered to discriminate between them, and prevent the unwanted removal of points from the ones that actually belong to the garment, and should ideally be kept as they are.

Besides, as the K-means algorithm picks up the cluster centroids for the first iteration randomly, sometimes, the colors might be clustered such that the noisy and correct data are mixed. In the cases that the majority is noisy or irrelevant, some of the correct data will be missed, and in the opposite case, some noise will be introduced into the resulting point cloud.

In order to overcome the above problems, a certain degree of tolerance is allowed, within which the cluster is still removed if it is not significantly far from the members of the color bank. This way, as aforementioned, some of the correct data might be lost, but the noise is removed with higher probability. The threshold required for the latter purpose is specified manually, which has been determined through trial-and-error.

Another virtue of denoising using the above strategy is that if some parts of the garment has occluded with the rest, it is easily detected and overcome. For example, if the noise due to a part of the garment is appearing on another one which has

a distinctively different color, after clustering, it will appear with a much smaller density, and easily disappear with normal denoising. Therefore, all the clusters that have been selected to be included in the final 3D model are first denoised separately, and then merged with the rest.

The algorithm requires configuration of two parameters, namely, the threshold if the centroid of the color cluster is further than which from any of the colors stored in the color bank it will get removed and the number of neighbors for the closest-neighbor-based denoising. Adjusting the foregoing parameters for every point cloud separately leads to better results, as there may be much differences between the closeness of the garment color types to the ones stored in the color bank. For example, if the garment is gray, which is relatively close to the colors stored in the color bank, then the threshold for removing the color clusters should be smaller, compared to the case where it has a distinctive one such as blue or red, in order that the resulting algorithm specifications would avoid removing so many color clusters.

Besides, the number of neighbors for the closest-neighbor-based denoising should be adjusted based on the density of the point cloud. More clearly, if the point cloud resulted from removing the irrelevant color clusters is dense enough, then it can be heavily denoised in order to obtain a fine shape which does not contain any outlier. In contrast, if many color clusters had been removed, and the point cloud is already thin, using a large number of neighbors might cause missing important information, and lead to a point cloud with see-through parts resulted from removing an unreasonable number of points from thinner regions of the garment. Fig. 7.5 represents the whole process of obtaining a clear point cloud from a raw RGB frame on a sample piece of garment.

7.5 Converting the Processed Point Cloud to a Mesh

Once the processed garment point cloud is obtained, it should be converted to a 3D model. First, the point cloud undergoes downsampling so as to reduce the point density, and only keep the points that are necessary for creating the subsequent surface to be fit onto them. In other words, if there are unnecessarily too many triangles, it will increase the computational cost.

As aforementioned, the downsampling strategy is to average the positions of the points in every cube of a certain size, and keep the resulting point as the representative of all. It should be noted that the size of the foregoing box should be adjusted based on the level of smoothness and details of the surface sought from the outset. More clearly, a larger bounding box will lead to a fewer number of points, and consequently, less details and less smoothness of the surface.

The next step is to triangulate the points for creating the surface. However, as



(a) A sample frame as a whole.



(b) The mannequin robot only.



(c) The segmented out garment.



(d) The denoised garment.

Figure 7.5: Obtaining a clean garment point cloud from a raw RGB-D frame.

each point is usually displaced during smoothing, the resulting point cloud might contain more than one point in some of the voxels. Therefore, it needs to be down-sampled once more, which prepares it to be fed into the triangulation algorithm as an input. Sample colored, triangulated meshes from the reconstructed garment models are shown in Figs.7.6 to 7.11.

The above stage is necessary since one of the parameters of triangulation checks for the closest point to every point, and does not allow it to be included in a triangle whose edges are considerably large, with respect to the latter quantity, meaning that if the points are too close to each other, they will not be able to get connected to the neighboring points which should otherwise have happened. The triangulation algorithm involves a set of parameters, including the aforemen-



Figure 7.6: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

tioned maximum relative distance between the neighboring points, the minimum and maximum of the angles in each triangle, the number of neighboring points to be checked, the maximum angle between the neighboring triangles and their maximum edge size, all of which should be carefully configured for the purpose of obtaining fine surfaces, such that in the most cases, they would lead to a reasonable performance and an acceptable level of quality of the results, which might not be possible in many cases, because of the considerable differences between the experimental scenarios.

As the triangulation algorithm requires a structured point cloud, we project it onto a mesh grid, which is created based on the minimum and maximum values of the location coordinates of the point cloud, with an adjustable step size, the smallest



Figure 7.7: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

possible amount of space for each instance of it a point could be found in the point cloud would be a cube with the aforementioned value as its length. The latter purpose is achieved through finding the closest point in the mesh grid, by rounding the location coordinate values at the closest possible one from the mesh grid.

Next, the point cloud undergoes triangulation and color interpolation and mapping for rendering. The points included in the garment model are discriminated from the rest of the points in the Cartesian, axis-aligned grid of the point cloud with a flag which is turned on for the points to which a point from the original point cloud is projected, where also the R, G and B color channel values are taken from the same indices in the original point cloud. More clearly, every voxel belonging to the foreground is labeled with 1, and the rest with -1. We sample the color from the unorganized point cloud by finding the closest point within it using the



Figure 7.8: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

Quickhull algorithm for convex hulls [52], which is accelerated by feeding in a triangulation of the coarse point cloud being obtained through Delaunay triangulation [198].

For removing the voxelization effect, the point cloud is passed through a low-pass filter. The isosurface is extracted using the Marching Cubes algorithm [216], where the triangulation representation is achieved through obtaining the simplices from the input. The surface normals are calculated based on the vertices of the isosurface volume. For painting the model, the edge colors are ignored, but the face color is found out through bilinear interpolation of the vertex colors.



Figure 7.9: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

7.6 Approximating Nonexistent Garment Size Categories

In order to avoid the burden of generating each of a high number of shapes individually, it may be feasible to create approximate representations of the garment size categories not present in the scanning database instead, where one of the challenges is to estimate the changes of the detailed shape of the garment in a certain size category based on another one while properly approximating the resulting effects on every major part of the garment. For example, when comparing the appearance of the garment between extremely small and extremely large size categories, the pattern of the variations of the looseness of the shape throughout the garment usually changes considerably, where in the latter, either strongest wrinkles or more noticeable looseness appear in lower parts than those of the lat-



Figure 7.10: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

ter. Therefore, the task of creating a realistic appearance cannot be accomplished through simply scaling all the parts. More clearly, not only should the scale factors be particularly tailored to every major part of the garment in every size category, but also the changes of the shape should be estimated using a more comprehensive nonlinear map that takes the aforementioned considerations into account properly. As a result, it may not be possible to approximate the shape in a missing size category solely based on a database of the sizes belonging to a single category, and thus one may need to consider using at least one smaller and one larger category of the sizes in an independent interpolation procedure so as to achieve the foregoing goal.



Figure 7.11: A sample colored, triangulated mesh from a reconstructed garment model being visualized from various viewpoints.

7.7 Conclusion

This chapter proposed a 3D garment reconstruction process to be used for virtual fitting purposes, which is based on merging multiple depth frames taken from a piece of garment in a certain pose from different viewpoints. As the first step, the raw Kinect frames were preprocessed and denoised, in order to segment the garment out, and obtain a clean depth map. In the post-processing stage, a statistical hole-filling algorithm was proposed, along with strategies meant for removing the extra points and smoothing the resulting point cloud. Finally, the refined point clouds were converted to colored, triangulated meshes.

CONCLUSION AND FUTURE WORK

In this chapter the materials discussed throughout the thesis will be reviewed, along with suggestions as to how the results can be further enhanced for 3D virtual fitting purposes, or possibly extended for other applications.

Conclusion

This thesis investigated the problem of enhancing the commercial efficiency and reliability of virtual fitting services offered by online garment retailers through making use of automated procedures based on mannequin robots and 3D models. The fundamental subjects discussed throughout the thesis included an automatic virtual fitting process utilizing mannequin robots, as well as a 3D modeling algorithm aiming at providing realistic representations of garments with a reasonable computational load.

In fact, the solution devised based mannequin robots is a fit-advising system intending to produce an approximate visualization of a piece of garment worn by a human by means of extracting the measurements using a laser scanner, and then finding an optimum set of control commands leading to a shape of the mannequin robot most realistically representing the body, according to the correspondences taken from a manually developed size dictionary.

On the other hand, the 3D modeling procedure proposed within the framework of the thesis is meant to create a detailed resemblance of the garment while taking a certain pose upon filming it from various viewpoints using the Kinect sensor, obtaining a raw depth map from each, and subsequently stitching them for acquiring the geometric properties to be further processed and enhanced in the course of filling the holes, smoothing the surface and converting it to a colored, triangulated mesh.

The model obtained through the foregoing procedure would then need to be considered as an element of a dictionary to be utilized as a basis for generating those supposedly corresponding to other body shapes and poses not present in the database, similarly to the algorithm applied to the outcomes of employing mannequin robots, which was partly studied throughout the thesis.

Future Work

Although the 3D object, scene and garment reconstruction procedures developed in the course of preparing the materials for this thesis were devised having in mind the underlying utility, i.e. 3D virtual fitting, the proposed framework is flexible enough to handle other similar tasks, e.g. in virtual or augmented reality, or for applications related to mobile devices.

More clearly, as virtual and augmented reality applications aim at replacing objects and scenes by their 3D models in order to simulate their real-world behavior, thereby reducing the underlying costs and the burden of manual work, the 3D reconstruction procedure proposed in this thesis could help do the foregoing task relatively conveniently, using a single Kinect sensor, as an efficient and affordable means.

On the other hand, mobile devices such as robots may require obtaining 3D models or measurements of the surrounding environments, when it comes to tasks including navigation or mapping. Thus by extending the procedures proposed throughout the thesis, similar applications could be developed for such devices, possibly making use of other types of sensors or cameras, which would comply with the hardware requirements and limitations as well.

As other possible directions for future work, after creating the model of the cloth, it needs to be put on the model of the body having been scanned, which will demand further processing of the material characteristics, intending to figure out how the final prospective result could be realistically determined and rendered. More precisely, on top of all the properties specific to either the cloth or the body model, there should be separate calculations aiming at finding out the behavior of the cloth being put on each particular body, which demands determining the forces, such as gravity.

The above, along with the characteristics of the material from which the cloth is made, plays a role in the parts of the cloth taking certain forms, e.g. hanging loosely from the associated parts of the body. Besides, the latter issues should be investigated in a more comprehensive manner when it comes to animating the cloth, which will have to include the effect of the dynamic properties and forces, presenting a duality with the static ones as well.

BIBLIOGRAPHY

(Ordered alphabetically)

- [1] 3d body scanner - vitronic - the machine vision people. <https://www.vitronic.com/industrial-and-logistics-automation/sectors/3d-body-scanner.html>, accessed: 2017-12-11
- [2] 3d laser scanning system with real-time registration and processing. <http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/48/>, accessed: 2017-12-23
- [3] 3dm analyst. <https://www.adamtech.com.au/3dm/Analyst.html>, accessed: 2017-12-12
- [4] 3dm calibcam. <https://www.adamtech.com.au/3dm/CalibCam.html>, accessed: 2017-12-12
- [5] Aicon 3d. <http://www.aicon3d.com/>, accessed: 2017-12-11
- [6] Artec 3d scanners. <https://www.artec3d.com/3d-scanner/artec-eva>, accessed: 2017-12-23
- [7] Artec studio. <https://www.artec3d.com/3d-software/artec-studio>, accessed: 2017-12-23
- [8] Bentley contextcapture. <https://www.bentley.com/en/products/brands/contextcapture>, accessed: 2017-12-26
- [9] blender.org - Home of the Blender project - Free and Open 3D Creation Software. <https://www.blender.org/>, accessed: 2016-12-26
- [10] Canon eos 5d mark ii. https://en.wikipedia.org/wiki/Canon_EOS_5D_Mark_II, accessed: 2017-12-22

- [11] Companion software to 3d terrestrial scanners. <http://www.riegl.com/products/software-packages/riscan-pro/>, accessed: 2017-12-23
- [12] Developing with kinect for windows. <https://developer.microsoft.com/en-us/windows/kinect/develop>, accessed: 2016-11-19
- [13] Dji phantom 3. <https://www.dji.com/phantom-3-pro>, accessed: 2017-12-27
- [14] Faro pointsense. <https://www.faro.com/products/construction-bim-cim/faro-pointsense/>, accessed: 2017-12-23
- [15] Geomagic studio. <http://support1.geomagic.com/Support/5605/5668/en-US/Article/Folder/336/Geomagic-Studio>, accessed: 2017-12-23
- [16] Human Solutions GmbH. http://www.human-solutions.com/group/front_content.php?idart=29, accessed: 2016-11-27
- [17] Inpho. <https://geospatial.trimble.com/products-and-solutions/inpho>, accessed: 2017-12-11
- [18] Leica cyclone 7.3. http://w3.leica-geosystems.com/downloads123/hds/hds/cyclone/brochures-datasheet/Cyclone_VIEWER_Pro_7.3_DS_en_IX.11_gzd2_en.pdf, accessed: 2017-12-24
- [19] MakeHuman: Open source tool for making 3d characters. <http://www.makehuman.org/index.php>, accessed: 2016-12-26
- [20] Marvelous designer. <http://www.marvelousdesigner.com/>, accessed: 2016-11-18
- [21] Matlab. <https://se.mathworks.com/products/matlab.html>, accessed: 2017-12-11
- [22] Micmac. <http://micmac.ensg.eu/index.php/Accueil>, accessed: 2017-12-11
- [23] Nikon d100. https://en.wikipedia.org/wiki/Nikon_D100, accessed: 2017-12-23
- [24] Openmp. <http://www.openmp.org/>, accessed: 2017-12-21

