Mixture of Gaussian Based Background Modelling for Crowd Tracking Using Multiple Cameras

Mohamed Abul Hassan[#1], Aamir Saeed Malik[#2], Walter Nicolas[#3], Ibrahima Faye[#4], Muhammad Tariq Mahmood[*5]

Centre for Intelligent Signal and Imaging Research (CISIR)

Department of Electrical & Electronic Engineering[1, 2, 3], Department of Fundamental and Applied Sciences[3], Universiti Teknologi PETRONAS[#], Malaysia

[2]aamir_saeed@petronas.com.my

Applied Soft Computing Lab[5], School of Computer Science and Engineering, Korea University of Technology and Education[*], Republic of Korea

***Abstract*— Visual surveillance system for tracking crowd using multiple cameras at dynamic backgrounds faces many challenges such as illumination variance, occultation, low spatial temporal resolution, sleeping person, shadows and camera noise. In this paper we address the issue of gradual and sudden illumination variance caused by movement of the sun and the clouds. We evaluate Mixture of Gaussian method and background modelling method for extracting foreground from the background for crowd related data base. We have evaluated the performance of the background model for sparse and dense crowds to evaluate the accuracy and efficiency of the model subjectively for crowd analytics based scenarios.**

I. Introduction

Visual surveillance system has become one of the most attractive research areas in the field of computer vision. The goal of video surveillance is to extract information from video footages collected by the video surveillance camera. The final objective is to enable the surveillance camera to automatically detect, recognize and track objects of interest and to understand and analyze their activities [1]. The applications of video surveillance extend; to both public and private sectors such as home land security, crime prevention, traffic control, of monitoring patients and children at home, where these applications are used at indoor and outdoor settings such as train stations, airports, malls, offices, parking lots, highways, hospitals and many other public locations [2].

The availability of low cost sensors and processors is one of the main reasons for the increasing interest in the field of video surveillance technology. The need for safety and security at public sectors is another reason for the intensive research in the video surveillance systems. Furthermore, the availability of huge number of cameras, which are operated at the expense of a great deal of man power, also demands for researches to develop an efficient system that can extract information from the real time footage and automatically detect, recognize and track objects of interest, so as to understand and analyze their activities where the entire process is shown in the “Fig. 1” in the form of a block diagram [1].

In crowd tracking systems understanding the crowd behavior prior to developing the system is essential. As the surveillance systems employed from place to place would yield different behavior characteristics of the crowd [3]. A crowd is a group of individuals who follows motion, in a form of a Pattern to other individuals at different scenes. There are several common patterns followed by the crowd such as bottleneck, departure, lane, arch/ ring and blocking. Acquiring the prior knowledge of the patterns will enhance the level of the optimization of the system.

The video surveillance system faces many challenges and limitations based on its environmental conditions. Mainly a surveillance system can be distinguished by two conditions; indoor and outdoor. At indoor conditions, a surveillance camera persists of static background, where the background is less prone to changes, while at outdoor conditions, the dynamic background is much more prone to change during the course of the day [4-6]. However, general surveillance systems have challenges such as gradual and sudden change of illumination, intra camera occlusion, stationary people, waking people, low spatial and temporal resolution, camera perturbation and noisy videos that bring several limitations to the robustness of the system.

