Exploring Object Segmentation Methods in Visual Surveillance for Human Activity Recognition

Sandipkumar M Vaniya, PhD Schalor
Department of Computer Science & Engineering,
Sathyabama University
Chennai, India
s.vaniya@gmail.com

Dr. B. Bharathi, Professor
Department of Computer Science & Engineering
Sathyabama University
Chennai, India
bharathi.cse@sathyabamauniversity.ac.in

Abstract— **Human activity recognition in video camera is the prime research topic in computer vision and machine learning since last many years. Visual cameras have been used in public and private place like railway station, shopping malls, airport, offices, schools and university, etc. to recognize threat in the scene. The automated visual surveillance system will help to catch suspect in the scene, person identification in distance and re-identification, traffic management, sports, human computer interface, etc. Generally speaking, the human activity recognition in visual surveillance divided into following stages: object (human or vehicle) segmentation, feature extraction, object classification, object tracking, and activity recognition. Hence, the robust and object segmentation method in video camera is a very important phase because the rest of methods are strongly rely on it. In this paper, we have studied the various methods and/or algorithms of object segmentation (human). We will also discuss the strength and weakness of algorithms, complexities in activity understanding and identify the possible future research challenges.**

Keywords—visual surveillance; object segmentation; human activity recognition; background model; multi-camera

I. INTRODUCTION

Visual surveillance system is used to monitor an environment (or scene) with multiple video cameras. To understand the object behavior in real-time environment requires the development of computer vision algorithms to detect, locate and track the interested objects as it moves through the environment. Intelligent visual surveillance system is directed on automatically identifying event of interest, alert police when any threat found and make entire surveillance system automatically. For example, surveillance systems play a major role in health monitoring for elderly aging people at home. Also it's helpful for observing human behavior and identifies illegal actions and events in public and private places. In complex and cluttered scene with moderate number of moving objects (e.g. 10-15) the problem of detecting and tracking the objects is significantly complicated by occlusion, where an object may be partially occluded or totally disappear from camera view. The basic philosophy of object segmentation is to divide the interested objects in unlabeled videos from the background scene automatically [1-3].

Object segmentation is the first step in automatic understanding human activities in visual surveillance because the rest of the processing framework activities, such as feature extraction, object classifications, object tracking and activity understanding are highly depended on it. Object Segmentation is the key concept in computer vision and used in many real type of applications such as traffic monitoring, gender identification, anomaly detection, etc. The process for the automated multiple camera video surveillance is as given in Fig. 1. The videos are captures by multiple video cameras. Although, there are different types of cameras are available in the market for the research purpose. But CCD cameras, night vision cameras and thermal cameras are widely used for research purpose. After taking input of video stream, the next stage is object segmentation. Object segmentation is required to build surrounding model and compute motion energy if cameras and background are continuously moving. This stage is very important because the rest of the phases are heavily depended on it. There is no such a general purpose algorithm which can work in every surrounding model. Depending upon the application domain and requirement of background scene, appropriate algorithm can apply. After object segmentation, object classification is another research domain in computer vision.

Many object classification methods have been presented since last two decades, such as shape-base, feature-base, contour-base, etc. The selection of object classification methods are largely depended on nature of application and environment conditions. Object tracking is the next phase after classification of objects. The idea behind object tracking is to track certain objects (human or vehicle) from the videos continuously. Hence, the designing robust object tracking method is also challenging task for computer vision engineers and scientist. Since last two decades, the various object tracking methods have presented. For example, region-base, shape-base, contour-base etc. Once the objects have tracked, the human activity understanding (detection and recognition) comes into the picture.

This paper is discussed the various object segmentation method along with strengths and weaknesses. The section II is described the various application and needs of visual surveillance system for public safety and security. The details of object (human) segmentation methods and some remarkable work has given in section III. However, it is impossible to

describe all state-of-the-art object segmentation methods but we have tried to present some popular methods. The last section is presented further research problems and difficulties in human activity recognition along with the summary of paper.

