Team Activity Analysis and Recognition Based on Kinect Depth Map and Optical Imagery Techniques

Vinayak Elangovan **Vinod K Bandaru** **Amir Shirkhodaie**

Center of Excellence for Battlefield Senor Fusion
Dept. of Mechanical & Manufacturing Engineering
Tennessee State University
TN, U.S.A.

ABSTRACT

Kinect cameras produce low-cost depth map video streams applicable for conventional surveillance systems. However, commonly applied image processing techniques are not directly applicable for depth map video processing. Kinect depth map images contain range measurement of objects at expense of having spatial features of objects suppressed. For example, typical objects' attributes such as textures, color tones, intensity, and other characteristic attributes cannot be fully realized by processing depth map imagery. In this paper, we demonstrate application of Kinect depth map and optical imagery for characterization of indoor and outdoor group activities. A Casual-Events State Inference (CESI) technique is proposed for spatiotemporal recognition and reasoning of group activities. CESI uses an ontological scheme for representation of casual distinctiveness of a priori known group activities. By tracking and serializing distinctive atomic group activities, CESI allows discovery of more complex group activities. A Modified Sequential Hidden Markov Model (MS-HMM) is implemented for trail analysis of atomic events representing correlated group activities. CESI reasons about five levels of group activities including: Merging, Planning, Cooperation, Coordination, and Dispersion. In this paper, we present results of capability of CESI approach for characterization of group activities taking place both in indoor and outdoor. Based on spatiotemporal pattern matching of atomic activities representing a known group activities, the CESI is able to discriminate suspicious group activity from normal activities. This paper also presents technical details of imagery techniques implemented for detection, tracking, and characterization of atomic events based on Kinect depth map and optical imagery data sets. Various experimental scenarios in indoors and outdoors (e.g. loading and unloading of objects, human-vehicle interactions etc.,) are carried to demonstrate effectiveness and efficiency of the proposed model for characterization of distinctive group activities.

Keywords: Group Activity Recognition, Sequential Hidden Markov Model, Team Members Interactions (TMI), Spatiotemporally Correlated Group Activities, Casual-Events State Inference (CESI), Kinect Depth Map.

1. INTRODUCTION

Video surveillance and Human behavior analysis is one of the ongoing active researches in image processing. Automatic recognition of human behaviors is an ambitious yet a challenging process in achievement of an Intelligent Surveillance System (ISS) for Department of Defense (DOD) and Department of Homeland Security (DHS). The main objective of ISS is to provide an early warning on potential threats while improving situational assessment in real-time. ISS facilitates the formulation and execution of preemptive activities to deter or forestall anticipated adversary courses of action [1-3]. An Intelligent Surveillance System (ISS) is a data/information collection strategy that emphasizes the ability of the system to linger on demand in an area to possibly provide situational assessment [3]. A typical Intelligent Surveillance Systems encompasses a number of activities. Most activities are asynchronous in nature. They may include environment sensing via hard (physical), soft (biological) information processing via receiving and filtering sporadic field briefs and reports. Another vein of activities may include data collection and conditioning, target detection, localization and tracking, event/entity characterization, and local and global-level data/information fusion for creating better situational awareness [3]. At each stage a spectrum of alternative models and techniques exist. Environment sensing deals with different aspects of sensing modalities such as: video signals, acoustic, radar, seismic, infrared and hyper spectral imagery, etc. The main objective of sensing stage is collecting raw data from the environment to be processed in the subsequent stages of the ISS. Data conditioning mainly, deals with noise filtering and data alignment. In target detection stage, surveillance anomalies, events of interest (EOIs) and targets of interest (TOIs) are determined.

Once an event has been detected, it is characterized in the stage of EOIs characterization. Data communication involves propagating messages from agent to agent over the network.

Human Activity recognition is a challenging problem in due to complexity involved in low-level processing, data alignment from different sensor sources, and fusion of soft (e.g., human intelligence) and hard (physical sensors) data, and interpretation of fragmented, yet spatiotemporally correlated and associated information that enhance Situational Awareness (SA) [4-6] . Human activities may involve human-human interactions, human-object interactions, human-environment interactions or multi-human-object interactions or multi-human-multi-object interactions [7]. However, less attention had been focused towards identifying cohesive patterns of Team Member's Interaction (TMI) towards suspicious activities. Significant efforts have been also devoted by previous researchers in understanding, comprehension, and reasoning of the information provided by soft and hard sensors through Intelligent Surveillance Systems to identify human group activity interactions [1-7]. Much work has been devoted on identifying low level activities like group running, group walking, group fighting, group gathering, group approaching, group chasing and etc. On the other hand, identifying human interactions as a team for high level activities have not been well addressed by previous researchers. High level activities may include operational activities involving loading and unloading of objects, exchange of baggage, exchange of vehicles, group dropping baggage in unattended environment, and similar group activities. Vehicles have been used as a primary source of transportation for pursuing many outdoor suspicious activities [8]. In the past, researchers had addressed human interactions with objects like suitcase, box, cups etc. in human activity recognition [7, 10]. In-depth research work had been carried on interpreting human vehicle interaction in field of automotive engineering but less attention had been focused in interaction between human and vehicle for detecting cohesive patterns of suspicious activities.