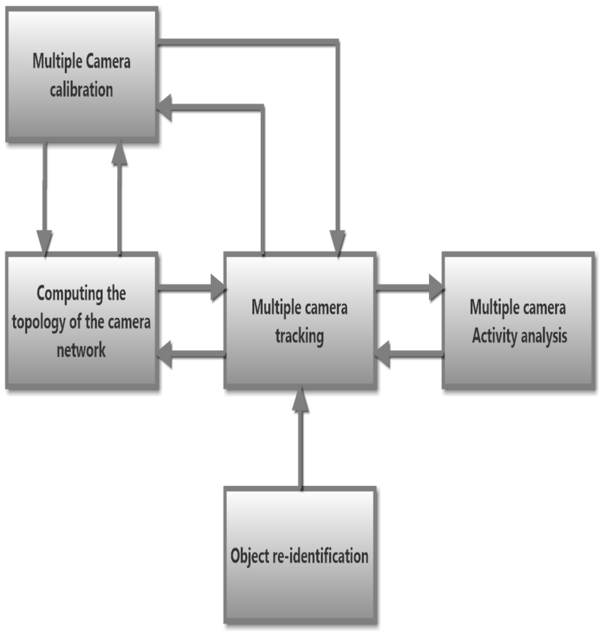

Fig. 1.Main processing units for multiple camera surveillance systems [1]

978-1-4799-4653-2/14/$31.00 © 2014 IEEE

In this paper we discuss the Mixture of Gaussian method (MOG) for the application in the discipline in crowd analytics which deals with behavior of crowds. Crowd analytics comprise of many surveillance applications such as crowd/objects detection, crowd tracking, crowd density estimation like motion estimation, direction and velocity estimation and counting etc. at real time. These applications are mostly deployed in complex dynamic environments.

Hence, these applications are often affected by gradual and sudden change of illumination, intra camera occlusion, and stationary people, waking people, low spatial and temporal resolution, camera perturbation and noisy videos which bring several limitations to the robustness of the system. The rest of the paper is organized as follows; section II provides an extensive study on the literature for the topic of video surveillance. Section III discusses the methodology which illustrates the algorithm and the mathematical model of MOG, while section IV presents the experimental results and discussion and finally section V concludes the paper.

II. Literature Review

An extensive study was carried out on the topic of video surveillance systems to provide a brief but rich description in this section; where the literature is organized based on the review on video surveillance systems, surveillance systems for tracking crowd, multiple camera tracking and background modeling of surveillance systems.

A. Video surveillance system

Video surveillance system is one of the most attractive fields in computer vision due to the fact that there are billions of cameras fixed at public and private sectors all over the world that are monitored by a large pool of man power. This is inefficient and thus, an intelligent system is in the requirement to be developed to provide analysis and protection to the subjects of the scene. Video surveillance systems are used for two applications; mainly crowd tracking and vehicle traffic surveillance. A video surveillance system frame work generally consists of background modeling, object segmentation, object classification, object tracking, behavior and activity analysis and object identification [1, 4].

Many surveillance systems have been developed to overcome the challenges faced by systems such as a robust background updating system for real time traffic surveillance and a sophisticated tracking algorithm for tracking humans under low resolution which are concentrated on background modeling, object segmentation and tracking. Detecting the amorphous and unstructured such as in objects behind fire, smoke and targets in deep turbulence using super resolution recovery by frames reconstruction systems focused on object segmentation and classification. Systems were developed for tracking, behavior and object identification using Trajectory-based activity analysis for visual surveillance, crowd tracking using activity-based semantic scene decomposition and Cross canonical correlation analysis for global Activity topology inference [2, 7].

B. Surveillance system for Crowd Tracking

Video surveillance systems for tracking crowds has been a point of interest for researchers as there are many challenges raised in tracking crowd and in which one of the most common and challenging issue is the illumination variance where the surveillance systems are always subjected to the gradual or sudden illumination variance. Moreover, the fact that the human being who is an object in a crowd consumes few pixels in the entire frame raises low resolution. While tracking crowds there are other issues such as occultation where objects are covered over by the other [8], then sleeping person or waking person who are classified as ghost objects and camera noise, are also some of the most redundant challenges faced by the surveillance system tracking.

Researchers over the past decade have developed many surveillance systems for overcoming these challenges using many methods. Template based matching methods using histogram and shape analysis of human body parts are used to overcome illumination variance and sleeping person detection, and background modeling techniques such as Mixture of Gaussians (MOG) and median filter is used to overcome illumination variance. Super resolution reconstruction, Observational model with intensity cues and practical filters are used to overcome issues in poor resolution. Many temporal differencing methods using Principal Component Analysis (PCA), pixel level multi models are used for shadow detection and overcoming camera noise [9].

