Adaptive Foreground Extraction for Crowd Analytics Surveillance on Unconstrained Environments

Mohamed Abul Hassan, Aamir Saeed Malik, Walter Nicolas, Ibrahima Faye

Centre for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia

Abstract. Background modeling is one of the key steps in any visual surveillance system. A good background modeling algorithm should be able to detect objects/targets under any environmental condition. The influence of illumination variance has been a major challenge in many background modeling algorithms. These algorithms produce poor object segmentation or consume substantial amount of computational time, which makes them not implementable at real time. In this paper we propose a novel background modeling method based on Gaussian Mixture Method (GMM). The proposed method uses Phase Congruency (PC) edge features to overcome the effect of illumination variance, while preserving efficient background/foreground segmentation. Moreover, our method uses a combination of pixel information of GMM and the Phase texture information of PC, to construct a foreground invariant of the illumination variance.

1 Introduction

Visual surveillance systems are essentially becoming the most attractive research areas in the field of computer vision. Their importance with respect to security and safety in public places is amongst the main reasons of growing attention in this field [1]. The accessibility to inexpensive devices and processors is an additional motivation for the promotion on investigation about visual surveillance systems. Considering that majority of visual surveillance, nevertheless, depends upon human intervention to monitor through video clips [2, 3], which is a tiresome and tedious task; keeping track of interesting events that will rarely take place. The large amount of data makes it humanly impossible to analyze and requires computer vision based solutions to automate the process.

Computer aided behavioral understanding of moving objects in a video is often quite challenging task. This requires extraction of related visual data, appropriate representation of information, as well as an interpretation of this visual information with respect to behavior learning and recognition [4]. In automated visual surveillance systems, one of the many key steps in video based human-activity recognition is usually to model the background, which often requires large expanse of processing time of the system.

Background modeling is comprised of foreground/background segmentation which yields information for various automated behavior understanding applications such as tracking, counting, direction estimation and velocity estimation of objects. Nevertheless background modeling at dynamic environments is a great challenge due to the variation in the nature of backgrounds. The motion of the sun along with the clouds in the sky is the greatest variation on the nature of the background which causes the influence of gradual and sudden illumination variance. The effect of illumination variance on the background often yields detection of false foreground mask due to shadows of the objects and false foreground detection due to sudden illumination variance [5, 6]. This provides misinterpreted information regarding the actual scene which would provide poor interpretation of the visual information for behavior learning and recognition applications.

The issue of background modeling for dynamic environments has been addressed by researches, who have proposed many background modeling algorithms based on statistical information [7, 8], fuzzy running average, clustering methods [9] and using neural network [10, 11]. However, most of these algorithms are based on specific environmental conditions, such as a specific time, place, or activity scenario and carries greater implementation complexity. Considering the implementation simplicity, Gaussian Mixture Model (GMM) is used as the background modeling algorithm in this paper. GMM is one of the most used and implemented background modeling algorithms at present due to its good compromise between computational efficiency and accuracy. However, this algorithm also suffers at above mentioned effects of illumination variance.

2   Related Work

Researches have extended the original GMM algorithm [12] to address the issue by proposing various methods such as, using multiple Gaussians by variable K Gaussian distribution [13, 14]. The adaptive background model for compensating sudden illumination variance uses variable assignment of K mean distribution or online variation of K mean distribution of the GMM model. Here the value of K varies between 3 to 5, the multiple Gaussian models, generated could distinguish between the illumination effected areas and eliminate false extraction of foreground mask. This modification dose improve the accuracy of the model but still the model has not been evaluated for different scenarios and is subjected to computational time inefficiency under the effect of sudden illumination and complex scenes.

Another method using phase texture information in place of pixel information [15]. The main advantage is that phase texture of an image is invariant to illumination. The phase texture was obtained using a Gabor filter and a phase based background subtraction was proposed by modelling the phase features independently using the MOG model and the foreground is extracted using distance transform applied at the binary image which is transformed in to a distance map by segmenting and thresholding the distance map to obtain the foreground.

The model proposed was able to compensate sudden illumination efficiently but was evaluated only for static backgrounds.