- [25] Optitex 3d suite. <http://optitex.com/case-study/endura/>, accessed: 2016-11-19
- [26] Padding Materials. http://www.ottobock.se/cps/rde/xbcr/ob_se_sv/materials_11_padding_materials.pdf
- [27] Photomodeler. <http://www.photomodeler.com/index.html>, accessed: 2017-12-07
- [28] Ptgui 9. <https://www.ptgui.com/news/2013/12/16/ptgui-9.1.9-released.html>, accessed: 2017-12-22
- [29] Sensefly albris. <https://www.sensefly.com/drones/albris.html>, accessed: 2017-12-27
- [30] Usb connection diagram. <http://www.ianstedman.co.uk/>, accessed: 2016-11-19
- [31] Wikipedia modernized. <http://www.wikiwand.com/>, accessed: 2017-10-12
- [32] Abbas, M.A., Setan, H., Majid, Z., Chong, A.K., Idris, K.M., Aspuri, A.: Calibration and accuracy assessment of leica scanstation c10 terrestrial laser scanner. In: *Developments in Multidimensional Spatial Data Models*, pp. 33–47. Springer (2013)
- [33] Abels, A., Kruusmaa, M.: Construction of a female shape-changing robotic mannequin. *Journal of Automation and Control Engineering* 1(2) (2013)
- [34] Abels, A., Kruusmaa, M.: Shape control of an anthropomorphic tailoring robot mannequin. *International Journal of Humanoid Robotics* 10(02), 1350002 (2013)
- [35] Adeli-Mosabbeeb, E., Fathy, M., Zargari, F.: Model-based human gait tracking, 3d reconstruction and recognition in uncalibrated monocular video. *The Imaging Science Journal* 60(1), 9–28 (2012)
- [36] Albarelli, A., Cosmo, L., Bergamasco, F., Torsello, A.: High-coverage 3d scanning through online structured light calibration. 2014 22nd International Conference on Pattern Recognition (2014)
- [37] Allegra, D., Gallo, G., Inzerillo, L., Lombardo, M., Milotta, F., Santagati, C., Stanco, F.: Low cost handheld 3d scanning for architectural elements acquisition. In: *Proceedings of the Conference on Smart Tools and Applications in Computer Graphics*. pp. 127–131. Eurographics Association (2016)

- [38] Alsadik, B., Gerke, M., Vosselman, G.: Visibility analysis of point cloud in close range photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(5), 9 (2014)
- [39] Alsadik, B., Gerke, M., Vosselman, G.: Automated camera network design for 3d modeling of cultural heritage objects. *Journal of Cultural Heritage* 14(6), 515–526 (2013)
- [40] Altman, S., Xiao, W., Grayson, B.: Evaluation of low-cost terrestrial photogrammetry for 3d reconstruction of complex buildings. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4 (2017)
- [41] Altomonte, M., Zerbato, D., Botturi, D., Fiorini, P.: Simulation of Deformable Environment with Haptic Feedback on GPU. In: *International Conference on Intelligent Robots and Systems (IROS)*. pp. 3959–3964. *IEEE/RSJ* (2008)
- [42] Amrullah, C., Suwardhi, D., Meilano, I.: Product accuracy effect of oblique and vertical non-metric digital camera utilization in uav-photogrammetry to determine fault plane. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* pp. 41–48 (2016)
- [43] Anbarjafari, G., Rasti, P., Daneshmand, M., Ozcinar, C.: Resolution enhancement based image compression technique using singular value decomposition and wavelet transforms. In: *Wavelet Transform and Some of Its Real-World Applications*, pp. 35–52. *Intech* (2015)
- [44] Anderson, I.A., Gisby, T.A., McKay, T.G., O'Brien, B.M., Calius, E.P.: Multi-functional dielectric elastomer artificial muscles for soft and smart machines. *Journal of Applied Physics* 112(4), 041101 (2012)
- [45] Angeles, J.: *Fundamentals of Robotic Mechanical Systems: Theory, Methods, and Algorithms*. Springer (2006)
- [46] Aubreton, O., Bajard, A., Verney, B., Truchetet, F.: Infrared system for 3d scanning of metallic surfaces. *Machine vision and applications* 24(7), 1513–1524 (2013)
- [47] Avots, E., Daneshmand, M., Traumann, A., Escalera, S., Anbarjafari, G.: Automatic garment retexturing based on infrared information. *Computers & Graphics* 59, 28–38 (2016)
- [48] Babahajiani, P., Fan, L., Kamarainen, J.K., Gabbouj, M.: Comprehensive automated 3d urban environment modelling using terrestrial laser scanning

- point cloud. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 10–18 (2016)
- [49] Bailey, T., Durrant-Whyte, H.: Simultaneous localization and mapping (slam): Part ii. *IEEE Robotics & Automation Magazine* 13(3), 108–117 (2006)
- [50] Balter, C., Finkelstein, B.: Online shopping with virtual modeling and peer review (May 31 2005), uS Patent 6,901,379
- [51] Baltsavias, E.P.: A comparison between photogrammetry and laser scanning. *ISPRS Journal of photogrammetry and Remote Sensing* 54(2), 83–94 (1999)
- [52] Barber, C.B., Dobkin, D.P., Huhdanpaa, H.: The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* 22(4), 469–483 (1996)
- [53] Beardsley, P.A., Zisserman, A., Murray, D.W.: Sequential updating of projective and affine structure from motion. *International journal of computer vision* 23(3), 235–259 (1997)
- [54] Beheshti, I., Maikusa, N., Daneshmand, M., Matsuda, H., Demirel, H., Anbarjafari, G.: Classification of alzheimer’s disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks. *Journal of Alzheimer’s Disease* pp. 1–10 (2017)
- [55] Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., Taubin, G.: The ball-pivoting algorithm for surface reconstruction. *IEEE transactions on visualization and computer graphics* 5(4), 349–359 (1999)
- [56] Bertsekas, D.P.: *Network Optimization: Continuous and Discrete Methods*, vol. 8. Athena Scientific (1998)
- [57] Besl, P.J., McKay, N.D.: Method for Registration of 3-D Shapes. In: *Robotics-DL Tentative*. pp. 586–606. International Society for Optics and Photonics (1992)
- [58] Bi, H., Zheng, W., Ren, Z., Zeng, J., Yu, J.: Using an unmanned aerial vehicle for topography mapping of the fault zone based on structure from motion photogrammetry. *International Journal of Remote Sensing* 38(8-10), 2495–2510 (2017)

- [59] Blomley, R., Weinmann, M.: Using multi-scale features for the 3d semantic labeling of airborne laser scanning data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV-2/W4 (2017)
- [60] Boehler, W., Marbs, A.: 3d scanning instruments. *Proceedings of the CIPA WG 6*, 9–18 (2002)
- [61] Boissonnat, J.D., Cazals, F.: Smooth surface reconstruction via natural neighbour interpolation of distance functions. In: *Proceedings of the sixteenth annual symposium on Computational geometry*. pp. 223–232. ACM (2000)
- [62] Bolotnikova, A., Rasti, P., Traumann, A., Lusi, I., Daneshmand, M., Noroozi, F., Samuel, K., Sarkar, S., Anbarjafari, G.: Block based image compression technique using rank reduction and wavelet difference reduction. In: *Seventh International Conference on Graphic and Image Processing*. pp. 981702–981702. International Society for Optics and Photonics (2015)
- [63] Bone, D.J.: Fourier fringe analysis: the two-dimensional phase unwrapping problem. *Applied optics* 30(25), 3627–3632 (1991)
- [64] Bonev, I.A.: *Geometric Analysis of Parallel Mechanisms*. Ph.D. thesis, Laval university, Quebec, Canada (Nov 2002)
- [65] Bornaz, L., Rinaudo, F.: Terrestrial laser scanner data processing. In: *XXth ISPRS Congress Istanbul* (2004)
- [66] Bougleux, S., Elmoataz, A., Melkemi, M.: Discrete regularization on weighted graphs for image and mesh filtering. In: *Scale Space and Variational Methods in Computer Vision*, pp. 128–139. Springer (2007)
- [67] Bright, B.C., Loudermilk, E.L., Pokswinski, S.M., Hudak, A.T., O’Brien, J.J.: Introducing close-range photogrammetry for characterizing forest understory plant diversity and surface fuel structure at fine scales. *Canadian Journal of Remote Sensing* 42(5), 460–472 (2016)
- [68] Brutto, M.L., Meli, P.: Computer vision tools for 3d modelling in archaeology. *International Journal of Heritage in the Digital Era* 1(1_suppl), 1–6 (2012)

- [69] Buffa, F., Pinna, A., Sanna, G.: A simulation tool assisting the design of a close range photogrammetry system for the sardinia radio telescope. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 3(5) (2016)
- [70] Bylow, E., Sturm, J., Kerl, C., Kahl, F., Cremers, D.: Real-time camera tracking and 3d reconstruction using signed distance functions. In: *Robotics: Science and Systems*. vol. 2 (2013)
- [71] Calvert, P.: Inkjet printing for materials and devices. *Chemistry of materials* 13(10), 3299–3305 (2001)
- [72] Cardou, P., Bouchard, S., Gosselin, C.: Kinematic-sensitivity Indices for Dimensionally Nonhomogeneous Jacobian Matrices. *IEEE Transactions on Robotics* 26(1), 166–173 (2010)
- [73] Carignan, M., Yang, Y., Thalmann, N.M., Thalmann, D.: Dressing animated synthetic actors with complex deformable clothes. In: *ACM Siggraph Computer Graphics*. vol. 26, pp. 99–104. ACM (1992)
- [74] Carrozza, M.C., Suppo, C., Sebastiani, F., Massa, B., Vecchi, F., Lazzarini, R., Cutkosky, M.R., Dario, P.: The SPRING Hand: Development of a Self-Adaptive Prosthesis for Restoring Natural Grasping. *Autonomous Robots* 16(2), 125–141 (2004)
- [75] Carvajal-Ramírez, F., Agüera-Vega, F., Martínez-Carricondo, P.J.: Effects of image orientation and ground control points distribution on unmanned aerial vehicle photogrammetry projects on a road cut slope. *Journal of Applied Remote Sensing* 10(3), 034004–034004 (2016)
- [76] Casella, V., Franzini, M.: Modelling steep surfaces by various configurations of nadir and oblique photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 3(1) (2016)
- [77] Catherwood, T., Winder, J., McIntosh, S., Winder, R.: 3d stereophotogrammetry: post-processing and surface integration. *The Imaging Science Journal* 63(2), 95–102 (2015)
- [78] Chadès, I., Martin, T.G., Nicol, S., Burgman, M.A., Possingham, H.P., Buckley, Y.M.: General rules for managing and surveying networks of pests, diseases, and endangered species. *Proceedings of the National Academy of Sciences* 108(20), 8323–8328 (2011)

