A Survey of Imagery Techniques for Semantic Labeling of Human-Vehicle Interactions in Persistent Surveillance Systems

Vinayak Elangovan
vinayake@gmail.com
Dept. of Mech. & Mfg Engr.
Tennessee State University
TN, U.S.A.

Amir Shirkhodaie
ashirkhodaie@tnstate.edu
Dept. of Mech. & Mfg Engr.
Tennessee State University
TN, U.S.A.

Abstract
Semantic labeling of Human-Vehicle Interactions (HVI) helps in fusion, characterization, and understanding cohesive patterns that when analyzed and reasoned, they may jointly reveal pertinent threats. Various Persistent Surveillance System (PSS) imagery techniques have been proposed in the past for identifying human interactions with various objects in the environment. Understanding of such interactions facilitates to discover human intentions and motives. However, without consideration of circumstantial context, reasoning and analysis of such behavioral activities is a very challenging and difficult task. This paper presents a current survey of related publications in the area of context-based HVI, in particular, it discusses taxonomy and ontology of HVI and presents a summary of reported robust imagery processing techniques for spatiotemporal characterization and tracking of human target in urban environments. The discussed techniques include model-based, shape-based and appearance-based techniques employed for identification and classification of objects. A detailed overview of major past research activities related to HVI in PSS with exploitation of spatiotemporal reasoning techniques applied to semantic labeling of the HVI is also presented.

Keywords: Human-Vehicle Interactions (HVI), Semantic Labeling, Imagery Data Processing, Sensor Data Fusion, Persistent Surveillance System (PSS), Spatiotemporal Reasoning, Human-Vehicle Taxonomy and Ontology.

1. INTRODUCTION

Various research activities have been undergone by previous researchers in improving the homeland security intelligence using live camera feed for fighting against crime and dangerous acts. Currently in many applications for detecting multiple suspicious activities through a real-time video feed, multiple analysts manually watch the live video stream to infer necessary suspicious information [1]. Vision analysis is expensive and also prone to errors since humans have some limitations in monitoring continuous signals [2]. Video surveillance and Human behavior analysis is one of the ongoing active researches in image processing [1-3]. Other research activities focus more on identifying spatiotemporal relations between human and vehicles for situation awareness.

Many outdoor apprehensive activities involve vehicles as their primary source of transportation to and from the scene where a suspicious plot is executed [30, 31]. Wheeled motor vehicles are used as major source for transporting in and take out ammunitions, supplies, and other suspicious belongings. Analysis of the Human-Vehicle Interactions (HVI) helps us to identify cohesive patterns of activities representing potential threats. Identification of such patterns can significantly improve situational awareness in PSS. For example, the Times Square Car bomb detection incident on May 1st, 2010 in New York had shown the importance of Soft Data. On the other hand, it also had increased fraught false alarms. Effective algorithm development for processing camera feeds can minimize human prone errors and also false alarms [4].

Human Activity recognition is a challenging problem in computer / artificial intelligence due to complexity involved in low-level processing, data alignment from different source (audio and camera) feeds and generating semantic messages. Human activities may involve human-human interactions, human-object interactions, human-environment interactions or multi-human-object interactions or multi-human-multi-object interactions. Various researchers had addressed object recognition like suitcase, box, cell phones etc. In-depth research work had been carried on interpreting human vehicle interaction in field of automotive engineering. Less attention had been addressed in interaction between human and vehicle for detecting cohesive patterns of suspicious activities. Understanding suspicious activities behaviors like smuggling in terms of artificial intelligence involves in-depth understanding of human-vehicle interactions and also addressing spatial-temporal relationship. Delivery or utility vehicle drivers may seem nervous or display non-compliant behavior such as insisting on parking close to a building or restricted area show an act of suspiciousness. For example U.S. Authorities investigates a man at a border crossing who seemed to be extremely nervous and was repeatedly glancing and checking on his vehicle's trunk [31]. In order to estimate the cohesive pattern of any suspicious act involving vehicle as a medium, it is highly vital to understand the human interactions with the vehicle. Pattern of interactions helps in identifying the involvement of suspicious act. For example, driving a truck or car with trunk open, dropping or picking up a bag from a trunk with a fast timely response, running out from a car, etc., creates suspiciousness in human behaviors.

This paper is organized as follows: Object Identification & Classification Techniques, Human & Object Tracking Techniques, Spatiotemporal Characterization for Human-Object Interactions, HVI Taxonomy & Ontology, followed by Conclusion, Acknowledgements and References.

2. OBJECT IDENTIFICATION & CLASSIFICATION TECHNIQUES

The need for object identification and classification technologies had gained a vital role in recent years in terms of increased security awareness in homeland security and in persistent surveillance systems. This section briefs the research activities put forth in the past for Object identification and classification techniques and draws a special attention towards vehicle detection and identification. Various methods had been proposed for object identification. In general the approaches can be classified as Model based approach, Shape based approach and Appearance based approach.