Most researchers had used image segmentation for extracting the fragmented details from the image. Image segmentation discriminates similarities such as intensity, color, textures, patterns etc. But in real time situations, issues like limited spatial resolution, noise, poor contrast features, overlapping intensities etc makes segmentation a difficult task [10]. To overcome this issue, we had proposed a method of target isolation and tracking of target and events using target zoning and image processing techniques. Silhouette-based posture analysis was proposed in [12] to estimate the posture of the detected person. To identify whether a person is carrying an object in upright-standing posture, dynamic periodic motion analysis and symmetry analysis are applied in [12]. Various motion detected algorithms had developed to detect a person or a part of a person. [9] had proposed a technique using image morphology operations for detecting people's head for people count. [9] made an assumption of normal width of a person and divides persons touching each other as soon as they detect the head. In our work, we employed image processing morphological operations and motion tracking algorithms to detect human head detection for counting number of people. In this paper, a robust approach had been proposed for detecting the TMI observations using Casual Event State Inference technique and TMI activities using a modified sequential Hidden Markov Model (HMM). This paper is organized as follows: TMI Framework, Casual Event State Inference (CESI), Modified Sequential – Hidden Markov Model (MS-HMM), Results & Discussion, Conclusion followed by Acknowledgements and References.

2. TMI FRAMEWORK

TMI activity prediction involves various processing in conjunction with spatiotemporal relations of events. The abstraction of Team Members Interaction framework is shown in Figure 1. Imagery source addressed in block-1 are optical images and kinect depth map images. The captured images from camera feed are preprocessed in block-2 to suppress undesired distortion in the image. In block-3, target of interest are detected, isolated and tagged by extracting relevant features.

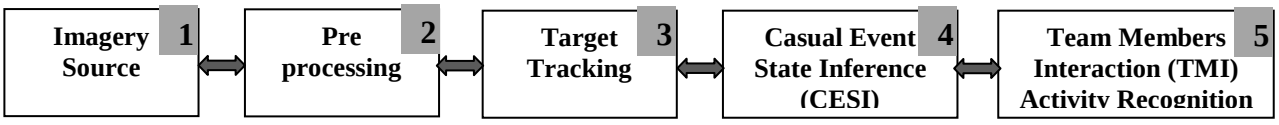


Figure 1: Team Members Activity Recognition Framework

In block-4, the information of detected targets is fed to Casual Event State Inference (CESI) for further processing. CESI categorizes interaction into Human-Vehicle Interactions or Human-Human Interactions or Human-Object Interactions.

CESI detects and identifies the atomic events of such interactions and converts the detected events to recognizable observations. In block-5, pipeline of observations from CESI are fed to a modified sequential - HMM for TMI activities prediction.

The proposed CESI initially categorizes human group activity into Non-Team Members Interaction and Team Members Interaction activities based on the spatiotemporal relationship between the team members and their targeted operational activity goal as shown in Figure 2.

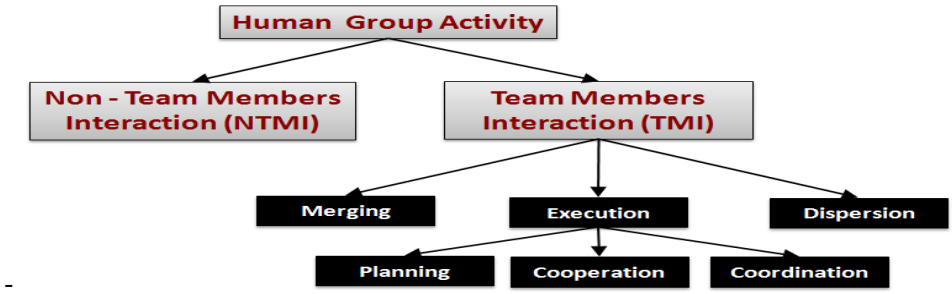


Figure 2: Team Member Interactions Stages

The activities focused under TMI in our work focus towards a targeted operational activity. For example, group of people running or group of people talking in a conference hall are not considered as a team members interaction. Team Members Interaction is categorized into three major activities i.e. Merging, Execution and Dispersion. Merging is considered as grouping or gathering or merging of individuals into a monitored environment. Execution is furthered categorized to Planning, Cooperation or Coordination activities. The final stage of a TMI is Dispersion i.e. team members departing either as individuals or in variable groups.

3. CASUAL EVENT STATE INFERENCE (CESI)

This section discusses in detail the process and roles involved in CESI. The inputs fed to CESI are the target location, target profile information, spatial information, number of targets and type of targets (i.e. vehicle or humans or objects etc). The main goal of CESI is to detect the atomic events from TMI interactions and transpose these events to evidential observations. CESI contains four major processing units as shown in Figure-3 and are discussed in detail below.