II. APPLICATIONS

1) Behavior Biometric: Biometric is the technique to recognize human automatically based upon physical characteristics (e.g. face, fingerprints, retina, iris). This is a very powerful techniques but the system is required to interact with person to collect physical data. Therefore, these type of the system called physical biometric and strongly depended on subject (human) interaction. In case of behavior biometrics, the physical interaction is not required. The system will measure the unique features of persons without physical interaction. The most popular research in behavior biometric is human gait recognition [4] in video camera.

2) Video Analytics: The basic idea behind the video analytics is to retrieve the interested information from the videos, such as finding a specific clips in videos. It is also called as content base image and/or video retrieval [5-6]. The video cameras continuously monitor the scene and identify the presence of human or vehicle and recognize their activities and interaction with other objects. For example, there is a system which automatically recognize vehicle number plate at toll collection booth. Another example is to recognize the person who is trying to move unrestricted area or region.

3) Public Safety, Security and Surveillance: Public safety and security from the terrorists or criminals is the important aspect for any nation. Safety and security can be achieve through the close observation by video cameras and understand their behavior and interactions automatically. Single camera is not enough to understand the entire scene due to limited in view and angles. Hence, the network of cameras [7] are most useful way to cover large view. Pedestrian detection [8] is the one of the active research topic in this area and many research contribution is done since last ten years. The second important aspect is to detect and recognize suspicious behavior [9]. It is very difficult to say that person in the scene is suspicious or not because it depends on application and nature of environment. In fall detection, the lying on stair and lying on bed, both consider as fall detection but context are different. Therefore, abnormal or violent behavior recognition [10-13] is also challenging task for researcher.

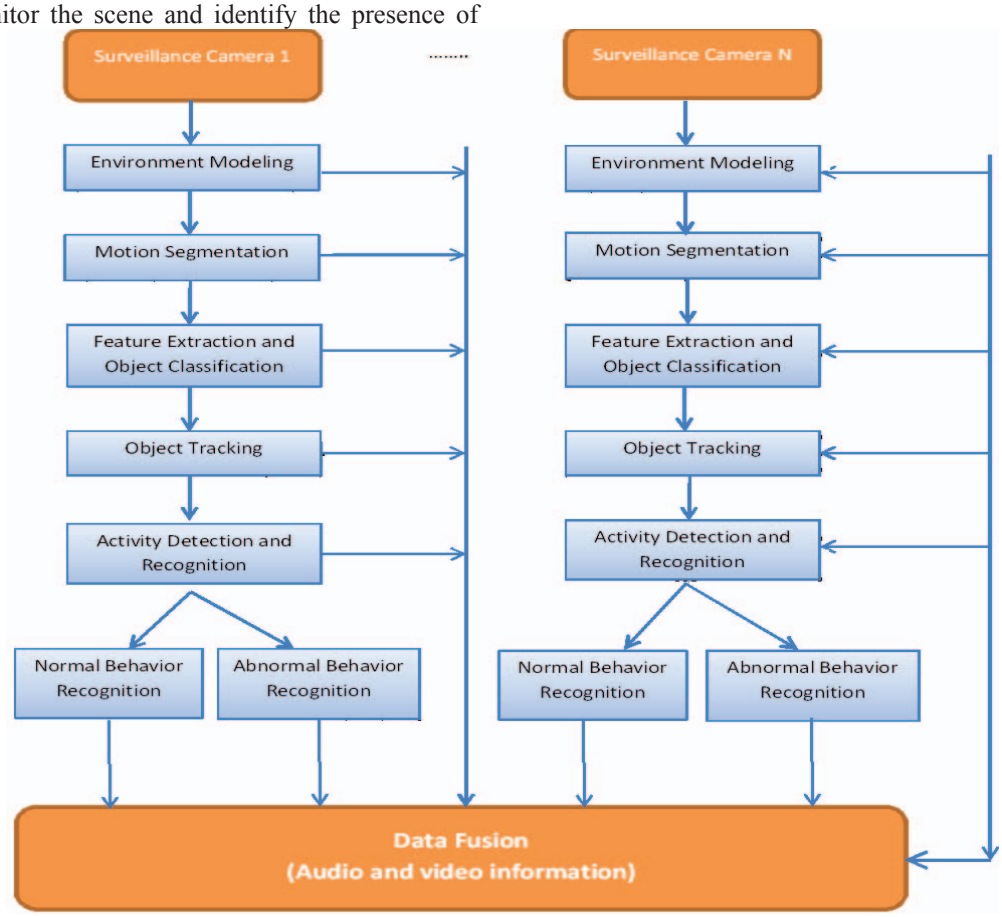

Fig. 1. Process for the automated multiple camera video surveillance [44]