C. Background Modeling of Surveillance Systems

Multiple camera tracking is one of the critical challenges in surveillance systems since each camera undergoes different illumination variance with respect to its view and is required to overcome the illumination variance and detects objects in its camera views, object re-identification in different camera views are another great challenge in multiple camera tracking especially when objects are supposed to be identified across blind areas, where there is a lack of coverage in the camera view and object detection through occlusion is another issue. However surveillance systems are designed for tracking objects through these issues using template matching cross canonical correlation and adoptive filtering [1, 2, 9].

Based on the above review of different surveillance systems for tracking crowds using single camera and multiple cameras it is understood that the need of an efficient background modeling algorithm is to extract the foreground from the background to overcome the issues of illumination variance, low resolution, sleeping person or waking person and camera noise. Over the years many background modeling algorithms are used such as MOG, Eigen background model, Median filter method, Kalman filter, Kernel density Estimation (KDE), wall flower, Temporal Differential, Hidden Macov Method (HMM), Bessian distribution and optical flow[8, 10-12]. Each of these methods is efficient from the other in its own manner in terms of speed, accuracy, memory requirement and nature of the background such as static and dynamic backgrounds which are usually indoor or outdoor. It is evident that the most significant challenge in this research is to develop a framework of background model which will extract the foreground for tracking crowds using multiple cameras though illumination variance with high computational efficiency and accuracy.

III. Methodology

Mixture of Gaussian (MOG) method is one of the most widely used background modeling algorithm. It is one of the dominant background modeling algorithms in line for its decent compromise among robustness in perilous situations and restraints [13, 14]. MOG functions are based on the probability density function of the Gaussians generated in the model. MOG background modeling algorithm is a pixel wise model which offers description of both background and foreground values by a likelihood of perceiving a certain pixel 'x' at a time 't' by means of mixture of Gaussians. The algorithm of MOG will be discussed using four main steps: initialization, modeling, similarity evaluation and updating background.

a. Initialization

Initially each pixel is characterized by intensity in the RGB color space and the probability of observing the current pixel is determined by the eq. (1) in a multi-dimensional scenario, where K is the amount of Gaussian spreading, $\omega_{i,t}$ is a weight related with i^{th} Gaussian on the time t with an average $\mu_{i,t}$ and standard deviation $\Sigma_{i,t}$, η is a Gaussian probability density function

$$p(x_t) = \sum_{i=1}^{k} \omega_{i,t} . \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \tag{1}$$

b. Modeling

Once the initialization process is completed, the foreground is detected and the parameters are updated using the criterion ratio $r_j=\omega_j/\sigma_j$ and the order of the Gaussians is determined following the ratio. This ordering depends on the background pixels which correspond to a high weight with a weak variance and a constant value for moving objects and the Gaussian are maintained on the threshold T in eq. (2)

$$B = \arg\min_b \left(\sum_{i=1}^{b} \omega_{i,t} > T\right) \tag{2}$$

c. Similarity Evaluation

When the background is subtracted based on the Gaussians, two scenarios occur,

Scenario 1: when a counterpart is found through one of the K Gaussians, the distribution is taken in place of background and the other by means of foreground.

Scenario 2: when there are no matches found with any of the K Gaussians the pixels are categorized in place of foreground.

d. Updating background

After the background and foreground are extracted successfully, and if the scenario 1 of similarity evaluation has taken place, the model is updated using the weight of Gaussians, mean and the standard deviation as shown in eq. (3) to (5), and for the scenario 2 least probable distribution is replaced with a new one, as shown in eq. (6) to (8).