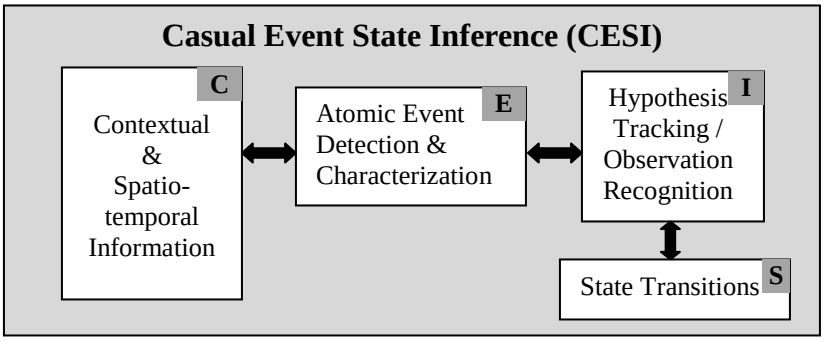
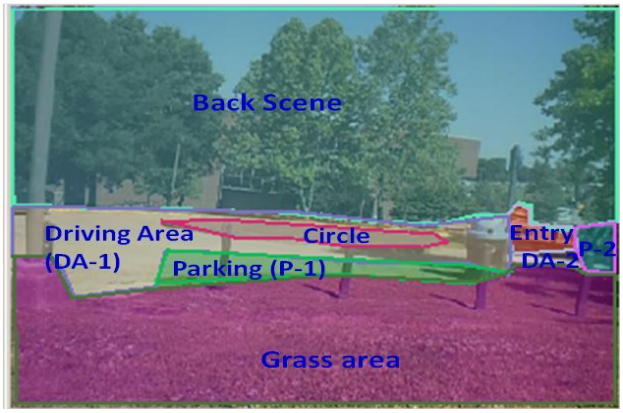


Figure 3: Casual Event State Inference (CESI) Framework

3.1 Contextual Information & Spatio-temporal Information

In order to monitor an environment and generate context-based semantic labels, we divide the monitored area into *n* complete connected zones and label each zone with a specific label in block-C. The target of interest (TOI) in the environment to be monitored can be humans or vehicles or objects that being carried by humans. Let the environment to be monitored be E_Z, then the total number of zones in the environment can be represented as shown in below Figure 4.



The mathematical expressions for environment zoning and target of interest is given below:

$$\boldsymbol{Environment\ Zones\ E\_Z} = \{\boldsymbol{Z_1}, \boldsymbol{Z_2}, \ldots \boldsymbol{Z_n}\}$$

$$\boldsymbol{Target\ of\ Interest\ TOI} = \{\boldsymbol{T\_H}, \boldsymbol{T\_V}, \boldsymbol{T\_O}\}$$

$$\boldsymbol{Human\ Targets\ T\_H} = \{\boldsymbol{T\_H_1}, \boldsymbol{T\_H_2}, \ldots \boldsymbol{T\_H_i}\}$$

$$\boldsymbol{Vehicle\ Targets\ T\_V} = \{\boldsymbol{T\_V_1}, \boldsymbol{T\_V_2}, \ldots \boldsymbol{T\_V_J}\}$$

$$\boldsymbol{Object\ Target\ T\_O} = \{\boldsymbol{T\_O_1}, \boldsymbol{T\_O_2}, \ldots \boldsymbol{T\_O_k}\}$$

Figure 4: Environment Monitoring & Target Detection

where 'n' is the total number of zones in the environment, T_H is the Human targets, T_V is the Vehicle targets, T_O is the Object targets, 'i' is the total number of human targets detected in E_Z, 'j' is the total number of vehicle targets detected in E_Z and 'k' is the total number of object targets detected in E_Z.

The location of every TOI is identified based on the zone, the target is located. Zoning of the environment helps in detecting and tracking a target abstractly in different connected zones through segmenting the background area which is to be monitored and labeling each zone. Zoning environment also helps in tracing out the suspiciousness involved in the activities. For every target, there exists zone information Z_m at any time t such that:

$$\forall \boldsymbol{TOI} \in \{\boldsymbol{T_H}, \boldsymbol{T_V}, \boldsymbol{T_O}\}, \exists\ \boldsymbol{Z_m} \in \boldsymbol{E\_Z}$$

where 'm' is the current zone of target of interest.

The spatiotemporal relationship between vehicle targets and human targets are detected using Allen's Temporal Logic 13 Base Relations namely [7]:

X before Y, X after Y, X meets Y, X met by Y, X overlaps Y, X overlapped by Y, X starts Y, X started by Y, X during Y, X contains Y, X finishes Y, X finished by Y, X equals Y, where symbols X and Y may be considered as vehicle targets or human targets, and observations respectively.

3.2 Atomic Event Detection & Characterization

In block-E, based on the detected targets and the contextual and spatiotemporal information of the activity taking / taken place, the interactions are classified into HVI or HHI or HOI or combinations of such interactions. Three different models are developed for detecting atomic events of HVI, HHI and HOI. The inputs, triggering elements and techniques used in each Interaction classifier are shown in Table-1.

Table-1: Interactions Classifier

Interactions Type	Inputs	Triggering Elements	Techniques	Output
HVI	Targets location, targets size, vehicle type, zone information	Human/s & vehicle in same zone or in very close proximity	Zoning of Vehicle (ZoV) technique, spatio-temporal correlation, ontology mapping	Postures, HVI atomic events
HHI	Target location, target size, zone information	Humans in same zone or neighboring zones	Measurement probing, spatio-temporal correlation, hypothesis matching	Postures, HHI atomic events
HOI	Target location, target size, zone information	Human/s & object are in very close proximity	Template matching, spatio-temporal correlation, hypothesis matching	HOI atomic events

3.3 Human-Vehicle Interaction (HVI):

Human-Vehicle Interactions are recognized using a vehicle zoning (ZoV) technique for atomic interactions detection and whereabouts of human(s) around the vehicle. By zoning surrounding of the vehicle and detecting human activities at such zones, one can ascertain type of potential/possible interactions that the human is involved with some degree of certainty. The details of HVI can be found in our previous paper [8]. By focusing on metaphysics of HVI, we developed an ontology system containing 100+ ontologies and 100+ hypotheses that links together what types of HVI are possible and what relations these events bear to one another to ensemble a situational awareness [8]. An example of HVI events detection and recognition is shown in Figure 5.