- [79] Chen, C., He, S., Chen, G., Cao, H.: Photogrammetry-based 3d printing reproduction method for oil paintings. *International Journal of Pattern Recognition and Artificial Intelligence* 32(03), 1854007 (2017)
- [80] Chen, C.Y., Shih, B.Y., Shih, C.H., Wang, L.H.: RETRACTED: Enhancing Robust and Stability Control of a Humanoid Biped Robot: System Identification Approach. *Journal of Vibration and Control* 19(8), 1199–1207 (2013)
- [81] Chen, T., Lensch, H.P., Fuchs, C., Seidel, H.P.: Polarization and phase-shifting for 3d scanning of translucent objects. In: *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. pp. 1–8. IEEE (2007)
- [82] Chen, Y., Medioni, G.: Object Modeling by Registration of Multiple Range Images. In: *Proceedings of the International Conference on Robotics and Automation*. pp. 2724–2729. IEEE (1991)
- [83] Chetverikov, D., Stepanov, D., Krsek, P.: Robust euclidean alignment of 3d point sets: the trimmed iterative closest point algorithm. *Image and Vision Computing* 23(3), 299–309 (2005)
- [84] Chittaro, L., Corvaglia, D.: 3d virtual clothing: from garment design to web3d visualization and simulation. In: *Proceedings of the eighth international conference on 3D Web technology*. pp. 73–ff. ACM (2003)
- [85] Christensen, J., Marks, J., Ngo, J.T.: Automatic Motion Synthesis for 3D Mass-Spring Models. *The Visual Computer* 13(1), 20–28 (1997)
- [86] Colombo, R., Pisano, F., Micera, S., Mazzone, A., Delconte, C., Carrozza, M.C., Dario, P., Minuco, G.: Robotic Techniques for Upper Limb Evaluation and Rehabilitation of Stroke Patients. *Transactions on Neural Systems and Rehabilitation Engineering* 13(3), 311–324 (2005)
- [87] Colomina, I., Molina, P.: Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 92, 79–97 (2014)
- [88] Cong, P., Xiong, Z., Zhang, Y., Zhao, S., Wu, F.: Dense single-shot 3d scanning via stereoscopic fringe analysis. *2013 IEEE International Conference on Image Processing* (2013)
- [89] Controzzi, M., Cipriani, C., Carrozza, M.C.: Design of Artificial Hands: A Review. In: *The Human Hand as an Inspiration for Robot Hand Development*, pp. 219–246. Springer (2014)

- [90] Cordier, F., Seo, H., Magnenat-Thalmann, N.: Made-to-measure technologies for an online clothing store. *IEEE Computer Graphics and Applications* 23(1), 38–48 (Jan 2003)
- [91] Cordier, F., Lee, W., Seo, H., Magnenat-Thalmann, N.: Virtual-try-on on the web. *Laval Virtual* (2001)
- [92] Cucci, D.: Accurate optical target pose determination for applications in aerial photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 3(3) (2016)
- [93] Cugini, U., Rizzi, C.: 3d design and simulation of men garments (2002)
- [94] Cui, Y., Chang, W., Nöll, T., Stricker, D.: KinectAvatar: Fully Automatic Body Capture Using a Single Kinect. In: *ACCV Workshops* (2). pp. 133–147. Citeseer (2012)
- [95] Cui, Y., Schuon, S., Thrun, S., Stricker, D., Theobalt, C.: Algorithms for 3d shape scanning with a depth camera. *IEEE transactions on pattern analysis and machine intelligence* 35(5), 1039–1050 (2013)
- [96] Cunliffe, A.M., Brazier, R.E., Anderson, K.: Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sensing of Environment* 183, 129–143 (2016)
- [97] Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. pp. 303–312. ACM (1996)
- [98] Da, T.K.F., Yvinec, M.: 3d alpha shapes. *CGAL user and reference manual* 4(2) (2013)
- [99] Dame, A., Prisacariu, V.A., Ren, C.Y., Reid, I.: Dense Reconstruction Using 3D Object Shape Priors. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1288–1295. IEEE (2013)
- [100] Daneshmand, M., Helmi, A., Avots, E., Noroozi, F., Alisinanoglu, F., Sait Arslan, H., Gorbova, J., Haamer, R.E., Ozcinar, C., Anbarjafari, G.: 3D Scanning: A Comprehensive Survey. *ArXiv e-prints* (Jan 2018)
- [101] Daneshmand, M., Aabloo, A., Anbarjafari, G.: Size-dictionary interpolation for robot’s adjustment. *Frontiers in bioengineering and biotechnology* 3 (2015)

- [102] Daneshmand, M., Aabloo, A., Ozcinar, C., Anbarjafari, G.: Real-time, Automatic Shape-Changing Robot Adjustment and Gender Classification. *Signal, Image and Video Processing* pp. 1–8 (2015)
- [103] Daneshmand, M., Aabloo, A., Ozcinar, C., Anbarjafari, G.: Real-time, automatic shape-changing robot adjustment and gender classification. *Signal, Image and Video Processing* 10(4), 753–760 (2016)
- [104] Daneshmand, M., Abels, A., Anbarjafari, G.: Real-time, automatic digital mannequin robot adjustment based on human body classification through supervised learning. *International Journal of Advanced Robotic Systems* 14(3), 1–9 (2017)
- [105] Daneshmand, M., Bilici, O., Bolotnikova, A., Anbarjafari, G.: Medical robots with potential applications in participatory and opportunistic remote sensing: A review. *Robotics and Autonomous Systems* (2017)
- [106] Daneshmand, M., Masouleh, M.T., Anbarjafari, G.: Kinematic sensitivity analysis of a 3-dof decoupled translational parallel mechanism with uncertainties in the passive joints. In: *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*. pp. 85–90. IEEE (2015)
- [107] Daneshmand, M., Masouleh, M.T., Anbarjafari, G.: Kinematic sensitivity analysis of parallel mechanisms by considering the effect of uncertainties in passive joints. *Modares Mechanical Engineering* 15(10), 78–88 (2016)
- [108] Daneshmand, M., Masouleh, M.T., Saadatzi, M.H., Ozcinar, C., Anbarjafari, G.: A robust proportion-preserving composite objective function for scale-invariant multi-objective optimization. *Scientia Iranica* 24(6), 2977–2991 (2017)
- [109] Daneshmand, M., Saadatzi, M.H., Kaloorazi, M.H.F., Masouleh, M.T., Anbarjafari, G.: Optimal design of a spherical parallel manipulator based on kinetostatic performance using evolutionary techniques. *Journal of Mechanical Science and Technology* 30(3), 1323–1331 (2016)
- [110] Daneshmand, M., Saadatzi, M.H., Masouleh, M.T.: Kinematic sensitivity and workspace optimization of planar parallel mechanisms using evolutionary techniques. In: *Robotics and Mechatronics (ICRoM), 2013 First RSI/ISM International Conference on*. pp. 384–389. IEEE (2013)
- [111] Daneshmand, M., Saadatzi, M.H., Masouleh, M.T.: Kinematic Sensitivity and Workspace Optimization of Planar Parallel Mechanisms Using

- Evolutionary Techniques. In: First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM). pp. 384–389. IEEE (2013)
- [112] Daneshmand, M., Tale Masouleh, M., Saadatzi, M.H.: Optimization of the kinematic sensitivity and the greatest continuous circle in the constant-orientation workspace of planar parallel mechanisms. *International Journal of Robotics, Theory and Applications* 4(1), 12–21 (2016)
- [113] Daneshmand, M., Traumann, A., Anbarjafari, G.: 3d size-estimation based on the geodesic distance measured by photogrammetric scanning device. In: *Proc. of 6th Int. Conf. on 3D Body Scanning Technologies*. pp. 221–231 (2015)
- [114] Daneshmand, M., Traumann, A., Anbarjafari, G.: 3d size-estimation based on the geodesic distance measured by photogrammetric scanning device. In: *Proceedings of the Sixth International Conference on 3D Body Scanning Technologies*. pp. 221–231 (2015), lugano, Switzerland
- [115] Dario, P., Hannaford, B., Menciassi, A.: Smart Surgical Tools and Augmenting Devices. *Transactions on Robotics and Automation* 19(5), 782–792 (2003)
- [116] De Boor, C., De Boor, C., Mathématicien, E.U., De Boor, C., De Boor, C.: *A practical guide to splines*, vol. 27. Springer-Verlag New York (1978)
- [117] DeWitt, J., Warner, T., Conley, J.: Comparison of dems derived from usgs dlg, srtm, a statewide photogrammetry program, aster gdem and lidar: implications for change detection. *GIScience and Remote Sensing* (2015)
- [118] Divivier, A., Trieb, R., Ebert, A., Hagen, H., Gross, C., Fuhrmann, A., Luckas, V., et al.: Virtual try-on topics in realistic, individualized dressing in virtual reality (2004)
- [119] Dorai, C., Wang, G., Jain, A.K., Mercer, C.: Registration and Integration of Multiple Object Views for 3D Model Construction. *Transactions on Pattern Analysis and Machine Intelligence* 20(1), 83–89 (1998)
- [120] Dou, M., Taylor, J., Fuchs, H., Fitzgibbon, A., Izadi, S.: 3d scanning deformable objects with a single rgbd sensor. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 493–501 (2015)
- [121] Drosos, V., Fidani, S., Manesis, C.: Use of photogrammetry and gis in the survey of natural areas of special protection: case study of aesthetic forest kouri almyrou magnisias. In: *First International Conference on Remote*

Sensing and Geoinformation of Environment. pp. 87950C–87950C. International Society for Optics and Photonics (2013)

- [122] Du, H., Xiong, W., Wang, H., Yuan, B., Wang, Z.: Configuration modeling and experimental verification with 3d laser scanning technology for a constrained elastica cable. *International Journal of Signal Processing, Image Processing and Pattern Recognition* 7(4), 363–370 (2014)
- [123] Duan, K.B., Keerthi, S.S.: Which is the best multiclass svm method? an empirical study. In: *International Workshop on Multiple Classifier Systems*. pp. 278–285. Springer (2005)
- [124] Duan, Y., Yan, L., Zhong, Y., Yao, J.: A new method of improving height accuracy of airborne photogrammetry using a multi-camera system. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. pp. 2019–2022. IEEE (2014)
- [125] Duguet, F., Drettakis, G., Girardeau-Montaut, D., Martinez, J.L., Schmitt, F.: A point-based approach for capture, display and illustration of very complex archeological artefacts. In: *International Symposium on Virtual Reality, Archeology and Cultural Heritage (VAST2004)*. pp. 105–114. Eurographics (2004)
- [126] El-Ashmawy, K.L.: A comparison between analytical aerial photogrammetry, laser scanning, total station and global positioning system surveys for generation of digital terrain model. *Geocarto International* 30(2), 154–162 (2015)
- [127] El-Omari, S., Moselhi, O.: Integrating 3d laser scanning and photogrammetry for progress measurement of construction work. *Automation in construction* 18(1), 1–9 (2008)
- [128] Evans, S., Jones, C., Plassmann, P.: 3d imaging for bite mark analysis. *The Imaging Science Journal* (2013)
- [129] Fangi, G., Malinverni, E., Tasseti, A.: The metric documentation of cham towers in vietnam by spherical photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 5, W1 (2013)
- [130] Fey, C., Rutzinger, M., Wichmann, V., Prager, C., Bremer, M., Zangerl, C.: Deriving 3d displacement vectors from multi-temporal airborne laser scanning data for landslide activity analyses. *GIScience & Remote Sensing* 52(4), 437–461 (2015)

- [131] Filliat, D.: A visual bag of words method for interactive qualitative localization and mapping. In: *Robotics and Automation, 2007 IEEE International Conference on*. pp. 3921–3926. IEEE (2007)
- [132] Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), 381–395 (1981)
- [133] Fitzgibbon, A.W., Cross, G., Zisserman, A.: Automatic 3D Model Construction for Turntable Sequences. In: *3D Structure from Multiple Images of Large-Scale Environments*, pp. 155–170. Springer (1998)
- [134] Fleishman, S., Cohen-Or, D., Silva, C.T.: Robust moving least-squares fitting with sharp features. In: *ACM Transactions on Graphics (TOG)*. vol. 24, pp. 544–552. ACM (2005)
- [135] Frank Steinbrucker, Christian Kerl, J.S., Cremers, D.: Large-scale multi-resolution surface reconstruction from rgb-d sequences. In: *Proceedings of the 2013 IEEE International Conference on Computer Vision*. IEEE (2013)
- [136] Fraser, C.S.: Automatic camera calibration in close range photogrammetry. *Photogrammetric Engineering & Remote Sensing* 79(4), 381–388 (2013)
- [137] Fritsch, F.N., Carlson, R.E.: Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis* 17(2), 238–246 (1980)
- [138] Fröhlich, C., Mettenleiter, M., et al.: Terrestrial laser scanning-new perspectives in 3d surveying. *International archives of photogrammetry, remote sensing and spatial information sciences* 36(Part 8), W2 (2004)
- [139] Fuciños, M., López, J., Pardo, X.M., Fdez-Vidal, X.R.: Fast implementation of a new radial symmetry measure for photogrammetry. In: *Iberian Conference on Pattern Recognition and Image Analysis*. pp. 221–228. Springer (2013)
- [140] Fujita, M., Kitano, H., Kageyama, K.: A Reconfigurable Robot Platform. *Robotics and Autonomous Systems* 29(2), 119–132 (1999)
- [141] Gagliardini, L., Caro, S., Gouttefarde, M., Wenger, P., Girin, A.: A Reconfigurable Cable-Driven Parallel Robot for Sandblasting and Painting of Large Structures. In: *Cable-Driven Parallel Robots*, pp. 275–291. Springer (2015)
- [142] Geibig, T., Shoykhetbrod, A., Hommes, A., Herschel, R., Pohl, N.: Compact 3d imaging radar based on fmcw driven frequency-scanning antennas. In: *Radar Conference (RadarConf), 2016 IEEE*. pp. 1–5. IEEE (2016)