Model-based approaches for object classification approximates the object as a collection of three dimensional geometrical primitives represented as boxes, spheres, cones, cylinders, generalized cylinders, surface of revolution etc., [16]. Developing an object model can be achieved by performing computer aided design (CAD) models or by extracting objects from real images. In [18] the models were constructed from real images taken at various camera depression angles and rotating objects horizontally**.** Researchers in [17] proposed a model-based tracker to identify aerodrome ground lighting for guiding the pilot onto the runway lighting pattern and taxi the aircraft into its respective terminal. The developed model-based approach matches a template of the ALS (approach lighting System) to set of extracted luminaries from the camera image and thereby estimating the camera's position. Modeling an object can be done by object-centered representation which uses object features like boundary curves, surfaces etc., or a view-centered representation which uses different view points to represent an object [19]. In [20] an approach had been developed for categorizing multi-view object from a video in which the topology of the images are identified and images are arranged by a connectivity graph employing temporal and spatial information of videos. The distinct features of images are extracted using e Scale-Invariant Feature Transform (SIFT) operator. Initially a neighborhood graph is generated for each instance of object by identifying Euclidean distance for each image pair and by applying constraint on physical proximity between successive video frames. Redundant informative images are removed by image interpolation or extrapolation. The generated neighborhood graphs are clustered together by a particular view point.

A shape-based approach represents an object through its shape or its contour. In [26], object silhouettes are extracted from each video frames using feature tracking, motion grouping and joint image segmentation. These object masks are matched and aligned to 3D silhouettes model for object recognition. In [40], a shape-based object detection and matching method named as Shape Band had been proposed. Shape band models an object within a bandwidth of its contour representing a deformable template. Matching is done with the template and object instance within the bandwidth. Whereas in [41], a boundary band method had been proposed which can extract object repetitions along with their deformations. The object locations were selected by matching the boundary band at every image pixel locations. They had used an edge-based approach for detecting and matching repeated elements. The local and global geometric information about object of interest are considered and Fast Fourier transform is used for performing the distance measure.

Silhouette-based posture analysis was proposed in [42] 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 [42]. W^4 segments objects from their silhouettes and also uses appearance models for tracking in subsequent frames. It can detect and track group of people and monitors human-object interactions like dropping and removing an object, and exchanging of bags. Events can also be detected by imitating human-vehicle behavior through semantic hierarchy implementing Coordinate, Behavior and Event class, determining the behavior of individual vehicle in detecting the traffic events on the road [43].

Appearance-based approach models only the appearance by extracting the local or global features of an object [16]. In appearance based approaches, typical methods used in object identification includes principle component analysis [22, 24], independent component analysis [23, 25], singular value decomposition, discrete cosine transform [29], Gabor wavelet features [29], discrete wavelet transform [29], neural networks based learning, evolutionary pursuit based learning, active appearance models, shape and appearance-based probabilistic models and combinations of several existing techniques. The proposed Appearance based approaches for object identification by most researchers in past mainly rely on dimensions equality such that the feature vectors represent same feature space. To overcome this, [5] had proposed a theory of subspace morphing which allows scale-invariant identification of object images which doesn't require scale normalization. In [29] the performance of discrete cosine transform, Gabor transform, discrete wavelet transform feature extraction methods are analyzed, and presented that discrete cosine transform method is more suitable for Gaussian mixture model in automatic image annotation though none of the three performs better in all classes for feature extraction. Motion-appearance based approach had been used for modeling a vehicle in our research work for Semantic Labeling of Human-Vehicle Interactions in Persistent Surveillance Systems [38] and image processing techniques for vehicle model is shown in Figure 1.

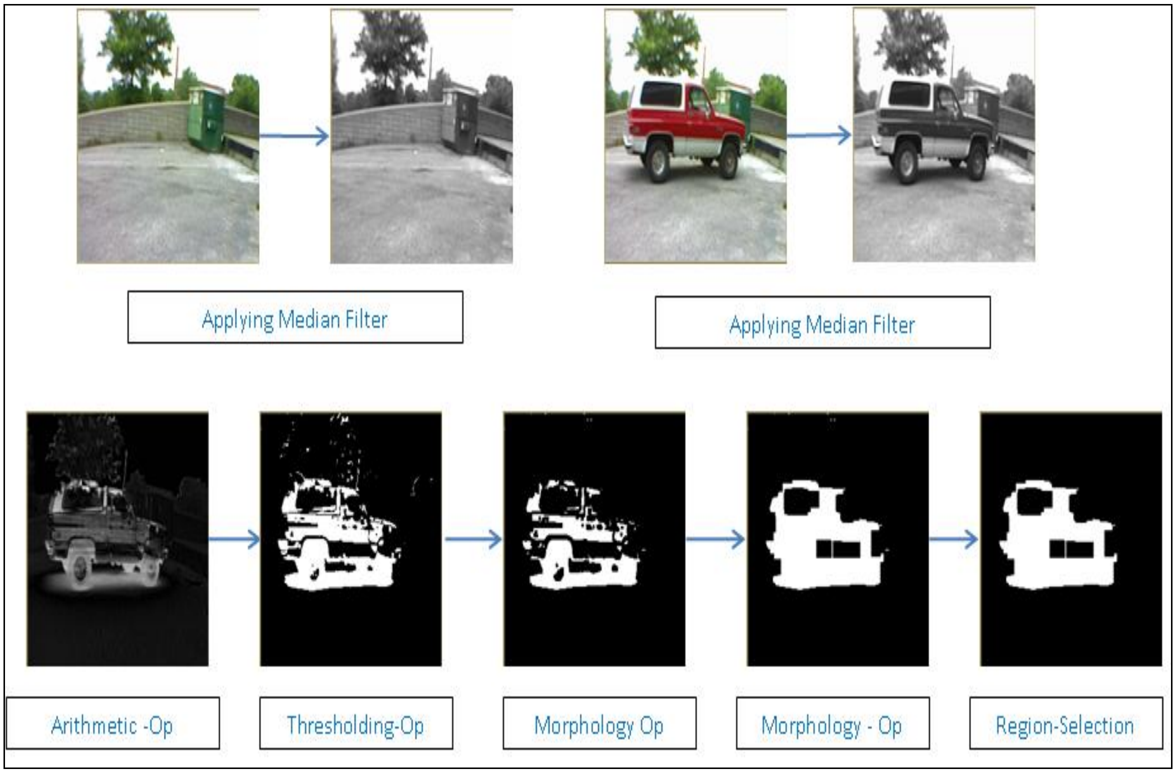

Figure 1: Image processing operations for modeling vehicle.