4) Healthcare Surveillance System: The motive behind the healthcare surveillance system is to understand and continuously tracking patient's activities, certain medical equipment and medical staffs. Healthcare surveillance is used to detect and identify patient and to track their movement in video camera. Healthcare surveillance raised three challenges

namely identification, tracking and storing information. Identification is related to who is allow in which areas of patient's room. Tracking plays a vital role to locate patient, medical staff and medical equipment. Storing the patient's information like x-ray, medical prescriptions, blood reports, etc. and fast indexing and searching are the most challenging research in healthcare surveillance. Some research direction and research contribution related to healthcare surveillance is given [14-24].

5) ***Casino and Gaming:*** Another popular application of computer vision is in gaming, casino and animation field. Human activity can also applicable to recognize sport activities [25] and dance activities [26]. In gaming [27], computer can learn the activities from the human and after that computer will create an avatar which will play on behalf of human and this makes life full of entertainments. In casinos, it is important to monitor the continuous flow of people, gambling table and crowd.

III. OBJECT SEGEMENTATION METHODS

The first step for designing visual surveillance system is object segmentation. Object segmentation is the process of taking interested object out from the images. There are two flavors of segmentation methods, one is static camera and another is dynamic (moving) camera.

1) Static Camera Segmentation: This types of camera are fixed on particular location like on wall or ceiling. Hence, the angle and viewpoint cannot be changed. Following are some approaches for static camera segmentation methods.

a) Background subtraction: This is one of the simple, popular and efficient method since late 1970s [28]. The concept of background subtraction method is to subtract the current picture or image from the given set of the images. The process can be done either pixel wise approach or entire region of interest (region-based) [29]. Some of the research paper, it is also known as foreground detection methods. Generally, we consider the foreground objects as human or vehicles. These foreground object needs to be extracted from the video footage along with reducing noise and other low level processing stage. Background subtraction method is enough to detect moving object if the video camera remain in fixed location and background remain idle. To implement background subtraction method programmatically, two important steps need to consider. First, initialization of background image in which initial background image is computed which called "Background model" and second is continuous updating of background image in order to cope up with changes. Background subtraction is generates binary image (also called foreground mask) by making difference of current video frame and background. This method is very simple and does not required high computational cost due to fixed in nature. Hence, this method is not suitable to real time environment because real time system come across the many challenges like changing of lighting condition frequently, illumination changes by rain or winter season, etc. Depending upon the system (indoor or outdoor), the algorithm changes are required to meet challenges. Since last many years, different modification and improvements have been carried out by researchers and several doctoral thesis published in the area of background subtraction methods [30]. Seki [31] proposes a method to handle dynamic scene and successfully discriminate object with ghost and shadow using statistical computation. The background subtraction computation cost is low if background is fixed but shows poor result in continuous updating model.

b) Gaussian model: Gaussian process is very important theory in probability and statistics. It is statistical model in which observation of data can be done in continuous domain. It is widely used for modeling dependent data observed over the time and space or both. It represent the image in the form of color value or grayscale vale over the time [30]. The limitation is that the real time image or videos contains moving objects, ghost, shadows, etc. Hence, the single Gaussian model is not enough to represent image. To overcome these limitations, the Gaussian Mixture Model (GMM) can represent image pixel value by sum of Gaussian density. It can easily represent the background which is consist of shadow and motion of objects [33]. The learning process of GMM is done using Expectation Maximization algorithm which compute the probability of pixel value in the image. The high computational cost is required for EM algorithm. In [34], k-means clustering algorithm have been used to reduce computational cost. The full explanation of Gaussian theory is out of scope of this paper.