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha \tag{3}$$

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1} \tag{4}$$

$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + (X_{t+1} - \mu_{i,t+1})(X_{t+1} - \mu_{i,t+1})^T \tag{5}$$

Where $\rho = \alpha . \eta(X_{t+1}, \mu_i, \Sigma_i)$

$$\omega_{k,t+1} = \text{Low prior weight} \tag{6}$$

$$\mu_{k,t+1} = X_{t+1} \tag{7}$$

$$\sigma^2_{k,t+1} = \text{Large initial variance} \tag{8}$$

IV. Experimental results

The experimentation phase was carried out subjectively where the model was implemented on MATLAB 2013a simulation platform in a Intel Core i7 processor with NVIDIA GeForce GT650 4GB Graphic card and a 4GB DDR3 RAM. The binary map of the background extraction was taken as a final result to determine the efficiency of the model. We used the PETS 2010 data set, which is specially developed for visual surveillance research on crowd behavior. It can be observed in "Fig. 2" to "Fig. 5" the background model for multiple camera view is obtained and evaluated for sparse and dense crowd scenarios and is evaluated for camera view 1 and 4,

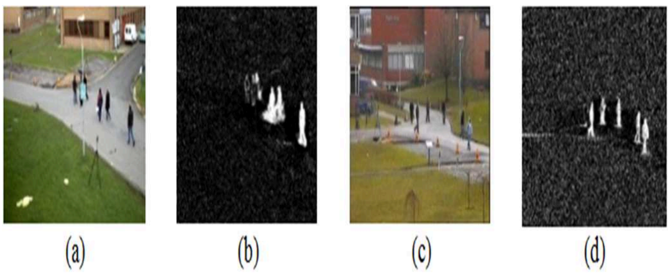

Fig. 2. Experimental results of background modeling using MOG for sparse crowd scenario, (a) camera view 1, (b) binary map of camera view 1, (c) camera view 4, (d) binary map of camera view 4.

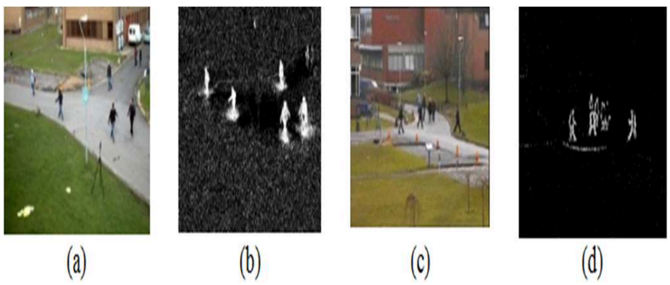

Fig. 3. Experimental results of background modeling using MOG for sparse crowd scenario, (a) camera view 1, (b) binary map of camera view 1, (c) camera view 4, (d) binary map of camera view

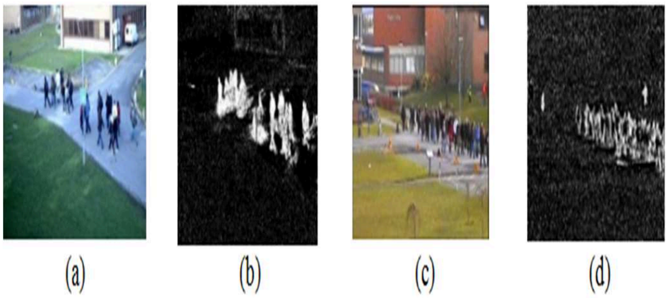

Fig. 4. Experimental results of background modeling using MOG for dense crowd scenario, (a) camera view 1, (b) binary map of camera view 1, (c) camera view 4, (d) binary map of camera view 4.