- [143] Georgii, J., Westermann, R.: Mass-Spring Systems on the GPU. *Simulation Modelling Practice and Theory* 13(8), 693–702 (2005)
- [144] Gil, A.L., Núñez-Casillas, L., Isenburg, M., Benito, A.A., Bello, J.J.R., Arbelo, M.: A comparison between lidar and photogrammetry digital terrain models in a forest area on tenerife island. *Canadian Journal of Remote Sensing* 39(05), 396–409 (2013)
- [145] Gloe, T., Böhme, R.: The dresden image database for benchmarking digital image forensics. *Journal of Digital Forensic Practice* 3(2-4), 150–159 (2010)
- [146] Gonçalves, J., Henriques, R.: Uav photogrammetry for topographic monitoring of coastal areas. *ISPRS Journal of Photogrammetry and Remote Sensing* (2015)
- [147] Gorte, B., Vargas-Luna, A., Sirmacek, B.: Camera-projector 3d scanning of a semi-submerged terrain in a flume. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-4/w1* pp. 13–18 (2013)
- [148] Granshaw, S.: Close range photogrammetry: Principles, methods and applications. *The Photogrammetric Record* 25(130), 203–204 (2010)
- [149] Grisetti, G., Stachniss, C., Grzonka, S., Burgard, W.: Toro project at openslam.org (2007)
- [150] Gruszczyński, W., Matwij, W., Cwiąkała, P.: Comparison of low-altitude uav photogrammetry with terrestrial laser scanning as data-source methods for terrain covered in low vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing* 126, 168–179 (2017)
- [151] Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B.: Using mobile laser scanning data for automated extraction of road markings. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 93–107 (2014)
- [152] Guo, J., Jiang, J., Wu, L., Zhou, W., Wei, L.: 3d modeling for mine roadway from laser scanning point cloud. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. pp. 4452–4455. IEEE (2016)
- [153] Han, B., Endreny, T.A.: River surface water topography mapping at sub-millimeter resolution and precision with close range photogrammetry: Laboratory scale application. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(2), 602–608 (2014)

- [154] Hanqing, M., Dong, W., Junfeng, S., Jiang, T.: A wideband 45 degree polarized electrical scanning antenna array manufactured by 3d printing with metals. In: Radar (RADAR), 2016 CIE International Conference on. pp. 1–3. IEEE (2016)
- [155] Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1), 100–108 (1979)
- [156] Harwin, S., Lucieer, A.: Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (uav) imagery. *Remote Sensing* 4(6), 1573–1599 (2012)
- [157] Hattab, A., Gonsler, I., Moreno, D., Taubin, G.: Differential 3d scanning. *IEEE Computer Graphics and Applications* 38(3), 43–51 (2017)
- [158] Hauswiesner, S., Straka, M., Reitmayr, G.: Free viewpoint virtual try-on with commodity depth cameras. In: *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*. pp. 23–30. ACM (2011)
- [159] Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In: *In the 12th International Symposium on Experimental Robotics (ISER*. Citeseer (2010)
- [160] Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research* 31(5), 647–663 (2012)
- [161] Herlihy, M., Luchangco, V., Moir, M., Scherer III, W.N.: Software transactional memory for dynamic-sized data structures. In: *Proceedings of the twenty-second annual symposium on Principles of distributed computing*. pp. 92–101. ACM (2003)
- [162] Hoffmeister, D., Zellmann, S., Kindermann, K., Pastoors, A., Lang, U., Bubenzler, O., Weniger, G., Bareth, G.: Geoarchaeological site documentation and analysis of 3d data derived by terrestrial laser scanning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(5), 173 (2014)
- [163] Holopainen, M., Vastaranta, M., Karjalainen, M., Karila, K., Kaasalainen, S., Honkavaara, E., Hyypä, J.: Forest inventory attribute estimation using airborne laser scanning, aerial stereo imagery, radargrammetry and

- interferometry-finnish experiences of the 3d techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(3), 63 (2015)
- [164] Horn, B.K.: Closed-Form Solution of Absolute Orientation Using Unit Quaternions. *JOSA A* 4(4), 629–642 (1987)
- [165] Hu, G., Peng, Q., Forrest, A.R.: Mean shift denoising of point-sampled surfaces. *The Visual Computer* 22(3), 147–157 (2006)
- [166] Huang, H., Brenner, C., Sester, M.: A generative statistical approach to automatic 3d building roof reconstruction from laser scanning data. *ISPRS Journal of photogrammetry and remote sensing* 79, 29–43 (2013)
- [167] Huang, H., Ma, X., Zhao, L., Du, H., Luo, H., Mao, X., Tang, M., Liu, Y.: Transmission line icing measurement on photogrammetry method. In: *Ninth International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2015)*. pp. 98150Z–98150Z. International Society for Optics and Photonics (2015)
- [168] Huang, H., Li, D., Zhang, H., Ascher, U., Cohen-Or, D.: Consolidation of unorganized point clouds for surface reconstruction. In: *ACM transactions on graphics (TOG)*. vol. 28, p. 176. ACM (2009)
- [169] Hubert, M., Debruyne, M.: Minimum covariance determinant. *Wiley interdisciplinary reviews: Computational statistics* 2(1), 36–43 (2010)
- [170] Impoco, G., Cignoni, P., Scopigno, R.: Closing gaps by clustering unseen directions. In: *Shape Modeling Applications, 2004. Proceedings*. pp. 307–316. IEEE (2004)
- [171] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., et al.: Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In: *Proceedings of the 24th annual ACM symposium on User interface software and technology*. pp. 559–568. ACM (2011)
- [172] Jiang, R., Tang, W., Wu, X., Fu, W.: A random forest approach to the setec-tion of epistatic interactions in case-control studies. *BMC bioinformatics* 10(Suppl 1), S65 (2009)
- [173] Jung, S.W.: Enhancement of image and depth map using adaptive joint trilateral filter. *Circuits and Systems for Video Technology, IEEE Transactions on* 23(2), 258–269 (2013)

- [174] Kadri, O., Mouss, L.H., Mouss, M.D.: Fault diagnosis of rotary kiln using svm and binary aco. *Journal of mechanical science and technology* 26(2), 601–608 (2012)
- [175] Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J.J., Dellaert, F.: isam2: Incremental smoothing and mapping using the bayes tree. *The International Journal of Robotics Research* p. 0278364911430419 (2011)
- [176] Kaess, M., Ranganathan, A., Dellaert, F.: isam: Incremental smoothing and mapping. *Robotics, IEEE Transactions on* 24(6), 1365–1378 (2008)
- [177] Kamberelis, G., Dimitriadis, G.: Collectively remembering tupac. *Afterlife as Afterimage: Understanding Posthumous Fame* 2, 143 (2005)
- [178] Kandare, K., Ørka, H.O., Dalponte, M., Næsset, E., Gobakken, T.: Individual tree crown approach for predicting site index in boreal forests using airborne laser scanning and hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation* 60, 72–82 (2017)
- [179] Kerschner, M.: Homologous twin snakes integrated in a bundle block adjustment. *International Archives of Photogrammetry and Remote Sensing* 32, 244–249 (1998)
- [180] Khan, W., Angeles, J.: The Kinetostatic Optimization of Robotic Manipulators: The Inverse and the Direct Problems. *Journal of Mechanical Design* 128(1), 168–178 (2006)
- [181] Kim, D.K., Kim, J.H., Kwon, H.J., Kwon, Y.H.: Fabrication of robot head module using contact-resistance force sensor for human-robot interaction and its evaluation. *Journal of mechanical science and technology* 26(10), 3269–3276 (2012)
- [182] Kim, D.M., Kim, H.J.: Digital single lens reflex (dslr) camera (Mar 26 2013), uS Patent 8,403,573
- [183] Kim, H.E., Hwang, S.S., Tan, A.C., Mathew, J., Choi, B.K.: Integrated approach for diagnostics and prognostics of hp lng pump based on health state probability estimation. *Journal of mechanical science and technology* 26(11), 3571–3585 (2012)
- [184] Kim, J., Forsythe, S.: Adoption of Virtual Try-on Technology for Online Apparel Shopping. *Journal of Interactive Marketing* 22(2), 45–59 (2008)
- [185] Kim, K., Lawrence, R.L., Kyllonen, N., Ludewig, P.M., Ellingson, A.M., Keefe, D.F.: Anatomical 2d/3d shape-matching in virtual reality: A user

- interface for quantifying joint kinematics with radiographic imaging. In: 3D User Interfaces (3DUI), 2017 IEEE Symposium on. pp. 243–244. IEEE (2017)
- [186] Kim, Y., Jang, J., Kim, W., Roh, T.S., Choi, D.W.: Multiple defect diagnostics of gas turbine engine using svm and rcga-based ann algorithms. *Journal of mechanical science and technology* 26(5), 1623–1632 (2012)
- [187] Kocak, D.M., Caimi, F.M., Jagielo, T.H., Kloske, J.: Laser Projection Photogrammetry and Video System for Quantification and Mensuration. In: OCEANS’02 MTS. vol. 3, pp. 1569–1574. IEEE (2002)
- [188] Kohlschütter, T.: Human body modelling by development of the automatic landmarking algorithm (2012)
- [189] Koller, D., Lindstrom, P., Ribarsky, W., Hodges, L.F., Faust, N., Turner, G.: Virtual gis: A real-time 3d geographic information system. In: Proceedings of the 6th conference on Visualization’95. p. 94. IEEE Computer Society (1995)
- [190] Konovalov, V., Clapés, A., Escalera, S.: Automatic Hand Detection in RGB-Depth Data Sequences. In: CCIA. pp. 91–100 (2013)
- [191] Kruusmaa, M., Abels, A.: Design of a Shape-Changing Anthropomorphic Mannequin for Tailoring Applications. In: International Conference on Advanced Robotics (ICAR). pp. 1–6. IEEE (2009)
- [192] Kühnapfel, U., Cakmak, H.K., Maaß, H.: Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers & graphics* 24(5), 671–682 (2000)
- [193] Kuzu, Y.: Volumetric object reconstruction by means of photogrammetry. Ph.D. thesis (2004)
- [194] Lai, K., Bo, L., Ren, X., Fox, D.: A large-scale hierarchical multi-view rgb-d object dataset. In: Robotics and Automation (ICRA), 2011 IEEE International Conference on. pp. 1817–1824. IEEE (2011)
- [195] Lai, X.b., Wang, H.s., Xu, Y.h.: A real-time range finding system with binocular stereo vision. *Int J Adv Robotic Sy* 9(27) (2012)
- [196] Laur, L., Daneshmand, M., Agoyi, M., Anbarjafari, G.: Robust grayscale watermarking technique based on face detection. In: Signal Processing and Communications Applications Conference (SIU), 2015 23th. pp. 471–475. IEEE (2015)

- [197] Le, A.V., Jung, S.W., Won, C.S.: Directional joint bilateral filter for depth images. *Sensors* 14(7), 11362–11378 (2014)
- [198] Lee, D.T., Schachter, B.J.: Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences* 9(3), 219–242 (1980)
- [199] Lee, S., Majid, Z., Setan, H.: 3d data acquisition for indoor assets using terrestrial laser scanning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, W1 (2013)
- [200] Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., et al.: The digital michelangelo project: 3d scanning of large statues. In: *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. pp. 131–144. ACM Press/Addison-Wesley Publishing Co. (2000)
- [201] Ley, A., Hänsch, R., Hellwich, O.: Automatic building abstraction from aerial photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV-2/W4 (2017)
- [202] Li, B., Schnabel, R., Klein, R., Cheng, Z., Dang, G., Jin, S.: Robust normal estimation for point clouds with sharp features. *Computers & Graphics* 34(2), 94–106 (2010)
- [203] Li, H.y., Wu, X.r.: The Construction of 3D-Mannequin-based on Human Body Characteristic Lines. In: *Second International Conference on Information and Computing Science*. vol. 2, pp. 341–344 (2009)
- [204] Li, J., Ye, J., Wang, Y., Bai, L., Lu, G.: Fitting 3D Garment Models onto Individual Human Models. *Computers & graphics* 34(6), 742–755 (2010)
- [205] Li, J., Chen, J.: A Mannequin-Modeling Method based on Section Templates and Silhouette Control. *International Journal of Clothing Science and Technology* 21(5), 300–310 (2009)
- [206] Li, R., Zou, K., Xu, X., Li, Y., Li, Z.: Research of interactive 3d virtual fitting room on web environment. In: *Computational Intelligence and Design (ISCID), 2011 Fourth International Symposium on*. vol. 1, pp. 32–35. IEEE (2011)
- [207] Li, W., Grossman, T., Fitzmaurice, G.: Gamicad: a gamified tutorial system for first time autocad users. In: *Proceedings of the 25th annual ACM symposium on User interface software and technology*. pp. 103–112. ACM (2012)