Several image processing techniques had been developed in past for identifying human and vehicle blobs in a given image. In general, blob detection is used to obtain regions of interest for further processing. In images of indoor scenes, human blobs are relatively large and so it is relatively easy to model the background to subtract the human objects. In contrary, outdoor scene images often have wide range of depth, and size of human blob and his/her interacting objects can be very small. [6] had developed appearance-based model using color correlogram to track humans and objects, segment human and objects even under partial occlusion. Color correlogram consists of a co-occurrence matrix expressing probability of finding a color 'c1' at a distance 'd' of a color 'c2' for specific distances. Maximum likelihood criterion had been used for classifying blobs from two merging blobs. Many others had considered the aspect ratios for classifying human and objects.

Various approaches have been developed earlier for vehicle detection using features extraction and learning algorithms. Some classical approaches [7-9] use background subtraction to extract motion features for moving vehicle detection. Wavelet transforms [10], Gabor filters [11], and vehicle templates can be used for vehicle detection given static images to extract texture features for detecting a vehicle object. [12] proposed a global color transform model and used vehicle edge map features for categorizing vehicle into eight orientations i.e. front, rear, left, right, front-left, rear-right by employing normalized cut spectral clustering (N-Cut).

3. HUMAN & OBJECT TRACKING TECHNIQUES

Classical approaches utilizes geometric transforms, signal extraction, noise reduction, and other techniques in detecting and tracking human and objects effectively using fixed cameras. Tracking of objects can be done in various ways like employing region based tracking, contour based tracking, feature based tracking, model based tracking etc., [3]. By dynamically extracting the foreground information i.e. employing periodic background subtraction, region based tracking can be done for moving objects [44, 45]. In contour based object tracking, objects were tracked through contour boundaries and dynamically updating the object contours by extracting the shapes of the objects which are computational efficient [46, 47]. Feature based tracking tracks objects by extracting feature vectors and mapping those with objects in successive images. Classical feature based tracking involves extracting global features [16, 48] or local features [3, 16] or based on dependence graphs [3]. By matching the model generated for objects, model based tracking algorithms track the objects either by using computer aided tools or through other computer vision techniques [3, 49, 50]. Detecting and tracking a human or vehicle can also done abstractly by zoning the environment in different connected zones through segmenting the background area which is to be monitored and labeling each zone as shown in Figure 2 [38].

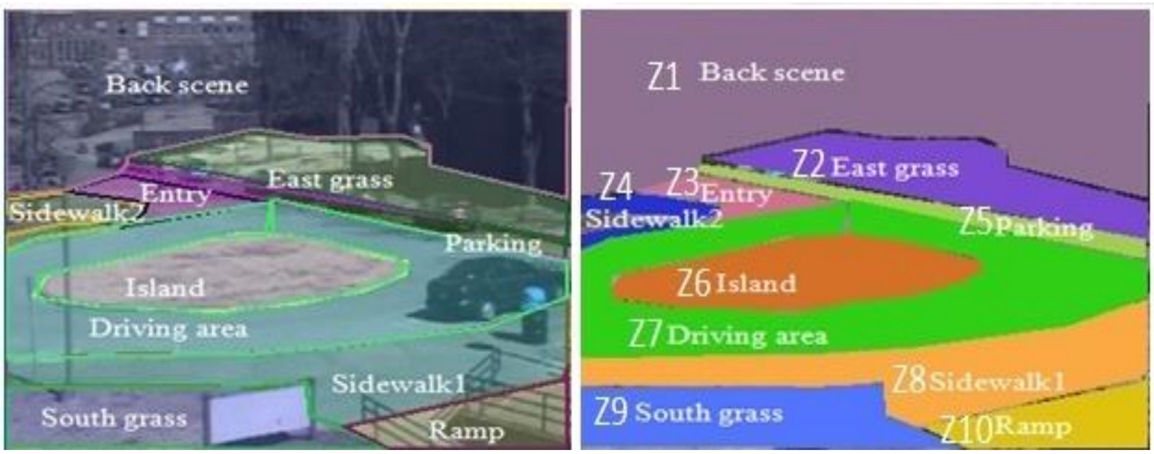

Figure-2: PSS Environment.

Researchers in [27] presented a patented novel segmentation algorithm technology, called ItBlt™, designed to identify and track cars at military check-points, civilian toll booths, and border crossing points. They could effectively locate and track a car object by discover approximate straight lines; performing image segmentation by applying taxonomy rules to find sets of lines that can be connected to form parts of cars like windshield, hood, front bumper area, trunk, etc.; outlining all cars in image by assembling segmented car parts; identifying unique features like dents, debris, and dirt patterns and using a priori information from each frame for accelerating the segmentation of subsequent frames.