The major steps followed in HVI image processing:

1. Vehicle location detection using environment zoning
2. Image differencing for vehicle isolation
3. Obtain vehicle model using binarization, blob processing and cavity filling
4. Perform Zoning of Vehicle (ZoV) technique to localize human around the vehicle
5. Segment human from background employing vehicle model as a template and do target isolation
6. Perform human posture tracking / classification and detect HVI atomic events
7. Generate HVI semantic annotations

Figure 5: HVI Event Identification (Hood Open & Walking)

3.4 Human-Human Interaction (HHI) and Human-Object Interaction:

Human-Human Interactions are recognized by isolating the detected targets and performing probe measurement technique for counting the heads and isolating the human targets. Local space correlation is performed on the isolated human targets to detect the connectivity in order to determine the interaction. For example, shaking hands can be determined if there exists a blob connectivity between two individuals. Identification of human postures enables in efficient detection of human-object interaction. For example, for an object removal, postures of human walking, standing, bending etc are required. Human postures are detected by performing template matching. Training sets of human postures are collected from real scenarios and also geometric transformations like scaling, rotation etc., are applied on the collected samples for classification process [18]. After performing human isolation, the isolated images are matched with the collected templates.

In template matching, the correlation between two images is performed and the mathematical expression is given below:

$$\rho = \frac{COV(X_1,X_2)}{\sigma_1,\sigma_2} \qquad COV(X_1,X_2) = E(X_1X_2) - \mu_{1,}\mu_{2,}$$

where $\boldsymbol{\rho}$ is the correlation coefficient between two images $\boldsymbol{X_1, X_2}$ and $\boldsymbol{\sigma_1, \sigma_2}$ are the standard deviations. The correlation factor ρ value varies between [-1,1]. If ρ is negative, then the two images are inversely correlated. If ρ is positive towards 1, then the images are strongly correlated. If ρ = 0, there is no correlation between the two images.

The major steps followed in HHI and HOI processing:

1. Human target detection using image differencing and environment zoning
2. Perform target isolation as shown in Figure 6
3. Detect number of targets
4. Identify target-target directional vector and classify Entity-Entity relationship
5. Perform human posture classification - template matching

6. Identify object profile through template matching and template differencing
7. Detect HHI and HOI events and generate semantic annotations

For detection of number targets, movements of individual and blob information is used for indoor and outdoor monitoring. For indoor, identification of number of head counts is also used as shown in Figure 7.

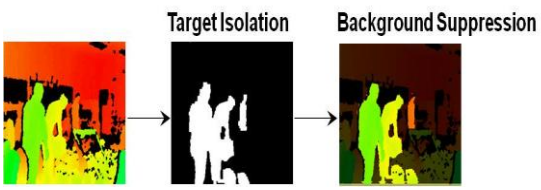


Figure 6: Target Isolation and Background Suppression

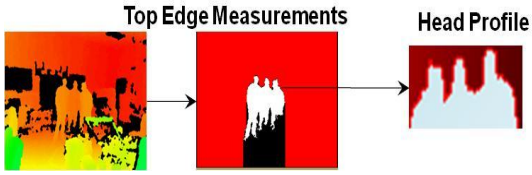


Figure 7: Extraction of Team Members Head Profile

The profile of head is extraction by used top edge measurements on a binary image i.e. the perpendicular distance from the top edge of the image to the top edge of the binary image is calculated at various sequential points. This feature vector is used to detect the pattern of head assuming the usual shape of head is oval or circle. Using this feature, the detection of separation point of the individuals is also detected.

3.5 Hypothesis Tracking / Observation Recognition

The atomic events of the discussed interactions are detected through HVI, HHI, HOI techniques. In block-E of CESI, pipelines of detected sequential atomic events are matched to an observation via Hypothesis tracking and Ontology mapping. An observation may be a combination of events or an evidential event. For example, person running or driver exiting from a car. Ontology developed for HVI is discussed in details in [8]. Figure 8 shows the process of Observations recognition and TMI activity recognition. In our work, we had developed hundreds of ontologies for Team Members Interactions. The matched observations are collected in a pipeline and mapped with the known ontologies through a Modified Sequential – Hidden Markov Model. State transition modeling is done for each sequential observation pair in the pipeline and an average of state transition rate of the evidential observation pair is computed to indicate the level of suspiciousness involved in the activity as depicted in Figure 8.