- [208] Li, Y., Sun, J., Tang, C.K., Shum, H.Y.: Lazy snapping. In: *ACM Transactions on Graphics (ToG)*. vol. 23, pp. 303–308. ACM (2004)
- [209] Li, Z., Jia, W., Mao, Z.H., Li, J., Chen, H.C., Zuo, W., Wang, K., Sun, M.: Anthropometric body measurements based on multi-view stereo image reconstruction. In: *35th Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*. pp. 366–369. IEEE (2013)
- [210] Li, Z., Yan, X., Yuan, C., Peng, Z.: Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed. *Journal of Mechanical Science and Technology* 26(8), 2413–2423 (2012)
- [211] Lindeberg, T.: Scale invariant feature transform. *Scholarpedia* 7(5), 10491 (2012)
- [212] Liu, L., Zhang, L., Xu, Y., Gotsman, C., Gortler, S.J.: A local/global approach to mesh parameterization. In: *Computer Graphics Forum*. vol. 27, pp. 1495–1504. Wiley Online Library (2008)
- [213] Liu, X.H., Wu, Y.W.: A 3d display system for cloth online virtual fitting room. In: *2009 WRI World Congress on Computer Science and Information Engineering* (2009)
- [214] Løkberg, O.J.: Electronic speckle pattern interferometry. In: *Optical Metrology*, pp. 542–572. Springer (1987)
- [215] López, J., Fuciños, M., Fdez-Vidal, X.R., Pardo, X.M.: Detection and matching of lines for close-range photogrammetry. In: *Iberian Conference on Pattern Recognition and Image Analysis*. pp. 732–739. Springer (2013)
- [216] Lorensen, W.E., Cline, H.E.: Marching Cubes: A High-Resolution 3D Surface Construction Algorithm. In: *ACM Siggraph Computer Graphics*. vol. 21, pp. 163–169. ACM (1987)
- [217] Lourakis, M.I., Argyros, A.A.: Sba: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)* 36(1), 2 (2009)
- [218] Low, K.L.: *Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration*. Chapel Hill, University of North Carolina (2004)
- [219] Lu, C.P., Hager, G.D., Mjølness, E.: Fast and globally convergent pose estimation from video images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(6), 610–622 (2000)

- [220] Lu, N., Wang, Q., Wang, S., Zhang, R.: The application of 3d laser scanning in the survey and measuring of guyue bridge of song dynasty in yiwu city. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(5), 185 (2015)
- [221] Lucet, G.: 3d survey of pre-hispanic wall painting with high resolution photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* (1), 191–196 (2013)
- [222] Luhmann, T., Fraser, C., Maas, H.G.: Sensor modelling and camera calibration for close-range photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 115, 37–46 (2016)
- [223] Lüsi, I., Anbarjafari, G., Meister, E.: Real-time mimicking of estonian speaker’s mouth movements on a 3d avatar using kinect 2. In: *Information and Communication Technology Convergence (ICTC), 2015 International Conference on*. pp. 141–143. IEEE (2015)
- [224] Madadi, M., Escalera, S., González, J., Roca, F.X., Lumbreras, F.: Multi-Part body Segmentation based on Depth Maps for Soft Biometry Analysis. *Pattern Recognition Letters* 56, 14–21 (2015)
- [225] Majasalmi, T., Korhonen, L., Korpela, I., Vauhkonen, J.: Application of 3d triangulations of airborne laser scanning data to estimate boreal forest leaf area index. *International Journal of Applied Earth Observation and Geoinformation* (2017)
- [226] Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A Review of 3D/2D Registration Methods for Image-Guided Interventions. *Medical image analysis* 16(3), 642–661 (2012)
- [227] Mayr, A., Rutzinger, M., Bremer, M., Geitner, C.: Mapping eroded areas on mountain grassland with terrestrial photogrammetry and object-based image analysis. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 3(5) (2016)
- [228] Meagher, D.: Geometric modeling using octree encoding. *Computer graphics and image processing* 19(2), 129–147 (1982)
- [229] Mechelke, K., Kersten, T.P., Lindstaedt, M.: Comparative investigations into the accuracy behaviour of the new generation of terrestrial laser scanning systems. *Proc. in the Optical* 3, 19–327 (2007)
- [230] Merlet, J.: Jacobian, Manipulability, Condition Number, and Accuracy of Parallel Robots. *Journal of Mechanical Design* 128, 199–206 (2006)

- [231] Mrovlje, J., Vrancic, D.: Distance measuring based on stereoscopic pictures. In: 9th International PhD Workshop on Systems and Control, Izola, Slovenia (2008)
- [232] Murtiyoso, A., Koehl, M., Grussenmeyer, P., Freville, T.: Acquisition and processing protocols for uav images: 3d modeling of historical buildings using photogrammetry. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4 (2017)
- [233] Nakamura, R., Izutsu, M., Hatakeyama, S.: Estimation method of clothes size for virtual fitting room with kinect sensor. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics. pp. 3733–3738. IEEE (2013)
- [234] Narváez, E.A.L., Narváez, N.E.L.: Point cloud denoising using robust principal component analysis. In: GRAPP. pp. 51–58 (2006)
- [235] Neubeck, A., Van Gool, L.: Efficient non-maximum suppression. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. vol. 3, pp. 850–855. IEEE (2006)
- [236] Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on. pp. 127–136. IEEE (2011)
- [237] Nex, F., Gerke, M., Remondino, F., Przybilla, H., Bäumker, M., Zurhorst, A.: Isprs benchmark for multi-platform photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(3), 135 (2015)
- [238] Nguyen, C.V., Izadi, S., Lovell, D.: Modeling kinect sensor noise for improved 3d reconstruction and tracking. In: 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on. pp. 524–530. IEEE (2012)
- [239] Noblet, V., Heinrich, C., Heitz, F., Armspach, J.P.: 3-D Deformable Image Registration: A Topology Preservation Scheme based on Hierarchical Deformation Models and Interval Analysis Optimization. *Transactions on Image Processing* 14(5), 553–566 (2005)
- [240] Nocerino, E., Menna, F., Remondino, F., Saleri, R.: Accuracy and block deformation analysis in automatic uav and terrestrial photogrammetry-lesson

- learnt. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, 5 (2013)
- [241] Nurunnabi, A., Belton, D., West, G.: Diagnostic-robust statistical analysis for local surface fitting in 3d point cloud data. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Science, Volumes I-3* pp. 269–274 (2012)
- [242] Nurunnabi, A., Belton, D., West, G.: Robust statistical approaches for local planar surface fitting in 3d laser scanning data. *ISPRS journal of photogrammetry and Remote Sensing* 96, 106–122 (2014)
- [243] Nurunnabi, A., West, G., Belton, D.: Outlier detection and robust normal-curvature estimation in mobile laser scanning 3d point cloud data. *Pattern Recognition* 48(4), 1404–1419 (2015)
- [244] Oguma, K.: High-speed video camera (Aug 30 2011), uS Patent App. 13/221,282
- [245] Pan, H., Shi, L., Min, Z., Dai, Y.: 3d modelling of elastic articulated objects and its parameter determination from image points. *The Imaging Science Journal* (2013)
- [246] Paquet, E., Viktor, H.L.: Adjustment of Virtual Mannequins through Anthropometric Measurements, Cluster Analysis and Content-based Retrieval of 3-D Body Scans (2007)
- [247] Peckar, W., Schnörr, C., Rohr, K., Stiehl, H.S.: Parameter-Free Elastic Deformation Approach for 2D and 3D Registration Using Prescribed Displacements. *Journal of Mathematical Imaging and Vision* 10(2), 143–162 (1999)
- [248] Pereira, F., Silva, C., Alves, M.: Virtual fitting room augmented reality techniques for e-commerce. In: *International Conference on ENTERprise Information Systems*. pp. 62–71. Springer (2011)
- [249] Piltan, F., Keshavarz, M., Badri, A., Zargari, A.: Design Novel Nonlinear Controller Applied to Robot Manipulator: Design New Feedback Linearization Fuzzy Controller with Minimum Rule base Tuning Method. *International Journal of Robotics and Automation* 3(1), 1–12 (2012)
- [250] Pjatkin, K., Daneshmand, M., Rasti, P., Anbarjafari, G.: Probability distribution function based iris recognition boosted by the mean rule. In: *Intelligent Computing and Internet of Things (ICIT), 2014 International Conference on*. pp. 47–50. IEEE (2015)

- [251] Poon, G., Yeung, Y.Y., Pang, W.M.: Enabling 3D Online Shopping with Affordable Depth Scanned Models. In: International Conference on Smart Computing (SMARTCOMP). pp. 150–155. IEEE (2014)
- [252] Prashanth, H., Shashidhara, H., KN, B.M.: Image scaling comparison using universal image quality index. In: Advances in Computing, Control, & Telecommunication Technologies, 2009. ACT'09. International Conference on. pp. 859–863. IEEE (2009)
- [253] Protopsaltou, D., Luible, C., Arevalo, M., Magnenat-Thalmann, N.: A body and garment creation method for an internet based virtual fitting room. In: Advances in Modelling, Animation and Rendering, pp. 105–122. Springer (2002)
- [254] Qin, R., Gruen, A.: 3d change detection at street level using mobile laser scanning point clouds and terrestrial images. *ISPRS Journal of Photogrammetry and Remote Sensing* 90, 23–35 (2014)
- [255] Rahrig, M., Luib, A.: Sri dalada maligawa-3d-scanning and documentation of the temple of the sacred tooth relic at kandy, sri lanka. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4 (2017)
- [256] Rao, Y.S., Deo, R., Nalini, J., Pillai, A.M., Muralikrishnan, S., Dadhwal, V.K.: Quality assessment of tandem-x dems using airborne lidar, photogrammetry and icesat elevation data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume II-8* (2014)
- [257] Rasti, P., Daneshmand, M., Alisinanoglu, F., Ozcinar, C., Anbarjafari, G.: Medical image illumination enhancement and sharpening by using stationary wavelet transform. In: Signal Processing and Communication Application Conference (SIU), 2016 24th. pp. 153–156. IEEE (2016)
- [258] Rasti, P., Daneshmand, M., Anbarjafari, G.: Statistical approach based iris recognition using local binary pattern. *DYNA-Ingeniería e Industria* 92(1) (2017)
- [259] Rasti, P., Lusi, I., Sahakyan, A., Traumann, A., Bolotnikova, A., Daneshmand, M., Kiefer, R., Aabloo, A., Anbarjafari, G., Demirel, H., et al.: Modified back projection kernel based image super resolution. In: Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on. pp. 161–165. IEEE (2014)
- [260] Rasti, P., Taşmaz, H., Daneshmand, M., Kiefer, R., Ozcinar, C., Anbarjafari, G.: Satellite image enhancement: Systematic approach for denoising and resolution enhancement. *Dyna-Colombia* 91(3), 326–329 (2016)

- [261] Reinhart, C., Breton, P.F.: Experimental validation of autodesk® 3ds max® design 2009 and daysim 3.0. *Leukos* 6(1), 7–35 (2009)
- [262] Remondino, F.: Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote Sensing* 3(6), 1104–1138 (2011)
- [263] Reshetyuk, Y.: Calibration of terrestrial laser scanners callidus 1.1, leica hds 3000 and leica hds 2500. *Survey Review* 38(302), 703–713 (2006)
- [264] Roncella, R., Forlani, G., Fornari, M., Diotri, F.: Landslide monitoring by fixed-base terrestrial stereo-photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(5), 297 (2014)
- [265] Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J.D., Breitkopf, U., Jung, J.: Results of the isprs benchmark on urban object detection and 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, 256–271 (2014)
- [266] Rusinkiewicz, S., Levoy, M.: Efficient Variants of the ICP Algorithm. In: *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling*. pp. 145–152. IEEE (2001)
- [267] Rusu, R.B., Blodow, N., Marton, Z., Soos, A., Beetz, M.: Towards 3d object maps for autonomous household robots. In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. pp. 3191–3198. IEEE (2007)
- [268] Sánchez-García, E., Balaguer-Beser, A., Pardo-Pascual, J.: C-pro: A coastal projector monitoring system using terrestrial photogrammetry with a geometric horizon constraint. *ISPRS Journal of Photogrammetry and Remote Sensing* 128, 255–273 (2017)
- [269] Scarelli, F.M., Sistilli, F., Fabbri, S., Cantelli, L., Barboza, E.G., Gabbianelli, G.: Seasonal dune and beach monitoring using photogrammetry from uav surveys to apply in the iczm on the ravenna coas (emilia-romagna, italy). *Remote Sensing Applications: Society and Environment* (2017)
- [270] Scheuring, M., Rezk-Salama, C., Barfuffl, H., Schneider, A., Greiner, G.: Augmented Reality based on Fast Deformable 2D-3D Registration for Image-Guided Surgery. In: *Medical Imaging*. pp. 436–445. International Society for Optics and Photonics (2002)