Researchers in [13] address tracking pedestrians and capturing pedestrian images and pedestrian activity recognition based on position and velocity. They track objects appearing in a digitized video sequence with the use of a mixture of Gaussians for background/foreground segmentation and a Kaman filter for tracking. They had addressed activities such as walking, running, loitering and falling to ground. They do not directly observe the motion of the pedestrian through articulated motion analysis and their system does not distinguish between objects moving at the same speed through different means, such as a bicyclist and a runner. An appearance-based model is used to track humans and objects and segment them, even under partial occlusion in [6]. This model is based on a color correlogram that consists of a co-occurrence matrix expressing the probability of finding a color c1 at a distance d of a color c2 for specified distances. To extract the moving regions in an image, temporal differencing can be employed where pixel-wise differences are estimated [3]. Temporal differencing is very adaptive to dynamic environments but blob closing needs to be applied. Connected Component Analysis can be used to cluster the regions where motions are detected.

Human-object interaction in terms of object removal and left behind can be identified by following way: When object blob splits into two where one of them is static and other is moving, then that static blob is an object that had been left behind. When a static blob is combined to a moving blob then that static blob is an object being removed. Tracking of an object can be done backwards for object left behind by analyzing previous images. When the environment is studied during the background modeling, an object dropped or removed by a person can be subsequently detected and considered as a foreground [28]. Appearance based object recognition systems typically operate by comparing a two dimensional image of object against a prototype and finding the closest match. Researchers in [14] proposed a method of categorizing hierarchy for generic object recognition using PCA based representation of appearance information. In [28] the human-object interactions address picking and dropping of an object. The consistent labeling is simply the color tagging the individual human and object. Further spatiotemporal relations have not been addressed for human-object interactions in identifying any suspicious activities.

4. SPATIOTEMPORAL CHARACTERIZATION FOR HUMAN-OBJECT INTERACTIONS

Interpreting human behaviors with respect to context involves spatial and temporal reasoning [33]. Two basic classical approaches for representing temporal information are Allen's Interval Algebra, and Vilian's and Kautz's Point Algebra [36]. Allen's temporal logic [34, 35] algorithm uses a composition table to reason about networks of relations. Thirteen atomic relations between two intervals have been by Allen namely before, after, meets, met, overlaps, overlapped, starts, started, during, contains, finishes, finished and equals as shown in Figure 3. Whereas the point algebra uses time points and three possible relations namely precedes, same and follows to describe interdependencies among them.

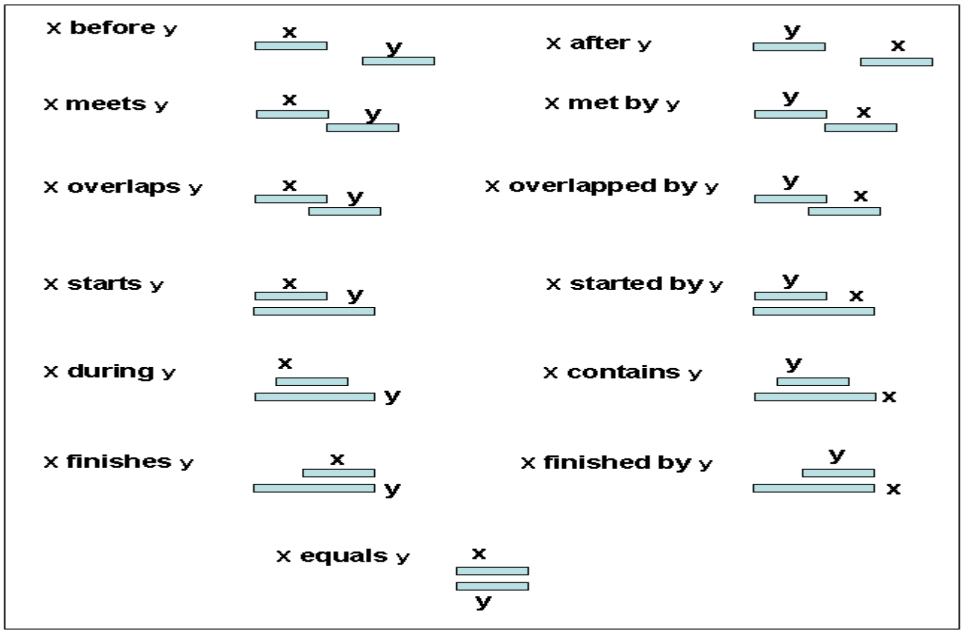

Figure 3: Allen's Temporal Logic 13 Base Relations.

Vilain and Kautz addressed that many interval relations can be expressed in terms of point relations by using end points of starting and endings of intervals [33] as shown in Figure 4. For example the time point of car driver side door being opened after the car parked precedes the time point of car trunk being opened.

Another qualitative approach for spatial representation and reasoning is Region Connection Calculus (RCC) which describes regions in Euclidean / topological space by 8 basic relations that are possible between two regions. The 8 basic relations proposed are disconnected (DC), externally connected (EC), equal (EQ), partially overlapping (PO), tangential proper part (TPP), tangential proper part inverse (TPPi), non-tangential proper part (NTPP), non-tangential proper part inverse (NTPPi) [33, 37] as shown in Figure 5.