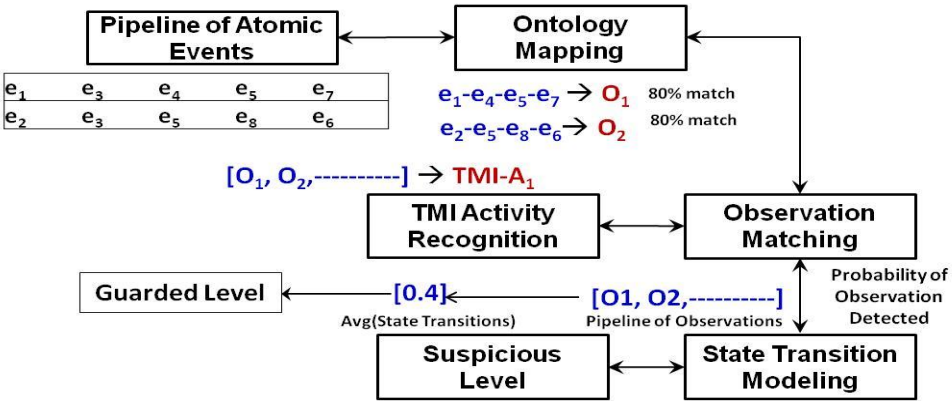


Figure **8**: Process of Observations & TMI Activity Recognitions

3.6 State Transitions

In block-S of CESI, the State Transitions (ST) between successive evidential events (i.e. observations) are modeled to determine the activities' temperament. A state transition model is developed which poses six categories of state transitions including: Human state transitions of Object handling, Visibility, Entity-entity relation, Human Postures, Human Kinematics and Distance to Target [16]. For example, Five different possible observations made in human handling objects (none, carrying small object, carrying large object, placing object, removing object), had been

considered. In our work, observations from two sequential frames were analyzed, observation (O) at frame *f*(*i*) and observation at frame *f*(*i*+1). For example in Human Object Handling state transition [16]: when two sequential frames were analyzed, observation (O) at frame f(i) evidences subject-1 with no object and observation at frame f(i+1) evidences subject-1 carrying an object as shown in Figure 9 at time t_6 and t_7. In general, the severity associated between two sequential observations can be computed using

$$.Severity\big(f(i) \rightarrow f(i+1)\big) = ST(O(f(i)) \rightarrow O(f(i+1))) \tag{3}$$

These sequential observations may involve a suspicious motive from human perspective. Analyzing the following state transitions helps in determining the probability of risk involved in a suspicious behavior as shown in Figure-8.

The severity between sequential frame decreases if a smooth state transition exists. For example in Object handing, the following is a sample for a smooth state transition.

none→carrying small object→placing object

The probability of an observation possibly occurring is computed by:

$$P(O(j)_{f(i+1)}) = P(O(j)_{f(i+1)}) + (1 - P(O(j)_{f(i)}))/2 \tag{4}$$

where O(j) is the observation of event j and $P(O(j)_{f(i)})$ is the current probability of observation j occurring in frame f(i).

More details of State transitions can be found in [7]. More assessments can be done if more sequential images are available for a given time frame. As the time frame between two images increases, the severity involved in state transitions decreases as there is a possibility of a smooth transition exists.

$$Time\ factor\ \alpha\big(f(i) \rightarrow f(i+1)\big) \uparrow, then\ ST(f(i) \rightarrow f(i+1)) \downarrow \tag{5}$$

In our work, we considered α=1. Furthermore, a Situation Awareness measure can be effectively calculated by modulating the spatiotemporal information and specific event importance subjective to the user. For example, the transition of observation (walking →running) elevates more in a vigilance environment than in a parking lot.

4 MODIFIED SEQUENTIAL-HIDDEN MARKOV MODEL

A Modified Sequential – HMM is developed for predicting the TMI activities from the traced evidential sequential observations. Hidden Markov Models attempts to model dynamical systems whose latest output depends only on the current state of the system. MS-HMM system infers the most likely sequence of states that can produce a given output sequence and also predicts the most likely next state of the model. The flow of process involved in the model is shown in Figure 9.

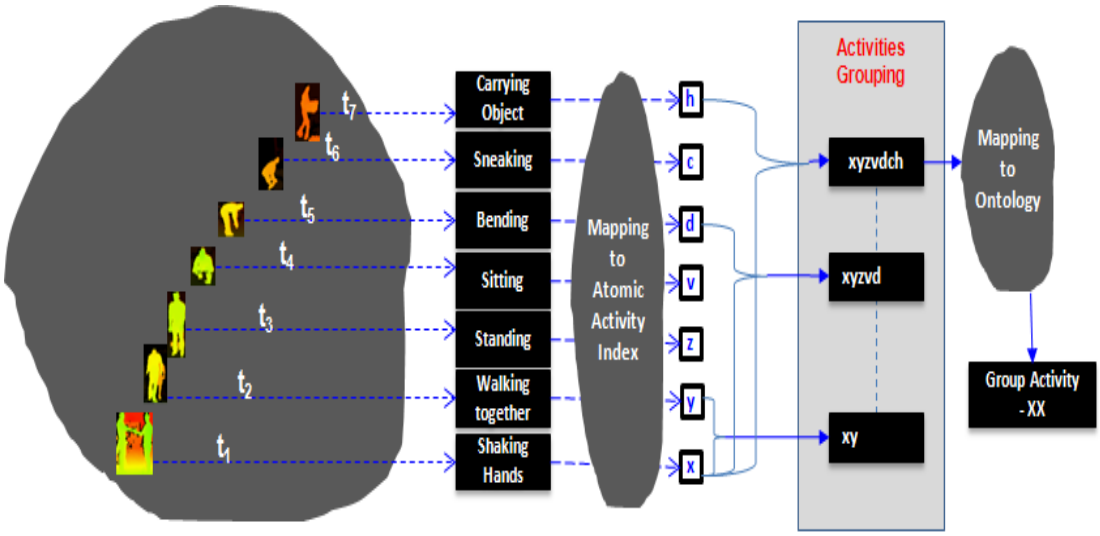


Figure 9: Human Activity Prediction through MS-HMM

MS-HMM calculates the probability of a given sequence of outputs originated from the system. For each evidential observation, it is assumed that there exists a state transition and an output emission transition. Hidden Markov Models can be seen as finite state machines where for each sequence unit observation there is a state transition and, for each state, there is a output symbol emission. Mathematical MS-HMM can be expressed as [20]:

$\lambda = (A, B, n)$ where A is an NxN state transition probability distribution matrix, B is an NxM observation symbol probability distribution and n is the initial state distribution. The developed system can accommodate both kinect and optical images for activity prediction. The atomic events from human interactions and observed state transitions are identified as detailed above. Grouping these relational observations and understanding the type of activity took place demands an effective prediction model. For this purpose, we proposed an Modified Sequential – Hidden Markov Model in combination with predefined TMI ontologies.

Captured images from a monitored area are preprocessed for detecting atomic events. Then these events are mapped to the atomic activity numeric index. Based on the emerging events, MS-HMM predicts the type of activity involved. As more events emerge, accuracy of prediction activities improves. Percentage of confidence involved in identifying an event through pre-processing is also fed back to system. If an event is identified with less confidence, for example: event detected from a single source, its corresponding event index is removed and fed back to the model for tuning the activity prediction for better confidence. System accommodates to predict pertinent human activities by eliminating the events detected with less confidence. Table-2a shows a sample of indexed events. Table-2b predefines sequence of human behavioral activities of certain tasks (e.g., loading or unloading of objects).

Table 2a: Events Index

Observations	Events ID
Vehicle Detected	10
Person Detected	20
Moving towards Trunk	67
Opening Trunk	63
Object Removed	51
Object Carried	54
Walking	31
Bending	33
Object Placed	53

Table 2b: Predefined Behavior model

Sequential Observations	Original Activity
10-20-67-51-54-31-33-53-31-67-51-54-31-33-53-31-67	Unloading
20-33-34-41-42-21-23-44-21-33-41-42-21-23	Unloading
10-20-42-21-33-34-44-21-41-42-21-10-33	Loading
20-42-21-33-34-44-21-41-42-21-33	Loading
10-20-42-21-23-44-21-10	Suspicious Object Dropping
20-21-23-41-42-21	Suspicious Object Dropping

Spatiotemporal realtionship and frequency of events occurrence are also enforced in our activity detection. Certain operational activities can be recognized based on the frequency of occurrence of Sub Activities (S_A). Detecting the frequency of occurrence determines the suspiciousness involved in the activity.

$$Frequency\ F(S\_A) > n\ \&\ Time\ elapse\ T < t, then\ Situation\ Awareness\ \uparrow$$

For eample if a person picks up an object, it is diagnosed as ‘Object Removed’ and if a person drops an object in a space-x, it is termed as ‘Object Placed’. If ‘Object Removed’ and ‘Object Placed’ happens more frequent, then the activity is identified as ‘Loading’ and ‘Unloading’ operations.

5 EXPERIMENTAL RESULTS

This section describes an experiment carried out for validation of TMI activity prediction in a context of human-vehicle, human-human and human-object interactions in a monitored environment as depicted in the Figure-11. The tested

loading and unloading experimental scenario is as follows: Vehicle-1 arrives at parking lot-1. Subject#1 and subject#2 get off the vehicle-1 and wait at the parking lot-1. Vehicle-2 arrives at parking lot-1 and parks near vehicle#1. Subject#3 gets off the vehicle-2 and meets subject#1 and subject#2. Subject#3 shakes hand with subject#1 and subject#2. Subject#3 opens the vehicle-2's trunk. All subjects unload a few square shaped box objects from vehicle-2's trunk to vehicle-1's trunk. Subject#2 and subject#3 gets into vehicle#2 and leaves the parking area (P-1). Though nothing peculiarly critical about this scenario, we intentionally set up this experiment to verify how well our ontology-based approach can assist in detecting a TMI activity in different stages of interactions between team member, vehicle and objects.

Figure 10 shows the target isolation of vehicle-1. Vehicle-1 is detected based on a motion detection algorithm and size factor of the object. The location of the vehicle-1 parked in parking area (P-1) is also detected using the Zoning of monitored environment as shown in Figure 4. For clarity of image details, the zoning of the monitored environment is not shown in below figure. The corresponding environment zoning is shown in Figure 4.

Image Processing steps for Target Isolation:

1. Apply Background differencing
2. Apply optimum thresholding technique
3. Apply 1 level erosion and remove smaller blobs
4. Apply 1 level dilation and image closing for target model
5. Apply exclusive morphology operation between binarized target model image and original image for isolating the target with RBG color

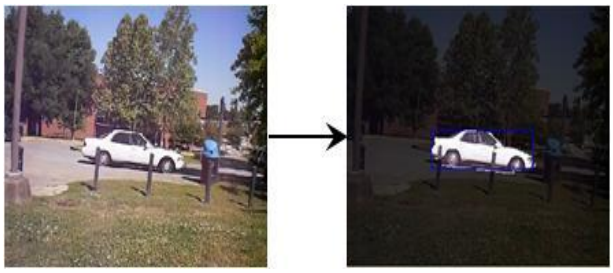


Figure 10: Vehicle-1 Target Isolation

Figure 11 shows a sample of optical images and the corresponding processed images in the experimented TMI activity scenario.