- [271] Schmidt, J., Eisele, L.: Stroboscopic illumination and dark rearing block the sharpening of the regenerated retinotectal map in goldfish. *Neuroscience* 14(2), 535–546 (1985)
- [272] Schmidt, T., Tyson, J., Galanulis, K.: Full-field dynamic displacement and strain measurement using advanced 3d image correlation photogrammetry: part 1. *Experimental Techniques* 27(3), 47–50 (2003)
- [273] Schneider, D.: Calibration of a rieg1 lms-z420i based on a multi-station adjustment and a geometric model with additional parameters. *Int. Arch. Photogramm. Remote Sens. Spat. Inf* 38, 177–182 (2009)
- [274] Sefercik, U.G., Glennie, C., Singhania, A., Hauser, D.: Area-based quality control of airborne laser scanning 3d models for different land classes using terrestrial laser scanning: sample survey in houston, usa. *International Journal of Remote Sensing* 36(23), 5916–5934 (2015)
- [275] Sengupta, S., Chaudhuri, P.: Virtual garment simulation. In: 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG) (2013)
- [276] Shahinpoor, M., Bar-Cohen, Y., Simpson, J., Smith, J.: Ionic polymer-metal composites (ipmcs) as biomimetic sensors, actuators and artificial muscles-a review. *Smart materials and structures* 7(6), R15 (1998)
- [277] Sheffer, A., Praun, E., Rose, K.: Mesh parameterization methods and their applications. *Foundations and Trends® in Computer Graphics and Vision* 2(2), 105–171 (2006)
- [278] Shen, J.K., Matuszewski, B.J., Shark, L., Moore, C.J.: Deformable Registration Using Spring Mass System with Cross-Section Correction. In: International Conference on Medical Information Visualisation-BioMedical Visualisation (MediVis). pp. 9–14. IEEE (2007)
- [279] Shi, B.Q., Liang, J., Liu, Q.: Adaptive simplification of point cloud using k-means clustering. *Computer-Aided Design* 43(8), 910–922 (2011)
- [280] Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 22(8), 888–905 (2000)
- [281] Sim, R., Roy, N.: Global a-optimal robot exploration in slam. In: Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on. pp. 661–666. IEEE (2005)

- [282] Singh, S.P., Jain, K., Mandla, V.R.: A new approach towards image based virtual 3d city modeling by using close range photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2(5), 329 (2014)
- [283] Stamos, I., Allen, P.: 3-d model construction using range and image data. In: *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on.* vol. 1, pp. 531–536. IEEE (2000)
- [284] Sturm, J., Bylow, E., Kahl, F., Cremers, D.: Copyme3d: Scanning and printing persons in 3d. In: *German Conference on Pattern Recognition.* pp. 405–414. Springer (2013)
- [285] Stylianou, A., O’Sullivan, J.D., Abrams, A., Pless, R.: Images don’t forget: Online photogrammetry to find lost graves. In: *Applied Imagery Pattern Recognition Workshop (AIPR), 2014 IEEE.* pp. 1–8. IEEE (2014)
- [286] Sun, Y., Sun, H., Yan, L., Fan, S., Chen, R.: Rba: Reduced bundle adjustment for oblique aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 121, 128–142 (2016)
- [287] Sun, Y., Zhao, L., Huang, S., Yan, L., Dissanayake, G.: L2-sift: Sift feature extraction and matching for large images in large-scale aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 91, 1–16 (2014)
- [288] Takaaki Shiratori, JÃ©rÃ©me Berclaz, M.H.C.S.T.L.Y.M., Shiller, S.: Efficient large-scale point cloud registration using loop closures. In: *3D Vision (3DV), 2015 International Conference on.* pp. 232–240. IEEE (2015)
- [289] Takasaki, H.: Moiré topography. *Applied optics* 9(6), 1467–1472 (1970)
- [290] Tamura, Y., Matsui, M., Pagnini, L.C., Ishibashi, R., Yoshida, A.: Measurement of wind-induced response of buildings using rtk-gps. *Journal of Wind Engineering and Industrial Aerodynamics* 90(12), 1783–1793 (2002)
- [291] Themistocleous, K., Ioannides, M., Agapiou, A., Hadjimitsis, D.: The methodology of documenting cultural heritage sites using photogrammetry, uav, and 3d printing techniques: the case study of asinou church in cyprus. *RSCy2015* pp. 953510–1 (2015)
- [292] Thielemans, K., Tsoumpas, C., Mustafovic, S., Beisel, T., Aguiar, P., Dikaios, N., Jacobson, M.W.: STIR: Software for Tomographic Image Reconstruction Release 2. *Physics in Medicine and Biology* 57(4), 867 (2012)

- [293] Thorne, J.B., Bobbitt, D.R.: Comparison of beer's law and thermal lens techniques for absorption measurements under conditions of high scattering backgrounds. *Applied spectroscopy* 47(3), 360–365 (1993)
- [294] Tong, J., Zhou, J., Liu, L., Pan, Z., Yan, H.: Scanning 3d full human bodies using kinects. *IEEE transactions on visualization and computer graphics* 18(4), 643–650 (2012)
- [295] Tong Liu, Xiaowei Zhang, Z.W., Yuan, Z.: A robust fusion method for rgb-d slam. In: *Chinese Automation Congress (CAC)*. pp. 474–481. IEEE (2013)
- [296] Toupin, R.A.: Saint-venant's principle. *Archive for Rational Mechanics and Analysis* 18(2), 83–96 (1965)
- [297] Toutin, T., Beaudoin, M.: Real-time extraction of planimetric and altimetric features from digital stereo spot data using a digital video plotter. *Photogrammetric engineering and remote sensing* 61(1), 63–68 (1995)
- [298] Tovée, M.J., Benson, P.J., Emery, J.L., Mason, S.M., Cohen-Tovée, E.M.: Measurement of body size and shape perception in eating-disordered and control observers using body-shape software. *British Journal of Psychology* 94(4), 501–516 (2003)
- [299] Traumann, A., Daneshmand, M., Escalera, S., Anbarjafari, G.: Accurate 3d measurement using optical depth information. *Electronics Letters* 51(18), 1420–1422 (2015)
- [300] Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment-a modern synthesis. In: *International workshop on vision algorithms*. pp. 298–372. Springer (1999)
- [301] Tsukada, K., Sekizuka, E., Oshio, C., Minamitani, H.: Direct measurement of erythrocyte deformability in diabetes mellitus with a transparent microchannel capillary model and high-speed video camera system. *Microvascular research* 61(3), 231–239 (2001)
- [302] Turner, D., Lucieer, A., Watson, C.: An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (uav) imagery, based on structure from motion (sfm) point clouds. *Remote Sensing* 4(5), 1392–1410 (2012)
- [303] Udin, W., Ahmad, A.: Assessment of photogrammetric mapping accuracy based on variation flying altitude using unmanned aerial vehicle. In: *IOP*

Conference Series: Earth and Environmental Science. vol. 18, p. 012027. IOP Publishing (2014)

- [304] Vaida, C., Plitea, N., Lese, D., Pisla, D.L.: A Parallel Reconfigurable Robot with Six Degrees of Freedom. In: Applied Mechanics and Materials. vol. 162, pp. 204–213. Trans Tech Publ (2012)
- [305] Valgma, L., Daneshmand, M., Anbarjafari, G.: Iterative closest point based 3d object reconstruction using rgb-d acquisition devices. In: Signal Processing and Communication Application Conference (SIU), 2016 24th. pp. 457–460. IEEE (2016)
- [306] Vauhkonen, J.: Reconstruction, quantification, and visualization of forest canopy based on 3d triangulations of airborne laser scanning point data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2(3), 255 (2015)
- [307] Verhoeven, G.: Taking computer vision aloft—archaeological three-dimensional reconstructions from aerial photographs with photoscan. Archaeological Prospection 18(1), 67–73 (2011)
- [308] Verma, V., Kumar, R., Hsu, S.: 3d building detection and modeling from aerial lidar data. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. vol. 2, pp. 2213–2220. IEEE (2006)
- [309] Volino, P., Magnenat-Thalmann, N.: Virtual clothing: Theory and practice. 2000 (2012)
- [310] Wang, C.C.: Parameterization and Parametric Design of Mannequins. Computer-Aided Design 37(1), 83–98 (2005)
- [311] Wang, C.C., Wang, Y., Yuen, M.M.: Feature based 3d garment design through 2d sketches. Computer-Aided Design 35(7), 659–672 (2003)
- [312] Wang, J., Zhang, X., Chen, H., Ding, S.: Relative pose measurement of satellite and rocket based on photogrammetry. In: Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. pp. 1117–1122. IEEE (2017)
- [313] Wang, J., Lindenbergh, R., Menenti, M.: Sigvox—a 3d feature matching algorithm for automatic street object recognition in mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing 128, 111–129 (2017)

- [314] Wang, Y., Zhang, Q., Zhou, Y.: Dense 3d mapping for indoor environment based on kinect-style depth cameras. In: Robot Intelligence Technology and Applications 3, pp. 317–330. Springer (2015)
- [315] Wei, W., Luo, X., Li, Z.: Layer-based Mannequin Reconstruction and Parameterization from 3D-Range Data. In: Advances in Geometric Modeling and Processing, pp. 498–504. Springer (2008)
- [316] Weise, T., Leibe, B., Van Gool, L.: Fast 3d scanning with automatic motion compensation. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. pp. 1–8. IEEE (2007)
- [317] Westoby, M., Brasington, J., Glasser, N., Hambrey, M., Reynolds, J.: 'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* 179, 300–314 (2012)
- [318] Whelan, T., Kaess, M., Johannsson, H., Fallon, M., Leonard, J.J., McDonald, J.: Real-time large-scale dense rgb-d slam with volumetric fusion. *The International Journal of Robotics Research* 34(4-5), 598–626 (2015)
- [319] Wink, A.M., Roerdink, J.B.: Denoising functional mr images: a comparison of wavelet denoising and gaussian smoothing. *IEEE transactions on medical imaging* 23(3), 374–387 (2004)
- [320] Wolk, R.M.: Utilizing google earth and google sketchup to visualize wind farms. In: Technology and Society, 2008. ISTAS 2008. IEEE International Symposium on. pp. 1–8. IEEE (2008)
- [321] Wu, C.: Visualsfm: A visual structure from motion system. <http://ccwu.me/vsfm/doc.html> (2011), accessed: 2018-01-15
- [322] Wu, J., Cui, Z., Sheng, V.S., Zhao, P., Su, D., Gong, S.: A comparative study of sift and its variants. *Measurement Science Review* 13(3), 122–131 (2013)
- [323] Xia, P., Zhu, X.j., Fei, Y.q.: Mechanical Design and Locomotion Control of a Homogenous Lattice Modular Self-Reconfigurable Robot. *Journal of Zhejiang University Science A* 7, 368–373 (2006)
- [324] Xiao, J., Owens, A., Torralba, A.: Sun3d: A database of big spaces reconstructed using sfm and object labels. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1625–1632 (2013)
- [325] Yan, J., Rui, W.: Parametric Design of Mannequins for Virtual Try-on System. In: 7th International Conference on Computer-Aided Industrial Design and Conceptual Design (CAIDCD). pp. 1–6. IEEE (2006)

- [326] Yang, B., Fang, L., Li, J.: Semi-automated extraction and delineation of 3d roads of street scene from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 79, 80–93 (2013)
- [327] Yang, B., Liu, Y., Dong, Z., Liang, F., Li, B., Peng, X.: 3d local feature bkd to extract road information from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 130, 329–343 (2017)
- [328] Yang, L., Zhang, L., Dong, H., Alelaiwi, A., El Saddik, A.: Evaluating and improving the depth accuracy of kinect for windows v2. *IEEE Sensors Journal* 15(8), 4275–4285 (2015)
- [329] Yastikli, N.: Documentation of cultural heritage using digital photogrammetry and laser scanning. *Journal of Cultural Heritage* 8(4), 423–427 (2007)
- [330] Yoshizawa, S., Belyaev, A., Seidel, H.P.: Smoothing by example: Mesh denoising by averaging with similarity-based weights. In: *Shape Modeling and Applications, 2006. SMI 2006. IEEE International Conference on*. pp. 9–9. IEEE (2006)
- [331] Yu, Q., Shang, Y.: *Videometrics: principles and researches*. Science Process Bei-jing, China pp. 71–71 (2009)
- [332] Yu, X., Zhang, T.: Application of terrestrial 3d laser scanning technology in spatial information acquisition of urban buildings. In: *Image, Vision and Computing (ICIVC), 2017 2nd International Conference on*. pp. 1107–1111. IEEE (2017)
- [333] Yuan, S., Li, M.: Fault diagnosis using binary tree and sphere-structured support vector machines. *Journal of mechanical science and technology* 26(5), 1431–1438 (2012)
- [334] Zai, D., Chen, Y., Li, J., Yu, Y., Wang, C., Nie, H.: Inventory of 3d street lighting poles using mobile laser scanning point clouds. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. pp. 573–576. IEEE (2015)
- [335] Zai, D., Guo, Y., Li, J., Luo, H., Lin, Y., Sun, Y., Huang, P., Wang, C.: 3d road surface extraction from mobile laser scanning point clouds. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. pp. 1595–1598. IEEE (2016)