C) Statistical model: Statistical model is the mathematical model which simplify and generate approximate data and predict future data from the approximations. To construct more complicated background model, statistical method compute each of the pixel value or group of the pixels and simultaneously updating background model during processes. The four different ways of pixel value representation is given in [35] and classify the pixel value depending upon brightness and chromaticity functions. Statistical approach is gaining popularity because it is more robust as compare to GMM and can work in changing lighting condition and shadows.

d) Segmentation by tracking: Segmentation by tracking is another way to construct the background model. GMM, statistical model and background subtraction methods are low-level segmentation methods because it only consider the pixel-value in images. Rather than considering pixel information, Brendal [36] has presented a method by tracking interested region in the frame with circular dynamic-time warping algorithm. The basic idea of DTW algorithms is to compute similarity between two time series sequences. For example, if one person is running fast than other person, DTW algorithm can applied to the temporal sequences. The some applications of DTW algorithm is speech recognition, signature recognition, speaker recognition, etc. In [37], another improved method by tracking spatial-color Gaussian mixture model is presented.

2) Dynamic (Moving) Camera Segmentation: As compare to static camera segmentation, the moving cameras are not fixed in particular angle. For example, the cameras install in running car, moving arms, etc. The moving cameras are also a hand held camera where the person moves. The basic difference is that static cameras are idle means stationary, the

frame of the camera is not moving but in case of moving camera, the frame of camera is moving which includes pan, tilt, zoom etc. Hence, to make object segmentation in moving camera is more challenging compare to static because motion of background and foreground objects are moving.

a) Temporal difference: The temporal difference is a very common method for moving camera segmentation. It is also called as reinforcement learning method. The challenge is to estimate camera motion as well as motion of objects in the scene. The temporal difference can detect motion of object by taking the difference of two consecutive frame. In this theory, the motion of camera need to be first estimate. Object segmentation in real time pan-tilt active camera using temporal difference method was proposed in [38]. Image mapping and morphological operators are applied to get the exact image from the different viewpoints and reduce the noise in image. Another improvement is presented in [39] to estimate pan-tilt camera using edge feature in two consecutive video frame. Temporal difference method is required to compute camera motion first which also raise the issue of noise in video frame. The computational cost of temporal difference method is less.

TABLE I. BRIEF SUMMARY OF OBJECT SEGMENTATION METHODS

Segmentation Method	**Static/Moving Camera**	**Observation**	**Remarks**
Simple Background Subtraction	Static	Sensitive to illumination changes	Point-based Method
Background model with Single Gaussian	Static	Need to specify number of Gaussian.	
Dynamic Background	Static	High computational cost	
Background Subtraction with GMM with EM algorithm	Static	High computational cost	
Background model with statistical parameters	Static	High computational cost.	
Segmentation by tracking -CDTW Algorithm	Static	High computational cost.	Region-based Method
Temporal Difference	Moving	Best in dynamic scene but poor in extracting all pixels	
Optical Flow	Moving	Sensitive to noise, High computational cost, Specialized Hardware requires	

b) Optical flow:

Optical flow is the method to measure the velocity (motion) of the objects. In computer vision, the optical flow method is used to quantify the motion of object in videos which is used for detection and tracking of objects. Optical flow can detect motion of object by applying flow vector of moving object with respect to temporal motion. In gait analysis [40], the method for extracting information relevant to articulated object is presented. Monotony operator have used for computing displacement vector field. The body parts of human body have tracked using contour-base tracking algorithm. Some remarkable work is also presented in [41-50] using MAP-MRF algorithm, Lucas-Kanade- Tomasi tracking algorithm. Optical flow method is very high in computational cost and also taking a long time to segment out the objects. For this reason, the optical flow methods requires specialized hardware (e.g. FPGA) to increase algorithm efficiency. We have presented the brief summary and latest survey on object tracking algorithms (see Table I).

IV. FUTURE RESEARCH

Development of viewpoint invariance method:

1) The recent surveillance systems are equipped with multiple cameras and very good in information storing capability. These multiple cameras will captures the videos or images from the multiple angle and it is very difficult to develop view invariant algorithms. The question is how we can fuse the information from the multiple cameras. Fusing information from the multiple cameras requires co-ordination among the devices in network which leads to distributed video surveillance system.