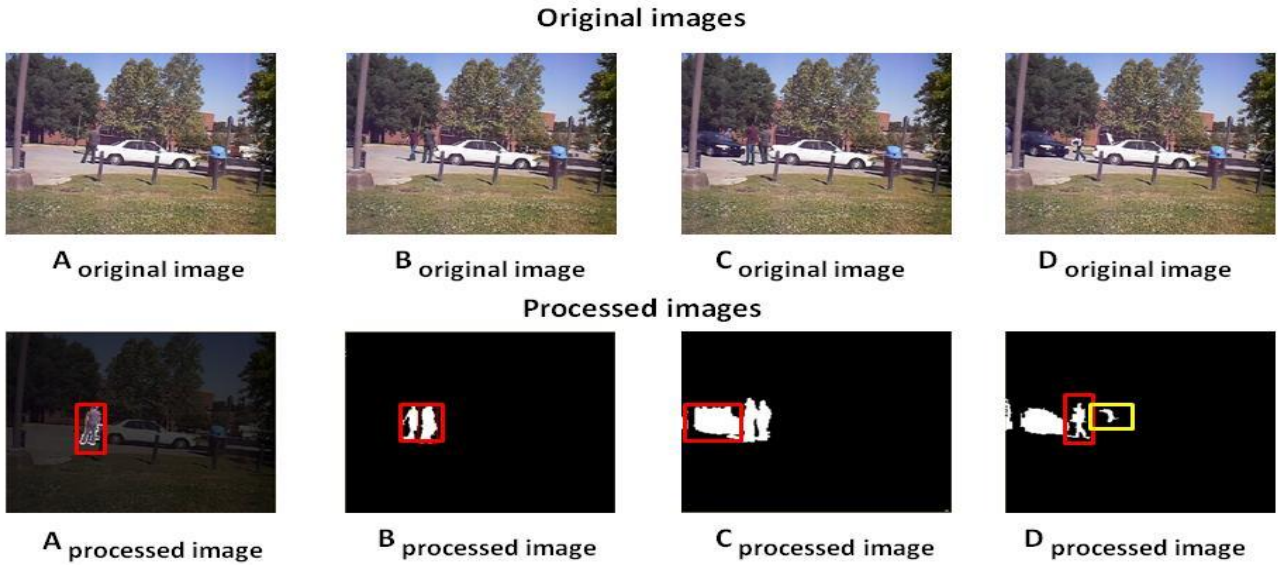


Figure 12: Original Image and Target and Activity Detected Processed Image

As seen in image-A, subject#1 is detected standing near the vehicle-1. Image-B shows the detection of subject#2 and subject#1 standing together. After processing sequential images and performing motion detection algorithms and factoring in the size of the object, vehicle#2 is detected parked in the driving area as shown in image-C. Image-D shows

the detection of vehicle-1 trunk open and detection of subject#1 carrying object towards vehicle-2. The detection of vehicle-2 hood open is identified using HVI ZoV technique as discussed in section 3.2. Through processing the sequential images, it was observed that subject#2 and subject#3 entered vehicle#2 and departed from the parking lot. Figure 13 shows the detection of subject#2 standing alone near the vehicle#1 in the parking lot.

Figure 13: Detection of Subject#2 standing near vehicle#1 after vehicle#2 left the parking lot

The following were detected in this TMI loading and unloading scenario:

Number of human targets detected → 3

Number of vehicle targets detected → 2

Sample of temporal relationships → subject#1 and subject#2 arrived before subject#3; vehicle-2 arrived after vehicle-1 and departed before vehicle-1.

Sample of Observation Frequencies → subject#3 unloaded object from vehicle-2 to vehicle-1 twice.

Sample of Observations detected → subject#2 got off from the vehicle-1 driver side door, confirms that subject#2 is driver of vehicle-1through HVI technique; subject#3 shakes hand with subject#1 and subject#2; and subject#1 departs with subject#3 in vehicle-2

Activity detected → Unloading and loading activity from vehicle#1 to vehicle#2

TMI unloading and loading of objects activity is detected based on sequential evidential observations and feeding to the MS-HMM as shown in Table-2a and Table-2b.

6 CONCLUSIONS

In this paper, we presented a Casual-Events State Inference (CESI) technique for spatiotemporal recognition and reasoning of team members interactions. Casual Events State Inference approach is employed with recognition of HVI, HOI and HHI activities that are spatiotemporally correlated and the state transition of observations. An outdoor experiment involving a loading and unloading operational activity is demonstrated. The developed MS-HMM can efficiently categorize the TMI activities based on the evidential observations. The major contribution of this paper is towards identifying high level operational TMI activities and not only focusing on low level activities through MS-HMM as discussed earlier. State transition modeling in conjunction with Casual Events Space Inference technique can efficiently and effectively measure the severity associated in a human activity and can facilitate generation of semantic annotations for reporting sensor observed activities [16]. Though the conducted experiments demonstrate the efficiency of TMI activity detection in optical images, various indoor experiments were also conducted using kinect depth map images. Through our indoor and outdoor experiments, it was found that kinect depth images can efficiently detect TMI activities in indoor under normal light conditions and also under complete darkness. Kinect images can also be used in outdoor under dark environments.