Various research activities had been carried in analyzing or interpreting human actions and suspicious events in context of persistent surveillance systems implementing the above mentioned qualitative approaches [51, 52]. In [32] detecting and tracking vehicles moving from an image sequence using spatial-temporal relations, when the vehicles to be tracked are at far distance from the camera have been addressed. Vehicle motion data was extracted to construct map to learn the expected area where vehicle can be observed moving and where vehicles could be occluded.

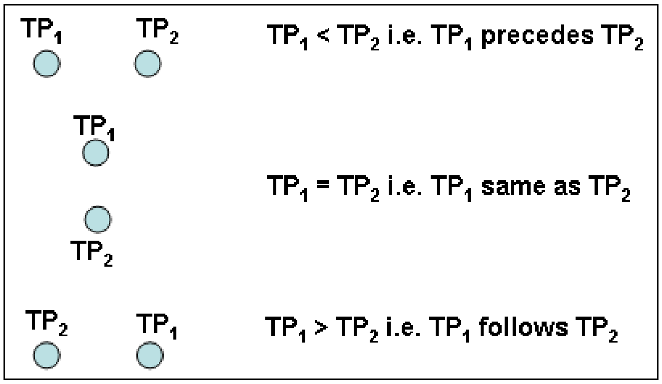

Figure 4: Vilain's and Kautz's 3 Possible Relations.

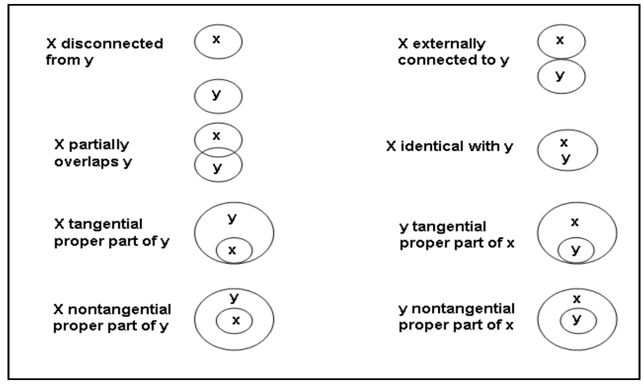

Figure 5: Region Connection Calculus 8 Relations

5. HVI TAXONOMY & ONTOLOGY

Development of taxonomy and ontology helps in narrowing the analysis in decision making of human-vehicle interactions. Various taxonomy based approach had been done by previous researchers in identifying human interaction with vehicle. Most of them have addressed in automating the vehicle [53, 54]. Very less attention has been drawn towards analyzing the pattern of activities in human-vehicle interaction. In [15] a synthetic 3D vehicle shape based model for localization of vehicles and specification of regions-of-interest i.e. front doors had been proposed. This paper deals with opening and closing of vehicle front doors. They haven't addressed the activities in other region of interest like opening and closing of hood, trunk, back doors etc. Complexity involved in human-vehicle interactions analysis are occlusion of human, deformity of vehicle shapes, orientation of vehicles, orientation of camera etc. For Vehicle localization, detection and foreground blob tracking is done by employing background elimination. For classifying vehicle blobs, the extracted blobs are matched with database of silhouettes of synthetic 3-D vehicle models. Similarity of blob is done by comparing the features of blob at each given frame. They had built synthetic 3D vehicle models for a sedan and SUV and then they extract 2D templates from 3D models. Motion analysis is done by identifying the optical flow. An adaptive Gaussian mixture model for background subtraction is employed. Connected component analysis and erosion / dilation are applied to the detected foreground pixels for extracting foreground blobs. Noisy blobs or blobs of no interest are removed by applying desired threshold. Ontology development in visual analysis systems is done by inspecting the relations between the entities, entities and object considering the spatial and temporal information [55 - 58]. In [55] ontology design for surveillance systems in a building had been discussed. A hierarchical ontology based framework had been presented in [56] to analyze events in dynamic scenes like person walk along sideway, car enter parking lot etc. The semantic structure they had followed is [Entity Word Attribute]. Though the events had been represented semantically, proper linguistic structure and reasoning with respect to spatial-temporal relations haven't been addressed. In [60] a rule-based system and state machine framework had been used to analyze the videos in three levels of context hierarchy i.e. attributes, events, and behaviors to identify the activities and classifying human-human interaction as one-to-one, one-to-many, many-to-one and many-to-many using an meeting ontology.

A thorough analysis had been in done in our current research work towards developing taxonomy for identifying human-vehicle interaction in persistent surveillance systems [38, 39] and the developed HVI taxonomy is shown in Figure 6.