2) The occlusion handling is also key research challenge in human activity recognition. Theoretically, we can say that the occlusion can be handle using multiple cameras but the question is that how can we combine the multiple views of images into single image plane.

3) The background subtraction is very simple and most common method used since last four decades. Most of the surveillance systems have built on background subtraction philosophy which requires constrains on background model. For example, some algorithms will only work on enough lighting conditions, partly occluded scene, etc. There is a need to design more robust and fast object segmentation methods which can work on dynamic scene because the real-time surveillance system is heavily depended on fast object detection, tracking and recognition. To do so, many issues like occlusion handling, viewpoint adjustment, fusion of information, noise in videos, understanding the context (what, when, where, etc.), posture variation, will required to solve at present. However, some level of improvement have done so far but not at significant level.

4) Most of the visual surveillance system can recognize single human activity. There is also a need to design a system which can understand crowd behavior. Crowd behavior understanding is an important research problem in computer vision. It deals with crowd management, automatic detection and recognition of suspicious activities, automatic controlling the flow of people in some public places, etc.

5) Abnormal or suspicious activity recognition is another research domain in computer vision. Detection abnormal behavior in dynamic scene is heavily depended on scene

context. For example, lying on sofa is very much different than lying on stair. Many commercial system have developed but majority of the system deals with constrained scene. There is a need to design robust algorithm for clutter and unconstrained scene which requires the fast learning process.

V. CONCLUSION

This paper describes the current-state-of-the art methods in object segment segmentation in visual surveillance for activity recognition. The strengths and weaknesses of algorithms is presented in the context of human activity recognition in static and dynamic (moving) camera also presented. Finally, we have shown the further research domains in object segmentation in visual surveillance.

REFERENCES

[1] Y. J. Lee, J. Kim, and K. Grauman. Key-segments for video object segmentation. In ICCV, 2011.

[2] P. Ochs and T. Brox. Higher order motion models and spectral clustering. In CVPR, 2012.

[3] T. Wang and J. Collomosse. Probabilistic motion diffusion of labeling priors for coherent video segmentation. IEEE Trans. Multimedia, 2012.

[4] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The HumanID gait challenge problem: Data sets, performance, and analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 2, pp. 162–177, 2005.

[5] Y. Rui, T. S. Huang, and S. F. Chang, "Image retrieval: current techniques, promising directions and open issues," Journal of Visual Communication and Image Representation, vol. 10, no. 4, pp. 39–62,1999.

[6] S. F. Chang, "The holy grail of content-based media analysis," IEEE Transactions on Multimedia, vol. 9, no. 2, pp. 6–10, 2002.

[7] Nakazawa, A.; Kato, H.; Inokuchi, S. Human Tracking Using Distributed Vision Systems. In Proceedings of IEEE Fourteenth International Conference on Pattern Recognition, Brisbane, Qld., Australia, 20 August 1998; Volume 1, pp. 593–596.

[8] Bodor, R.; Jackson, B.; Papanikolopoulos, N. Vision-based Human Tracking and Activity Recognition. In Proceedings of the 11th Mediterranean Conference on Control and Automation, Rhodes, Greece, 18–20 June 2003; Volume 1, pp. 18–20.

[9] Fiaz, M.K.; Ijaz, B. Vision based Human Activity Tracking using Artificial Neural Networks. In Proceedings of IEEE International Conference on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, 15–17 June 2010; pp. 1–5.

[10] Sempena, S.; Maulidevi, N.U.; Aryan P.R. Human Action Recognition Using Dynamic Time Warping. In IEEE International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, 17–19 July 2011; pp. 1–5.

[11] Ribeiro, P.C.; Santos-Victor, J. Human Activity Recognition from Video: Modeling, Feature Selection and Classification Architecture. In Proceedings of the International Workshop on Human Activity Recognition and Modelling (HAREM), Oxford, UK, 9 September 2005; Volume 1, pp. 61–70.

[12] Niu, W.; Long, J.; Han, D.; Wang, Y. Human Activity Detection and Recognition for Video Surveillance. In Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 27–30 June 2004; Volume 1, pp. 719–722.