- [336] Zhang, J., Chesi, G.: Multiview stereo object reconstruction with a one-line search method. *Journal of Electronic Imaging* 22(2), 023019–023019 (2013)
- [337] Zhang, J., Wang, X.: Digital photogrammetry of chinese early aerial photo and application in morphotectonics mapping of tanlu active fault zone. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*. pp. 2924–2926. IEEE (2013)
- [338] Zhang, L., Liu, L., Gotsman, C., Huang, H.: Mesh reconstruction by meshless denoising and parameterization. *Computers & Graphics* 34(3), 198–208 (2010)
- [339] Zhang, M., Lu, Z.: Hybrid Elastic Registration Using Constrained Free-Form Deformation. In: *International Conference on Medical Image Analysis and Clinical Applications (MIACA)*. pp. 75–78. IEEE (2010)
- [340] Zhang, Q., An, P., Zhang, Y., Shen, L., Zhang, Z.: Efficient depth map compression for view rendering in 3d video. *The Imaging Science Journal* 61(4), 385–395 (2013)
- [341] Zhang, Y., Gibson, G.M., Hay, R., Bowman, R.W., Padgett, M.J., Edgar, M.P.: A Fast 3D Reconstruction System with a Low-Cost Camera Accessory. *Scientific Reports* 5 (2015)
- [342] Zhang, Z., Cheng, M., Chen, X., Zhou, M., Chen, Y., Li, J., Nie, H.: Turning mobile laser scanning points into 2d/3d on-road object models: Current status. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International*. pp. 3524–3527. IEEE (2015)
- [343] Zhou, Y., Yu, Y., Lu, G., Du, S.: Super-segments based classification of 3d urban street scenes. *International Journal of Advanced Robotic Systems* 9(6), 248 (2012)
- [344] Zhou, Z., Shu, B., Zhuo, S., Deng, X., Tan, P., Lin, S.: Image-based clothes animation for virtual fitting. In: *SIGGRAPH Asia 2012 Technical Briefs*. p. 33. ACM (2012)
- [345] Zlatanova, S., Rahman, A.A., Shi, W.: Topological models and frameworks for 3d spatial objects. *Computers & Geosciences* 30(4), 419–428 (2004)
- [346] Zolotukhin, A., Safonov, I., Kryzhanovskii, K.: 3d reconstruction for a scanning electron microscope. *Pattern recognition and image analysis* 23(1), 168–174 (2013)

ACKNOWLEDGMENTS

My greatest gratitude goes out to all the favors and supports I received from my marvelous supervisor, Prof. Gholamreza Anbarjafari (Shahab), working under whose guidance has always been fantastic for me, as well as the other awesome supervisor of mine, Prof. Alvo Aabloo.

With a special mention to the iCV Research Group, which has ever been an excellent place to work.

I am also grateful to my colleagues, including, in an alphabetical order, remarkable individuals Ahmed Helmi, Andre Litvin, Egils Avots, Fatemeh Noroozi, Hasan Sait Arslan, Iris Lüsi, Jelena Gorbova and Rain Eric Haamer, to name a few.

I am greatly thankful to my parents and my sister, who provided me with significant emotional backing along the way.

Last but never the least, my thesis project was partly funded by the Estonian Research Council Grant PUT (PUT638). I also acknowledge partial financial support from Fits.Me through University of Tartu R&D contract LLTTI16056.

KOKKUVÕTE (SUMMARY IN ESTONIAN)

VIRTUAALSE PROOVIKABIINI 3D KEHAKUJUDE JA ROBOTI JUHTIMISALGORITMIDE UURIMINE

Virtuaalne riide proovimine on üks põhilistest teenustest, mille pakkumine võib suurendada rõivapoodide edukust, sest tänu sellele lahendusele väheneb füüsilise töö vajadus proovimise faasis ning riide proovimine muutub kasutaja jaoks mugavamaks. Samas pole enamikel varem välja pakutud masinnägemise ja graafika meetoditel õnnestunud inimkeha realistlik modelleerimine, eriti terve keha 3D modelleerimine, mis vajab suurt kogust andmeid ja palju arvutuslikku ressursi. Varasemad katsed on ebaõnnestunud põhiliselt seetõttu, et ei ole suudetud korralikult arvesse võtta samaaegseid muutusi keha pinnal. Lisaks pole varasemad meetodid enamasti suutnud kujutiste liikumisi realistlikult reaalses visualiseerida. Käesolev projekt kavatses kõrvaldada eelmainitud puudused nii, et rahuldada virtuaalse proovikabiini vajadusi. Välja pakutud meetod seisneb nii kasutaja keha kui ka riide skaneerimises, analüüsimises, modelleerimises, mõõtmete arvutamises, orientiiride paigutamises, mannekeenidelt võetud 3D visuaalsete andmete segmenteerimises ning riide mudeli paigutamises ja visualiseerimises kasutaja kehal. Selle projekti käigus koguti visuaalseid andmeid kasutades 3D laserskanerit ja Kinecti optilist kaamerat ning koostati nendest andmebaas. Neid andmeid kasutati välja töötatud algoritmide testimiseks, mis peamiselt tegelevad riide realistliku visuaalse kujutamise ja suuruse pakkumise süsteemi täiendamisega virtuaalse proovikabiini kontekstis.

APPENDICES

Morteza Daneshmand

Curriculum Vitae

Estonian ID: 38804210125
iCV Research Group
Institute of Technology, University of Tartu
Nooruse 1-139, 50411 Tartu, Estonia
✉ mortezad@ut.ee
📄 icv.tuit.ut.ee



Education

- 2014–2018 **Doctor of Philosophy, Engineering and Technology.**
University of Tartu, Tartu, Estonia
Specialty: Physical and Information Technology
Supervisors: Alvo Aabloo and Gholamreza Anbarjafari
Thesis: Realistic 3D Virtual Fitting Room
GPA: 4.53/5
- 2011–2014 **Master of Science, Technological Interdisciplinary.**
University of Tehran, Tehran, Iran
Specialty: Mechatronics Engineering
Supervisors: Mehdi Tale Masouleh and Mohammad Bagher Menhaj
Thesis: Kinematic Sensitivity and Workspace Optimization of Planar Parallel Mechanisms
GPA: 19.20/20 (Ranked second out of 18)
- 2006–2011 **Bachelor of Science, Electrical Engineering.**
Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
Specialty: Control
Supervisor: Mohammad Bagher Menhaj
Thesis: Implementation of Distributed Formation Control of Multi-Agent Systems on e-puck Mobile Robots
GPA: 14.06/20

Work and Research Experiences

- 2016–Present **Researcher in 3D Modelling, iCV Research Group.**
Retexturing of images and garment 3D model creation
Funded by Rakuten - Fits.me (2 years R&D Contract)
Total Budget: 105,000 EUR
- 2016–Present **Consultant in 3D Modelling, iCV Research Group.**
Head Scan Using Mobile Phones (3 months R&D Contract)
Total Budget: 16,000 EUR
- 2016 **Researcher in 3D Modelling, iCV Research Group.**
HTC VIVE VR (3 months R&D Contract)
Total Budget: 10,000 EUR
- 2016 **Researcher in Machine Learning, iCV Research Group.**
Age and Gender estimation
Funded by Super Appli (3 months R&D Contract)
Total Budget: 18,000 EUR

- 2015–Present **Researcher in Machine Learning**, *iCV Research Group*.
Enhancing Human-Computer Interaction using Multimodal Emotion Recognition based on Gesture, Face, and Speech (4 years R&D Contract)
Total Budget: 112,800 EUR
- 2015–2016 **Project Manager**, *iCV Research Group*.
Automatic drawing (7 months R&D Contract)
Total Budget: 37,000.00 EUR
- 2014–2015 **Specialist in 3D Modelling**, *iCV Research Group*.
Self-Calibrating Test Prototype of Shape-Changing Mannequin
Funded by Software Technology and Applications Competence Center (STACC) (1 year R&D Contract)
Total Budget: 56,000 EUR

Research Interests

- 3D reconstruction and modeling
- Optimization
- Visualization and computer graphics
- Image processing and identity recognition
- Control, robotics and mechatronics
- Design and kinetostatic of parallel mechanisms

Computer skills

Languages, libraries and software **MATLAB, C/C++, Maple, JavaScript, HTML, CSS, OpenCV, Festo, Delta and Mitsubishi PLCs, Blender, Microsoft Office, Latex, Beamer, SPSS, Eviews.**

OS **Microsoft Windows, Linux.**

Additional Career Information and R&D-Related Works

- 2015–2018 **Chair of the University of Tartu IEEE Student Branch.**
- 2014–Present **Member of the University of Tartu RoboCup Team (Philosopher Estonia).**
- 2012–Present **Student member of the IEEE.**
- 2014 **PLC program, SHEE (Self-deployable Habitat for Extreme Environments) [FP7].**
- 2012–2014 **Student member of the Robotics Society of Iran (RSI).**
- 2013–2014 **Member of the Iranian Society of Instrumentation and Control Engineers.**

Teaching Assistance

- Autumn 2017 **Data Analysis and Computational Methods with MATLAB, University of Tartu.**
- Spring 2017 **Advances in Automation Systems and Engineering, University of Tartu.**
- Autumn 2015 **Robotics Technology, University of Tartu.**

Autumn 2012 **Advanced Robotics, University of Tehran.**
Spring 2013
Autumn 2012 **Physics-I (Mechanics), Sharif University of Technology, Tehran, Iran.**
Autumn 2012 **Mechatronics II, University of Tehran.**
Autumn 2012 **Static and Strength of Materials, University of Tehran.**
Spring 2012 **Advanced Automatic Control, University of Tehran.**
Autumn 2011 **Signals and Systems Analysis, University of Tehran.**
Autumn 2011 **Electronics-I, University of Tehran.**

TOEFL Scores (Administration: 21/12/2013)

Reading, 30.

Listening, 25.

Speaking, 22.

Writing, 29.

Total, 106.

GRE General Scores (Administration: 11/09/2013)

Verbal Reasoning, Scaled Score: 156, % Below: 71.

Quantitative Reasoning, Scaled Score: 169, % Below: 97.

Analytical Writing, Scaled Score: 4.0, % Below: 56.

References

Gholamreza Anbarjafari, University of Tartu, shb@ut.ee, (+372) 737 4855.

Alvo Aabloo, University of Tartu, alvo.aabloo@ut.ee, (+372) 737 5534.

Mehdi Tale Masouleh, University of Tehran, m.t.masouleh@ut.ac.ir, (+98) 21 6111 8413.

Mohammad Bagher Menhaj, Amirkabir University of Technology, menhaj@aut.ac.ir, (+98) 21 6454 3347.

Journal Articles

- E. Avots, M. Daneshmand, A. Traumann, S. Escalera, and G. Anbarjafari. " Automatic garment retexturing based on infrared information ". In: *Computers & Graphics* 59 (2016), pp. 28–38.
- I. Beheshti, N. Maikusa, M. Daneshmand, H. Matsuda, H. Demirel, and G. Anbarjafari. " Classification of alzheimer's disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks ". In: *Journal of Alzheimer's Disease* 60.1 (2017), pp. 295–304.
- M. Daneshmand, M. T. Masouleh, M. Saadatzi, C. Ozcinar, and G. Anbarjafari. " A robust proportion-preserving composite objective function for scale-invariant multi-objective optimization ". In: *Scientia Iranica. Transaction B, Mechanical Engineering* 24.6 (2017), pp. 2977–2991.
- M. Daneshmand, A. Aabloo, and G. Anbarjafari. " Size-dictionary interpolation for robot's adjustment ". In: *Frontiers in bioengineering and biotechnology* 3 (2015), p. 63.