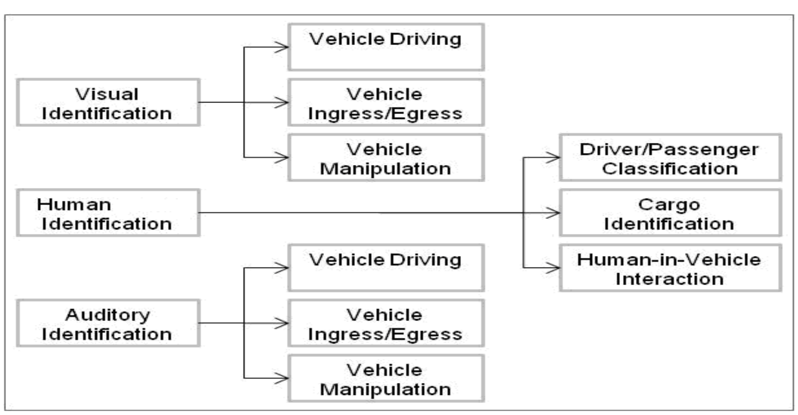

Figure 6: Human-Vehicle Interaction (HVI) Taxonomy

Visual and auditory identification are the data collected through camera sensors and acoustic sensors respectively. We had used HVI taxonomy as a means for recognizing different types of HVI activities. HVI taxonomy may comprise multiple threads of ontological patterns. By spatiotemporal linking of ontological patterns, a HVI pattern is hypothesized to pursue a potential threat situation [38, 39]. The data collection for identifying vehicle suspicious behaviors can be gathered either through visual identification or auditory identification or through human. Each category is further branched out to define more atomic HVI relationships, not shown here in brevity of space limitation. One such category i.e. Vehicle Manipulation had been shown in Figure 7. The Vehicle Manipulation can be classified as opening / closing, exterior manipulation and cargo manipulation. Each of these is further classified. For example exterior manipulation is further branched to explain where exactly the manipulation takes place and thereby raising the severity of suspiciousness.

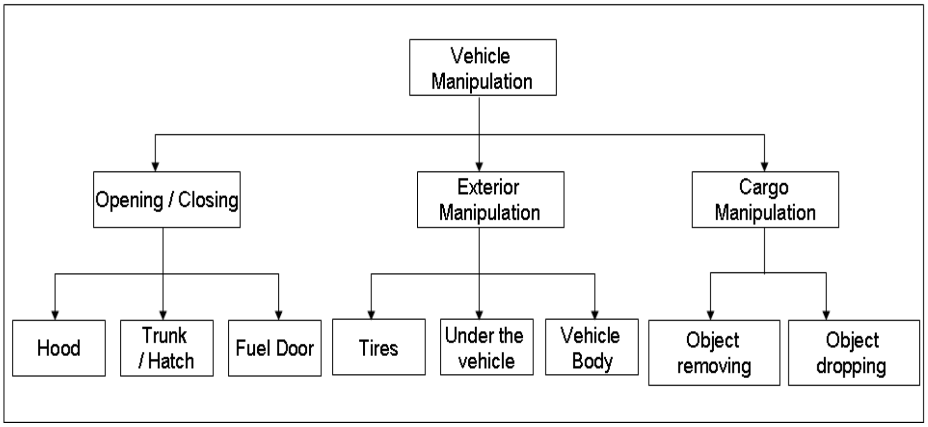

Figure 7: Taxonomy for Vehicle Manipulation

Figure 8 shows the image processing technique applied for an example of such interaction. Detection of events is done by foreground extraction. The extracted foreground is highlighted in the original image through arithmetic operation by performing RGB image averaging. As seen below the detection of human near the passenger door and opening of truck trunk is detected [38].

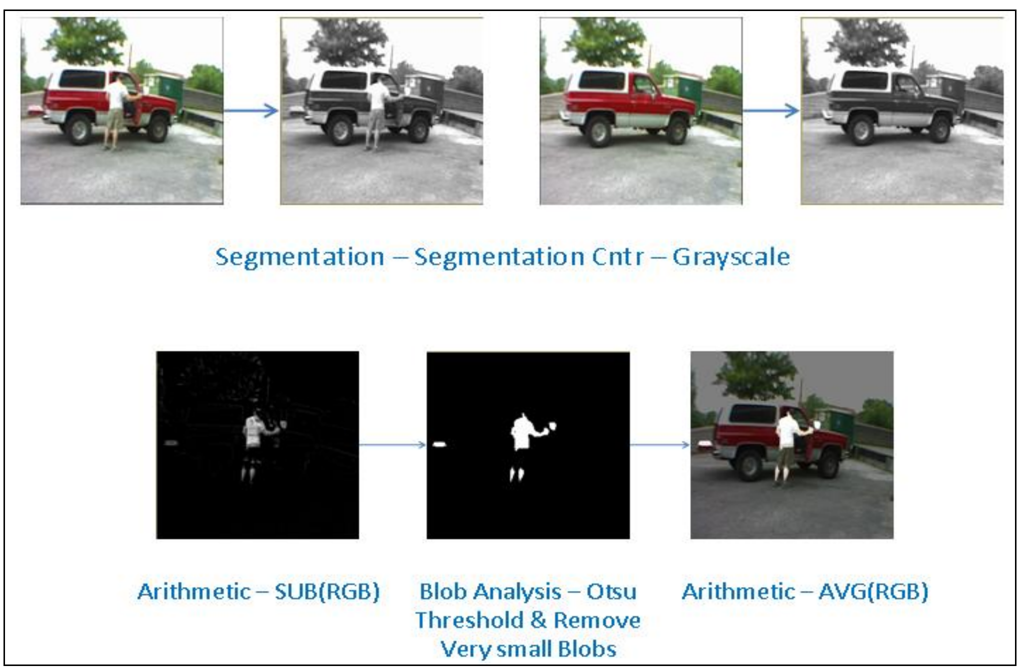

Figure 8: Example of detected Human-Vehicle Interaction (HVI)

6. CONCLUSION

This survey paper had addressed the current major techniques used for identifying and classifying objects; and also discussed tracking of humans and objects in terms of surveillance and security issues. Major image processing techniques involved in spatiotemporal characterization is also discussed here. Though not much work done in area of Human-Vehicle Interactions in Persistent Surveillance Systems, a current survey and related work are also presented in this paper.