[13] Moënne-Loccoz, N.; Brémond, F.; Thonnat, M. Recurrent Bayesian network for the recognition of human behaviors from video. In Proceedings of the 3rd International Conference on Computer Vision Systems (ICVS), Graz, Austria, 1–3 April 2003; pp. 68–77.

[14] Kuo, Y.; Lee, J.; Chung, P. A visual context-awareness-based sleeping-respiration measurement system. IEEE Trans. Inf. Technol. Biomed. 2010, 14, 255–265.

[15] Gao, J.; Hauptmann, A.G.; Bharucha, A.; Wactlar, H.D. Dining Activity Analysis Using a Hidden Markov Model. In Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR), Cambridge, UK, 23–26 August 2004; Volume 2, pp. 915–918.

[16] Huynh, H.H.; Meunier, J.; Sequeira, J.; Daniel, M. Real time detection, tracking and recognition of medication intake. World Acad. Sci. Eng. Technol. 2009, 60, 280–287.

[17] Foroughi, H.; Rezvanian, A.; Paziraee, A. Robust Fall Detection Using Human Shape and Multi-Class Support Vector Machine. In Proceedings of the IEEE Sixth Indian Conference on Computer Vision, Graphics & Image Processing (ICVGIP), Bhubaneswar, India, 16–19 December 2008; pp. 413–420.

[18] Foroughi, H.; Aski, B.S.; Pourreza, H. Intelligent Video Surveillance for Monitoring Fall Detection of Elderly in Home Environments. In Proceedings of the IEEE 11th International Conference on Computer and Information Technology (ICCIT), Khulna, Bangladesh, 24–27 December 2008; pp. 219–224.

[19] Foroughi, H.; Yazdi, H.S.; Pourreza, H.; Javidi, M. An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Multi-class Support Vector Machine. In Proceedings of IEEE 4th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 28–30 August 2008; pp. 83–90.

[20] Foroughi, H.; Naseri, A; Saberi, A.; Yazdi, H.S. An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Neural Network. In Procedings of IEEE 9th International Conference on Signal Processing (ICSP), Beijing, China, 26–29 October 2008; pp. 1499–1503.

[21] Lühr, S.; Venkatesh, S.; West, G.; Bui, H.H. Explicit state duration HMM for abnormality detection in sequences of human activity. PRICAI 2004: Trends Artif. Intell. 2004, 3157, 983–984.

[22] Duong, T.V.; Phung D.Q.; Bui, H.H.; Venkatesh, S. Human Behavior Recognition with Generic Exponential Family Duration Modeling in the Hidden Semi-Markov Model. In Proceedings of IEEE 18th International Conference on Pattern Recognition (ICPR), Hong Kong, China, 20–24 August 2006; Volume 3, pp. 202–207.

[23] Liu, C.; Chung, P.; Chung, Y.; Thonnat, M. Understanding of human behaviors from videos in nursing care monitoring systems. J. High Speed Netw. 2007, 16, 91–103.

[24] Li, Y.; Miaou, S.; Hung, C.K.; Sese, J.T. A Gait Analysis System Using two Cameras with Orthogonal View. In Proceedings of IEEE International Conference on Multimedia Technology (ICMT), Hangzhou, China, 26–28 July 2011; pp. 2841–2844.

[25] Luo, Y.; Wu, T.; Hwang, J. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks. Comput. Vis. Image Underst. 2003, 92, 196–216.

[26] Shechtman, E.; Irani, M. Space-time Behavior Based Correlation. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 405–412.

[27] Huo, F.; Hendriks, E.; Paclik, P.; Oomes, A.H.J. Markerless Human Motion Capture and Pose Recognition. In Proceedings of the 10th IEEE Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), London, UK, 6–8 May 2009; pp. 13–16.

[28] H. Kautz, "A fonnal theory of plan recognition," Ph.D. dissertation, University of Rochester, 1987.

[29] Piccardi, M. Background Subtraction Techniques: A Review. In Proceedings of IEEE International Conference on Systems, Man and Cybernetics, The Hague, The Netherlands, 10–13 October 2004; Volume 4, pp. 3099–3104.

[30] Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfinder: Real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. 1997, 19, 780–785.