- M. Daneshmand, A. Aabloo, C. Ozcinar, and G. Anbarjafari. " Real-time, automatic shape-changing robot adjustment and gender classification ". In: *Signal, Image and Video Processing* 10.4 (2016), pp. 753–760.
- M. Daneshmand, A. Abels, and G. Anbarjafari. " Real-time, automatic digi-tailor mannequin robot adjustment based on human body classification through supervised learning ". In: *International Journal of Advanced Robotic Systems* 14.3 (2017), p. 1729881417707169.
- M. Daneshmand, O. Bilici, A. Bolotnikova, and G. Anbarjafari. " Medical robots with potential applications in participatory and opportunistic remote sensing: a review ". In: *Robotics and Autonomous Systems* 95 (2017), pp. 160–180.
- M. Daneshmand, A. Helmi, E. Avots, F. Noroozi, F. Alisinanoglu, H. S. Arslan, J. Gorbova, R. E. Haamer, C. Ozcinar, and G. Anbarjafari. " 3d scanning: a comprehensive survey ". In: *arXiv preprint arXiv:1801.08863* (2018).
- M. Daneshmand, M. T. Masouleh, and G. Anbarjafari. " Kinematic sensitivity analysis of parallel mechanisms by considering the effect of uncertainties in passive joints ". In: *Modares Mechanical Engineering* 15.10 (2016), pp. 78–88.
- M. Daneshmand, M. H. Saadatzi, M. H. F. Kaloorazi, M. T. Masouleh, and G. Anbarjafari. " Optimal design of a spherical parallel manipulator based on kinetostatic performance using evolutionary techniques ". In: *Journal of Mechanical Science and Technology* 30.3 (2016), pp. 1323–1331.
- M. Daneshmand, M. Tale Masouleh, and M. H. Saadatzi. " Optimization of the kinematic sensitivity and the greatest continuous circle in the constant-orientation workspace of planar parallel mechanisms ". In: *International Journal of Robotics, Theory and Applications* 4.1 (2016), pp. 12–21.
- M. Daneshmand, A. Traumann, and G. Anbarjafari. " 3d size-estimation based on the geodesic distance measured by photogrammetric scanning device ". In: (2015).
- P. Rasti, M. Daneshmand, and G. Anbarjafari. " Statistical approach based iris recognition using local binary pattern ". In: *Dyna* 92.1 (2017), pp. 76–81.
- P. Rasti, H. Taşmaz, M. Daneshmand, R. Kiefer, C. Ozcinar, and G. Anbarjafari. " Satellite image enhancement: systematic approach for denoising and resolution enhancement ". In: *Dyna-Colombia* 91.3 (2016), pp. 326–329.
- A. Traumann, M. Daneshmand, S. Escalera, and G. Anbarjafari. " Accurate 3d measurement using optical depth information ". In: *Electronics Letters* 51.18 (2015), pp. 1420–1422.

Book Chapters

- G. Anbarjafari, P. Rasti, M. Daneshmand, and C. Ozcinar. " Resolution enhancement based image compression technique using singular value decomposition and wavelet transforms ". In: *Wavelet Transform and Some of Its Real-World Applications*. Intech, 2015, pp. 35–52.

Conference Papers

- A. Bolotnikova, P. Rasti, A. Traumann, I. Lusi, M. Daneshmand, F. Noroozi, K. Samuel, S. Sarkar, and G. Anbarjafari. " Block based image compression technique using rank reduction and wavelet difference reduction ". In: *Seventh International Conference on Graphic and Image Processing (ICGIP 2015)*. Vol. 9817. International Society for Optics and Photonics. 2015, p. 981702.
- M. Daneshmand, M. T. Masouleh, and G. Anbarjafari. " Kinematic sensitivity analysis of a 3-dof decoupled translational parallel mechanism with uncertainties in the passive joints ".

- In: *Ubiquitous Robots and Ambient Intelligence (URAI), 2015 12th International Conference on*. IEEE. 2015, pp. 85–90.
- M. Daneshmand, M. H. Saadatzi, and M. T. Masouleh. “ Kinematic sensitivity and workspace optimization of planar parallel mechanisms using evolutionary techniques ”. In: *Robotics and Mechatronics (ICRoM), 2013 First RSI/ISM International Conference on*. IEEE. 2013, pp. 384–389.
 - M. F. Kaloorazi, S. Esfahani, M. T. Masouleh, and M. Daneshmand. “ Dimensional synthesis of planar cable-driven parallel robots via interval analysis ”. In: *CCToMM Symposium, IFToMM, Montreal, Quebec, Canada*. 2013.
 - L. Laur, M. Daneshmand, M. Agoyi, and G. Anbarjafari. “ Robust grayscale watermarking technique based on face detection ”. In: *Signal Processing and Communications Applications Conference (SIU), 2015 23th*. IEEE. 2015, pp. 471–475.
 - K. Pjatkin, M. Daneshmand, P. Rasti, and G. Anbarjafari. “ Probability distribution function based iris recognition boosted by the mean rule ”. In: *Intelligent Computing and Internet of Things (ICIT), 2014 International Conference on*. IEEE. 2015, pp. 47–50.
 - P. Rasti, M. Daneshmand, F. Alisinanoglu, C. Ozcinar, and G. Anbarjafari. “ Medical image illumination enhancement and sharpening by using stationary wavelet transform ”. In: *Signal Processing and Communication Application Conference (SIU), 2016 24th*. IEEE. 2016, pp. 153–156.
 - P. Rasti, I. Lusi, A. Sahakyan, A. Traumann, A. Bolotnikova, M. Daneshmand, R. Kiefer, A. Aabloo, G. Anbarjafari, H. Demirel, et al. “ Modified back projection kernel based image super resolution ”. In: *Artificial Intelligence, Modelling and Simulation (AIMS), 2014 2nd International Conference on*. IEEE. 2014, pp. 161–165.
 - L. Valgma, M. Daneshmand, and G. Anbarjafari. “ Iterative closest point based 3d object reconstruction using rgb-d acquisition devices ”. In: *Signal Processing and Communication Application Conference (SIU), 2016 24th*. IEEE. 2016, pp. 457–460.

ELULOOKIRJELDUS

Isikuandmed

Nimi	Morteza Daneshmand
Sünniaeg ja -koht	21. aprill 1988 Qom, Iraan
Kodakondsus	iraanlane
Perekonnaseis	vallaline
Keelteoskus	pärsia, inglise
Aadress	Nooruse 1, Tartu 50411 Tartumaa Eesti
Kontaktandmed	+372 56 372 718 md@icv.tuit.ut.ee

Haridustee

2014–	Tartu Ülikool, Tehnika ja tehnoloogia doktorikraad omandamisel
2011–2014	Tehrani Ülikool, MSc Mehhatroonika
2006–2011	Amirkabiri Tehnikaülikool (Tehran Polytechnic), BSc Elektrotehnika
1999–2006	Iranian National Organization for Development of Exceptional Talents (NODET), keskharidus
1994–1999	Jahad Elementary School, põhiharidus

Teenistuskäik

2014–2015	Tarkvara Tehnoloogia Arenduskeskus OÜ, 3D modelleerimise spetsialist
-----------	--

DISSERTATIONES TECHNOLOGIAE UNIVERSITATIS TARTUENSIS

1. **Imre Mäger.** Characterization of cell-penetrating peptides: Assessment of cellular internalization kinetics, mechanisms and bioactivity. Tartu 2011, 132 p.
2. **Taavi Lehto.** Delivery of nucleic acids by cell-penetrating peptides: application in modulation of gene expression. Tartu 2011, 155 p.
3. **Hannes Luidalepp.** Studies on the antibiotic susceptibility of *Escherichia coli*. Tartu 2012, 111 p.
4. **Vahur Zadin.** Modelling the 3D-microbattery. Tartu 2012, 149 p.
5. **Janno Torop.** Carbide-derived carbon-based electromechanical actuators. Tartu 2012, 113 p.
6. **Julia Suhorutšenko.** Cell-penetrating peptides: cytotoxicity, immunogenicity and application for tumor targeting. Tartu 2012, 139 p.
7. **Viktoryia Shyp.** G nucleotide regulation of translational GTPases and the stringent response factor RelA. Tartu 2012, 105 p.
8. **Mardo Kõivomägi.** Studies on the substrate specificity and multisite phosphorylation mechanisms of cyclin-dependent kinase Cdk1 in *Saccharomyces cerevisiae*. Tartu, 2013, 157 p.
9. **Liis Karo-Astover.** Studies on the Semliki Forest virus replicase protein nsP1. Tartu, 2013, 113 p.
10. **Piret Arukuusk.** NickFects–novel cell-penetrating peptides. Design and uptake mechanism. Tartu, 2013, 124 p.
11. **Piret Villo.** Synthesis of acetogenin analogues. Asymmetric transfer hydrogenation coupled with dynamic kinetic resolution of α -amido- β -keto esters. Tartu, 2013, 151 p.
12. **Villu Kasari.** Bacterial toxin-antitoxin systems: transcriptional cross-activation and characterization of a novel *mqsRA* system. Tartu, 2013, 108 p.
13. **Margus Varjak.** Functional analysis of viral and host components of alphavirus replicase complexes. Tartu, 2013, 151 p.
14. **Liane Viru.** Development and analysis of novel alphavirus-based multi-functional gene therapy and expression systems. Tartu, 2013, 113 p.
15. **Kent Langel.** Cell-penetrating peptide mechanism studies: from peptides to cargo delivery. Tartu, 2014, 115 p.
16. **Rauno Temmer.** Electrochemistry and novel applications of chemically synthesized conductive polymer electrodes. Tartu, 2014, 206 p.
17. **Indrek Must.** Ionic and capacitive electroactive laminates with carbonaceous electrodes as sensors and energy harvesters. Tartu, 2014, 133 p.
18. **Veiko Voolaid.** Aquatic environment: primary reservoir, link, or sink of antibiotic resistance? Tartu, 2014, 79 p.
19. **Kristiina Laanemets.** The role of SLAC1 anion channel and its upstream regulators in stomatal opening and closure of *Arabidopsis thaliana*. Tartu, 2015, 115 p.

20. **Kalle Pärn.** Studies on inducible alphavirus-based antitumour strategy mediated by site-specific delivery with activatable cell-penetrating peptides. Tartu, 2015, 139 p.
21. **Anastasia Selyutina.** When biologist meets chemist: a search for HIV-1 inhibitors. Tartu, 2015, 172 p.
22. **Sirle Saul.** Towards understanding the neurovirulence of Semliki Forest virus. Tartu, 2015, 136 p.
23. **Marit Orav.** Study of the initial amplification of the human papillomavirus genome. Tartu, 2015, 132 p.
24. **Tormi Reinson.** Studies on the Genome Replication of Human Papillomaviruses. Tartu, 2016, 110 p.
25. **Mart Ustav Jr.** Molecular Studies of HPV-18 Genome Segregation and Stable Replication. Tartu, 2016, 152 p.
26. **Margit Mutso.** Different Approaches to Counteracting Hepatitis C Virus and Chikungunya Virus Infections. Tartu, 2016, 184 p.
27. **Jelizaveta Geimanen.** Study of the Papillomavirus Genome Replication and Segregation. Tartu, 2016, 168 p.
28. **Mart Toots.** Novel Means to Target Human Papillomavirus Infection. Tartu, 2016, 173 p.
29. **Kadi-Liis Veiman.** Development of cell-penetrating peptides for gene delivery: from transfection in cell cultures to induction of gene expression *in vivo*. Tartu, 2016, 136 p.
30. **Ly Pärnaste.** How, why, what and where: Mechanisms behind CPP/cargo nanocomplexes. Tartu, 2016, 147 p.
31. **Age Utt.** Role of alphavirus replicase in viral RNA synthesis, virus-induced cytotoxicity and recognition of viral infections in host cells. Tartu, 2016, 183 p.
32. **Veiko Vunder.** Modeling and characterization of back-relaxation of ionic electroactive polymer actuators. Tartu, 2016, 154 p.
33. **Piia Kivipõld.** Studies on the Role of Papillomavirus E2 Proteins in Virus DNA Replication. Tartu, 2016, 118 p.
34. **Liina Jakobson.** The roles of abscisic acid, CO₂, and the cuticle in the regulation of plant transpiration. Tartu, 2017, 162 p.
35. **Helen Isok-Paas.** Viral-host interactions in the life cycle of human papillomaviruses. Tartu, 2017, 158 p.
36. **Hanna Hõrak.** Identification of key regulators of stomatal CO₂ signalling via O₃-sensitivity. Tartu, 2017, 160 p.
37. **Jekaterina Jevtuševskaja.** Application of isothermal amplification methods for detection of *Chlamydia trachomatis* directly from biological samples. Tartu, 2017, 96 p.
38. **Ülar Allas.** Ribosome-targeting antibiotics and mechanisms of antibiotic resistance. Tartu, 2017, 152 p.
39. **Anton Paier.** Ribosome Degradation in Living Bacteria. Tartu, 2017, 108 p.
40. **Vallo Varik.** Stringent Response in Bacterial Growth and Survival. Tartu, 2017, 101 p.

41. **Pavel Kudrin.** In search for the inhibitors of *Escherichia coli* stringent response factor RelA. Tartu, 2017, 138 p.
42. **Liisi Henno.** Study of the human papillomavirus genome replication and oligomer generation. Tartu, 2017, 144 p.
43. **Katrin Krõlov.** Nucleic acid amplification from crude clinical samples exemplified by *Chlamydia trachomatis* detection in urine. Tartu, 2018, 118 p.
44. **Eve Sankovski.** Studies on papillomavirus transcription and regulatory protein E2. Tartu, 2018, 113 p.