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] 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.

[2] 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.

[3] 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.

[4] http://blog.newsweek.com/blogs/declassified/archive/ 2010/05/07/what-can-intelligence-agencies-do-to-spot-threats-like-the-times-square-bomber.aspx.

[5] Zhongfei (Mark) Zhanga, Rohini K. Sriharib, “Subspace morphing theory for appearance based object identification”, Computer Science Department, Watson School, State University of New York, Binghamton, NY 13902, USA, Center of Excellence for Document Analysis and Recognition (CEDAR), State University of New York, Buffalo, NY 14228, USA, 2002 Pattern Recognition Society. Published by Elsevier Science Ltd.

[6] M. Balcells-Capellades, D. DeMenthon and D. Doermann, An appearance-based approach for consistent labeling of humans and objects in video, Pattern Anal Appl 7 (2004) (4), pp. 373–385.

[7] R. Cucchiara, M. Piccardi, and P. Mello, “Image analysis and rule-based reasoning for a traffic monitoring,” IEEE Trans. on ITS, vol.1, no.2, pp.119-130, June 2002.

[8] G. L. Foresti, V. Murino, and C. Regazzoni, “Vehicle recognition and tracking from road image sequences,” IEEE Trans. on Vehicular Technology, vol.48, no.1, pp.301-318, Jan. 1999.

[9] V. Kastinaki, V. Zervakis, and K. Kalaitzakis, “A survey of video processing techniques for traffic applications,” Image and Vision Computing, vol.21, no.4, pp.359-381, 2003.

[10] J. Wu, X. Zhang, and J. Zhou, “Vehicle detection in static road images with PCA-and- wavelet-based classifier,” IEEE of the international conference on ITS, pp.740-744, 2001.

[11] Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection using Gabor filters and support vector machines,” IEEE International Conference on Digital Signal Processing, vol.2, pp.1019-1022, 2002.

[12] Jui-Chen Wu, Jun-Wei Hsieh, Yung-Sheng Chen, and Cheng-Min Tu, “Vehicle Orientation Detection Using Vehicle Color and Normalized Cut Clustering”, MVA 2007 IAPR Conference, Japan.

[13] R. Bodor, B. Jackson and N.P. Papanikolopoulos, "Vision-Based Human Tracking and Activity Recognition," Proc. of the 11th Mediterranean Conf. on Control and Automation, Jun. 2003.

[14] Gaurav Jain , Mohit Agarwal , Ragini Choudhury , Montbonnot St. Martin , Santanu Chaudhury, Appearance based Generic Object Recognition, Proc. of ICVGIP 2000 - Indian Conference on Computer Vision, Graphics and Image Processing

[15] Jong T. Lee, M. S. Ryoo1, and J. K. Aggarwal, “View Independent Recognition of Human-Vehicle Interactions using 3-D Models”

[16] Roth, P.M. & Winter, M. (2008) Survey of Appearance-based Methods for Object Recognition, Technical Report ICG-TR-01/08, Graz University of Technology, Institute for Computer Graphics and Vision

[17] James H. Niblock, Jian-Xun Peng, Karen R. McMenemy and George W. Irwin, Autonomous Model-Based Object Identification & Camera Position Estimation with Application to Airport Lighting Quality Control, VISAPP 2008 - International Conference on Computer Vision Theory and Applications.

[18] Li, B., Chellappa, R., Zheng, Q., and Der, S., "Model-based temporal object verification using video", IEEE Transactions on Image Processing, vol. 10, no 6, pp. 897–908, 2001.

[19] Pope, A. R.: Model-Based Object Recognition - A Survey of Recent Research. Technical Report TR-94-04, University of British Columbia (1994).

[20] H. Noor, S. Noor, S.H. Mirza, "Using Video for Multi-view Object Categorization in Security Systems" SPIE Defense, Security and Sensing Symposium, April 13 – 17, 2009, Orlando, Florida, USA.

[21] Pope, A. R.: Model-Based Object Recognition - A Survey of Recent Research. Technical Report TR-94-04, University of British Columbia (1994).

[22] Ian T. Jolliffe. Principal Component Analysis. Springer, 2002.

[23] Aapo HyvÄarinen, Juha Karhunen, and Erkki Oja. Independent Component Analysis. John Wiley & Sons, 2001.

[24] Danijel Sko·caj. Robust subspace approaches to visual learning and recognition. PhD thesis, University of Ljubljana, Faculty of Computer and Information Science, 2003.

[25] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski. Face recognition by independent component analysis. IEEETrans. on Neural Networks, 13(6):1450-64, 2002.

[26] A. Toshev, A. Makadia, and K. Daniilidis. Shape-based object recognition in videos using 3d synthetic object models. In Proc. of CVPR, pages 288–295, 2009.

[27] Thomas Martel, Mike Negin and John Freyhof, Novel Image Segmentation as a Method for Conducting Forensic Analysis of Video Imagery, IEEE Conference on Technologies for Homeland Security, Boston, Massachusetts, APRIL 26, 2005

[28] M. Balcells, D. DeMenthon, and D. Doermann, "An appearance-based approach for consistent labeling of humans and objects in video, pattern and application", *Pattern Analysis & Application*, vol. 7, no. 4, pp. 373-385, Dec. 2004.

[29] Rukun Hu, Shuai Shao, Ping Guo, Investigating Visual Feature Extraction Methods for Image Annotation, Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA - October 2009

[30] http://publicintelligence.info/TSAhighwaysthreat.pdf, Accessed on April 12th, 2011.