[31] Seki, M.; Fujiwara, H.; Sumi, K. A Robust Background Subtraction Method for Changing Background. In Proceedings of Fifth IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, USA, 4–6 December 2000; pp. 207–213.

[32] R. Jain and H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, pp. 206-214, 1979.

[33] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in Thirteenth Corif. on Uncertainty in Artificial Intelligence, 1997, pp. 175-81.

[34] Permuter, H.; Francos, J.; Jermyn, I. A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recogn. 2006, 39, 695–706.

[35] Horprasert, T.; Harwood, D.; Davis, L.S. A statistical approach for real-time robust background subtraction and shadow detection. IEEE ICCV 1999, 99, 1–19.

[36] Brendel, W.; Todorovic, S. Video Object Segmentation by Tracking Regions. In proceedings of IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 833–840.

[37] Yu, T.; Zhang, C.; Cohen, M.; Rui, Y.; Wu, Y. Monocular Video Foreground/Background Segmentation by Tracking Spatial-color Gaussian Mixture Models. In Proceedings of IEEE Workshop on Motion and Video Computing (WMVC), Austin, TX, USA, 23–24 February 2007; p. 5

[38] Murray, D.; Basu, A. Motion tracking with an active camera. IEEE Trans. Pattern Anal. Mach. Intell. 1994, 16, 449–459.

[39] Kim, K.K.; Cho, S.H.; Kim, H.J.; Lee, J.Y. Detecting and Tracking Moving Object Using an Active Camera. In Proceedings of IEEE 7th International Conference on Advanced Communication Technology (ICACT), Phoenix Park, Dublin, Ireland, 21–23 February 2005; Volume 2, pp. 817–820.

[40] D. Meyer, et al., "Model based extraction of articulated objects in image sequences for gait analysis," in Image Processing, 1997. Proceedings., International Conference on, 1997, pp. 78-81.

[41] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, pp. 1778-1792,2005.

[42] Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with An Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, B.C., Canada, 24–28 August 1981.

[43] Shi, J.; Tomasi, C. Good Features to Track. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 21–23 June 1994; pp. 593–600.

[44] F.Bremond, M. Thonnat, and M. Zuniga, "Video-understanding framework for automatic behavior recognition," Behavior Research Methods, Vol. 30, No. 3, pp. 416-426, 2006

[45] Wang, L.; Hu, W.; Tan, T. Recent developments in human motion analysis. Pattern Recognit. 2003, 36, 585–601.

[46] Amer, A.; Regazzoni, C. Introduction to the special issue on video object processing for surveillance applications. Real Time Imaging 2005, 11, 167–171.

[47] O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 865-878, Nov. 2012.

[48] J. Varadarajan and J. Odobez, "Topic models for scene analysis and abnormality detection," in Proc. IEEE 12th Int. Conf. Comput. Vision Workshops, Sep. 27–Oct. 4, 2009, pp. 1338–1345.

[49] X. Zhang, H. Liu, Y. Gao, and D. H. Hu, "Detecting abnormal events via hierarchical Dirichlet processes," in Proc. 13th Pacific-Asia Conf.Knowledge Discovery Data Mining, Apr.27–30, 2009, pp. 278–289.

[50] A. Chianese, V. Moscato, and A. Picariello, "Detecting abnormal activities in video sequences," in Proc. Ambi-Sys workshop on Ambient media delivery and interactive television, pp. 1–8, 2008.

[51] Y. Chen, G. Liang, K. L. Ka, and Y. Xu,"Abnormal behavior detection by multi-SVM-based Bayesian network," in Proc. Int. Conf. Inf. Acquisition, Jul. 9–11, 2007, pp. 298–303.

[52] Y. Wang, K. Huang, and T. Tan, "Abnormal activity recognition in office based on R transform," in Proc. IEEE Int. Conf. Image Process., 2007, pp. I-341–I-344.

[53] C.-K. Lee, M.-F. Ho, W.-S. Wen, and C.-L. Huang, "Abnormal event detection in video using N-cut clustering," in Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process., 2006, pp. 407–410.

[54] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in Proc. IEEE Conf. Comput. Vision Pattern Recog., 2009, pp. 1446–1453.