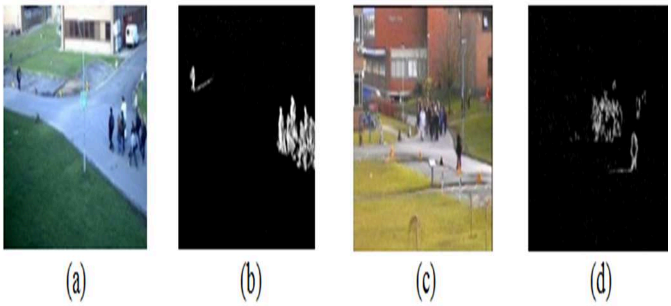

Fig. 5. Experimental results of background modeling using MOG for dense crowd scenario, (a) camera view 1, (b) binary map of camera view 1, (c) camera view 4, (d) binary map of camera view 4

From the experimental results, it can be witnessed that MOG adapted efficiently in dynamic environment, and was able to efficiently segregate the individual objects at sparse and dense scenarios. The model also was able to detect small objects from pixel range of 10×35 to 20×70, which provided the accuracy of segmentation of the model. Though the model was affected by sudden illumination which resulted in extracting false foreground mask and failed to differentiate the shadows of the foreground objects.

V. Conclusion

Visual surveillance for crowd analytics is an emerging research topic which has brought a great deal for interest of researches in the field of computer vision and pattern recognition. In this paper we discussed an overview of visual surveillance with multiple camera tracking. We investigated Mixture of Gaussian (MOG) method and re-implemented the model for evaluating its functionality on the PETS 2010 data set to determine the performance of MOG for sparse and dense crowd scenarios.

As such, we were able to evaluate Mixture of Gaussian method for crowd scenarios subjectively. The Model was robust under a dynamic environment and was efficient in object detection and segregation. However, the model suffered under sudden illumination variance and was unable to differentiate shadows of the foreground objects. In conclusion this paper provides an insight to the area of visual surveillance for the application of crowd analytics and provides evaluation on the performance of MOG method for crowd scenarios under dynamic environments.

Acknowledgment

This work is supported by PRGS grant (Ref: 1/12/TK02/UTP/02/05), Ministry of Education (MoE), Malaysia

References

[1] X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern Recognition Letters, vol. 34, pp. 3-19, 2013.

[2] O. J. Alper Yilmaz, Mubarak Shah, "Object Tracking: A Survey," ACM Comput. Surv, vol. 38, p. 45, Dec. 2006 2006.

[3] J. Ferryman and A. Ellis, "PETS2010: Dataset and Challenge," in Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, 2010, pp. 143-150.

[4] C. Chung-Hao, Y. Yi, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Camera Handoff with Adaptive Resource Management for Multi-camera Multi-target Surveillance," in Advanced Video and Signal Based Surveillance, 2008. AVSS '08. IEEE Fifth International Conference on, 2008, pp. 79-86.

[5] S. Denman, A. Bialkowski, C. Fookes, and S. Sridharan, "Determining operational measures from multi-camera surveillance systems using soft biometrics," in Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on, 2011, pp. 462-467.

[6] S. Jie, S. K. Zhou, and R. Chellappa, "Tracking Algorithm Using Background-Foreground Motion Models and Multiple Cues," in Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, 2005, pp. 233-236.

[7] M. Mueller, P. Karasev, I. Kolesov, and A. Tannenbaum, "A video analytics framework for amorphous and unstructured anomaly detection," in Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011, pp. 2945-2948.

[8] C. C. Loy, X. Tao, and G. Shaogang, "Multi-camera activity correlation analysis," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009, pp. 1988-1995.

[9] T. Bouwmans, "Recent Advanced Statistical Background Modeling for Foreground Detection - A Systematic Survey," 2011.

[10] H. Weiming, T. Tieniu, W. Liang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 34, pp. 334-352, 2004.

[11] Y. Tao, P. Quan, L. Jing, and S. Z. Li, "Real-time multiple objects tracking with occlusion handling in dynamic scenes," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, pp. 970-975 vol. 1.

[12] I. Haritaoglu, D. Harwood, and L. S. Davis, "W⁴: real-time surveillance of people and their activities," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, pp. 809-830, 2000.

[13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., 1999, p. 252 Vol. 2.

[14] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, 2004, pp. 28-31 Vol.2.