[31] https://www.ihssnc.org/portals/0/Building_on_Clues_Strom.pdf, Accessed on April 12th, 2011.

[32] Teal M, Ellis T. Spatial-Temporal Reasoning on Object Motion. In Proceedings of the British Machine Vision Conference. BMVA Press, 1996 pp. 465-474.

[33] Guesgen, H., Marsland, S.: Spatio-temporal reasoning and context awareness. Handbook of Ambient Intelligence and Smart Environments (2010) 609-634

[34] Allen J (1983) Maintaining knowledge about temporal intervals. Communications of the ACM 26:832–843

[35] Güsgen, H.W.: Spatial reasoning based on Allen's temporal logic. TR-89-049, International Computer Science Institute, Berkeley 1989.

[36] R.Wetprasit, A. Sattar, and L. Khatib. Reasoningwith multi-point events. In LectureNotes in Artificial Intelligence 1081; Advances in AI, Proceedings of the eleventh biennial conference on Artificial Intelligence, pages 26–40, Toronto, Ontario, Ca, 1996. Springer. The extended abstract published in Proceedings of TIME-96 workshop, pages 36-38, Florida, 1996. IEEE.

[37] Anthony G. Cohn, Brandon Bennett, John Gooday, Micholas Mark Gotts: Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. GeoInformatica, 1, 275–316, 1997.

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

[39] Vinayak Elangovan and Amir H. Shirkhodaie, Adaptive Characterization, Tracking, and Semantic Labeling of Human-Vehicle Interactions Via Multi-modality Data Fusion Techniques, SPIE Defense, Security and Sensing, April 2011, Orlando, Florida.

[40] BAI, X., LI, Q. N., LATECKI, L. J., LIU, W. Y., AND TU, Z. W. 2009. Shape band: A deformable object detection approach. In Proc. of CVPR, 1335–1342.

[41] Cheng, M.M., Zhang, F.L., Mitra, N.J., Huang, X., Hu, S.M.: Repfinder: Finding approximately repeated scene elements for image editing. ACM ToG (Proc. SIGGRAPH) 29 (2010)

[42] 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.

[43] 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.

[44] K. Karmann and A. Brandt, "Moving object recognition using an adaptive background memory," in Time-Varying Image Processing and Moving Object Recognition, V. Cappellini, Ed. Amsterdam, The Netherlands: Elsevier, 1990, vol. 2.

[45] Y. Huang, T.S. Huang, and H. Niemann, "Segmentation-based object tracking using image warping and kalman filtering," in IEEE Signal Processing Society 2002 International Conference on Image Processing, Rochester, New York, 2002.

[46] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," IEEE Transactions, Pattern Analysis and Machine Intelligence, vol. 23, pp. 349–361, Apr. 2001.

[47] Y. Wu and T. S. Huang, "A co-inference approach to robust visual tracking," in Proceedings International Conference in Computer Vision, vol. II, 2001, pp. 26–33.

[48] B. Schiele, "Vodel-free tracking of cars and people based on color regions," in Proceedings of IEEE International Workshop Performance Evaluation of Tracking and Surveillance, Grenoble, France, 2000, pp. 61–71.

[49] I. A. Karaulova, P. M. Hall, and A. D. Marshall, "A hierarchical model of dynamics for tracking people with a single video camera," in Proc.British Machine Vision Conf., 2000, pp. 262–352.

[50] T. Zhao, T. S. Wang, and H. Y. Shum, "Learning a highly structured motion model for 3D human tracking," in Proc. Asian Conf. Computer Vision, Melbourne, Australia, 2002, pp. 144–149.

[51] M. Marszalek, I. Laptev, and C. Schmid. Actions in context. In CVPR, 2009.

[52] S. W. Joo and R. Chellappa. Attribute grammar-based event recognition and anomaly detection. In CVPRW, 2006.

[53] Yanco, H.A. and J.L. Drury (2002). "A taxonomy for human-robot interaction." In Proceedings of the AAAI Fall Symposium on Human-Robot Interaction, November 2002, Falmouth, MA. AAAI Technical Report FS-02-03, 111–119.

[54] H.A. Yanco, J. Drury, Classifying Human–Robot Interaction: An Updated Taxonomy, IEEE Conference on SMC, 2004.

[55] U. Akdemir, P. Turaga, and R. Chellappa, "An ontology based approach for activity recognition from video," Proc. of ACMMM, pp. 709–712, 2008.

[56] L. Xin and T. Tan, "Ontology-based hierarchical conceptual model for semantic representation of events in dynamic scenes," Proc. of IEEE PETS, pp. 57–64, 2005.

[57] J. SanMiguel, J. Martinez, and A. Garcia, "An ontology for event detection and its application in surveillance video," Proceedings of IEEE AVSS, pp. 220-225, 2009.

[58] B. Vrusias, D. Makris, J. Renno, "A framework for ontology enriched semantic annotation of cctv video" Proc. of IEEE WIAMIS, pp. 5–8, 2007.

[59] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, P. V.-K., and M. Strintzis, "Knowledge-assisted semantic video object detection," IEEE TCSVT, 15(10):1210–1224, 2005.

[60] A. Hakeem and M. Shah, "Ontology and taxonomy collaborated framework for meeting classification," Proc. of the IEEE ICPR, pp. 219–222, 2004.