ACKNOWLEDGMENTS

This work has been supported by a Multidisciplinary University Research Initiative (MURI) grant (Number W911NF-09-1-0392) for "Unified Research on Network-based Hard/Soft Information Fusion", issued by the US Army Research Office (ARO) under the program management of Dr. John Lavery.

REFERENCES

[1] Marco A. Pravia, Olga Babko-Malaya, Michael K. Schneider, James V. White, and Chee-Yee Chong, "Lessons Learned in the Creation of a Data Set for Hard/Soft Information Fusion," 12th International Conference on Information Fusion, July 6th – 9th, 2009, Seattle, WA.

[2] D.L. Hall, J. Llinas, M. McNeese, and T. Mullen, "A Framework for Dynamic Hard/Soft Fusion," Proc. 11th Int. Conf. on Information Fusion, Cologne, Germany, July 2008.

[3] Rababaah H., and Shirkhodaie A., "A Survey of Intelligent Visual Sensors for Surveillance Applications," IEEE Sensor Application Symposium, February 6-8, (2007).

[4] Joshua Candamo, Matthew Shreve, Dmitry B, Goldgof, Deborah B. Sapper, and Rangachar Kasturi, "Understand Transit Scenes: A Survey on Human Behavior – Recognition Algorithms" in IEEE Transactions on Intelligent Transportation Systems, Vol.11, No.1, March 2010.

[5] N. Sulman, T. Sanocki, D. Goldgof, and R. Kasturi, "How effective is human video surveillance performance?" in Proc. Int. Conf. Pattern Recog., 2008, pp. 1–3.

[6] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. Syst., Man, Cybern.C, Appl. Rev., vol. 34, no. 3, pp. 334–352, Aug. 2004.

[7] Vinayak Elangovan, and Amir H. Shirkhodaie, "A Survey of Imagery Techniques for Semantic Labeling of Human-Vehicle Interactions in Persistent Surveillance Systems," SPIE Defense, Security and Sensing, Orlando, Florida, April (2011).

[8] Vinayak Elangovan, and Amir H. Shirkhodaie, "Context-Based Semantic Labeling of Human-Vehicle Interactions in Persistent Surveillance Systems," SPIE Defense, Security and Sensing, Orlando, Florida, April (2011).

[9] D. Merad, K.-E. Aziz, N. Thome, "Fast People Counting Using Head Detection from Skeleton Graph," avss, pp.233-240, 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010.

[10] Stephen Gould, Tianshi Gao, Daphne Koller, "Region-based Segmentation and Object Detection," Proceedings of Advances in Neural Information Processing Systems (NIPS), 2009.

[11] T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications ", Applied Imagery Pattern Recognition Workshop, 37th IEEE AIPR pp1-8 Oct 2008.

[12] Ismail Haritaoglu, David Harwood and Larry S. Davis, "W4: Real-Time Surveillance of People and Their Activities," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, August 2000.

[13] Shunsuke Kamijo, Masahiro Harada and Masao Sakauchi, "An Incident Detection System Based on Semantic Hierarchy," IEEE Intelligent Transportation Systems Conf., Oct. 3rd–6th, 2004, Washington, DC.

[14] Shirkhodaie, A., "Rababaah, H. " Multi-Layered Impact Modulation for Context-based Persistent Surveillance Systems, " in the SPIE 2010 Defense, Security, and Sensing Conference, Multi-Sensor, Multi-Source Information Fusion: Architectures, Algorithms, and Application, April 2010, Orlando, FL.

[15] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," IEEE Pervasive Computing, vol. 9, pp. 48–53, 2010.

[16] Vinayak Elangovan, and Amir H. Shirkhodaie, "Recognition of Human Activity Characteristics based on State Transitions Modeling Technique," SPIE Defense, Security and Sensing, Baltimore, Maryland, April (2012).

[17] Rababaah, H., and Shirkhodaie, A., “Feature Energy Assessment Map (FEAM): a Novel Model for Multi-Modality Multi-Agent Information Fusion in Large-Scale Intelligent Surveillance Systems, Networks,” SPIE Defense, Security, and Sensing Conference, Multi-Sensor, Multi-Source Information Fusion: Architectures, Algorithms, and Application, paper 7345-19, April 13-17, 2009, Orlando, FL.

[18] Rababaah H., Shirkhodaie A., ”Human Posture Classification for Intelligent Visual Surveillance Systems, “ in the SPIE Defense and Security, Visual Analytics for Homeland Defense and Security, March 17-20, 2008, Orlando, FL.

[19] Xiquan Yang, Ye Zhang, Na Sun, Deran Kong, “Research on Method of Concept Similarity Based on Ontology,” Proc. of the 2009 International Symposium on Web Information Systems and Applications (WISA’09), China, May 22-24, 2009, pp. 132-135.

[20] L. R. Rabiner, "A Tutorial on Hidden Markov Models, and Selected Applications in Speech Recognition," *Proc. IEEE*, Vol. 77, No. 2, pp. 257--286, Feb. 1989.

[21] Amir Shirkhodaie, Vinayak Elangovan, Aaron Rababaah, “Acoustic Semantic Labeling and Fusion of Human-Vehicle Interactions”, SPIE Defense, Security and Sensing, Orlando, Florida, April (2011).