Manuel et al.[16] proposed Mixture of Merged Gaussian Algorithm (MMGA) using a combination of GMM and Real-Time Dynamic Ellipsoidal Neural Networks (RTDENN), as an updating mechanism. This model uses the conventional GMM to perform the background modeling using different color spaces (i.e. RGB or HSV). The modelled background pixels are stored for every frame, and is used for the RTDENN updating mechanism. The Gaussians are managed based on the comparison to the previous pixel information, where the existence of the Gaussians are decided by merging or creating Gaussian distributions. Eventually the updating mechanism using RTDENN functions based on the excitement of the neurons, which is the inverse of the square root of the Mahalanobis distance between the mean of the sample vectors which excite the neuron and each neuron.

Zezhi, C. and Ellis, T. proposed an illumination compensation background model by self-adaptive background modeling method [17]. This background modeling method followed [18] the, recursive GMM using multidimensional Gaussian kernel density transform. The author of this proposed method [17] used the recursive GMM along with a spatial temporal filter, which suppressed noise and compensated the illumination using median of quotient. This model was evaluated for various surveillance application scenarios alongside with crowd related datasets. However from the evaluated sequences it was observed that the model wasn't evaluated on more divers' environmental conditions.

These proposed methods have been developed in the recent past to effectively solve the issue of sudden illumination variance. However we would implement these methods along with our proposed method in section 4, and evaluate them for various scenarios. In this paper we propose a method to address the issue of illumination compensation for GMM with PC edge features to model the background. The motivation leading to formulate the background modeling method utilizing GMM and Phase congruency edge detector is due to the fact that phase texture of an image is invariant to illumination variance. Therefore, we developed the model using GMM to determine the foreground of the scene using the pixel intensity information while using phase congruency edge detector to extract the phase edge and corner information of the scene.The method utilizes the pixel intensity information and phase edge information of foreground/background to achieve efficient object segmentation, shadow removal and sudden illumination compensation.

3   Proposed Method

Our proposed method for dynamic background/foreground modeling uses a combination of Gaussian Mixture Method (GMM)[12] and Phase Congruency (PC) [19, 20] edge features. These are widely used computational tool in detecting background/foreground. However, we use a novel approach on using these methods as features/characteristics for accurate object detection and segregation un-

der dynamic environmental conditions. For each pixel of a frame X at time t in an image sequence, the pixel characteristic is determined based on the intensity of the monochromatic color space. Then, the multidimensional background is modeled based on weighted sum of the probability of observing current pixel values for K Gaussian distributions is given by,

$$P(X_t) = \sum_{i=1}^{K=3} \omega_{i,t}.\eta\left(X_t, \mu_{i,t}, \Sigma_{i,t}\right) \tag{1}$$

Where η is the probability density function, K is the number of Gaussian distributions, $\omega_{i,t}$ is a weight associated with i^{th} Gaussian at the time t with a mean $\mu_{i,t}$ and standard deviation $\Sigma_{i,t}$. The weighted sum of the probability distribution is initialized and at instance of time t the current pixel value X_t is verified if it's in range of the standard deviation of the Gaussian distribution. The pixel in range would be classified as matching to one of the K Gaussian distributions. In this case when a Gaussian distribution matches a pixel X_t, the parameters of the pixels K distributions will be updated by two scenarios. Scenario 1: for unmatched Gaussians distributions mean $\mu_{i,t}$ and standard deviation $\Sigma_{i,t}$ will be unchanged. Scenario 2: for matched Gaussians distributions $\mu_{i,t}$ and $\Sigma_{i,t}$ will be updated as shown in (Eq. 2) to (Eq. 4)

$$mu_{i,t} = (1-\rho)\,\mu_{i,t-1} + \rho.X_t \tag{2}$$

$$\Sigma_{i,t} = (1-\rho)\,\Sigma_{i,t-1} + \rho\left|(X_t - \mu_{i,t})^T (X_t - \mu_{i,t})\right|_{diag} \tag{3}$$

$$\rho = \alpha.\eta\left(X_t\left|\mu_{i,t-1}, \Sigma_{i,t-1}\right|\right) \tag{4}$$

Once the parameters are initialized, the first foreground detection is performed and the above parameters are updated for time $t + 1$ using a criterion ratio, $r_i = \frac{\omega_i}{\sigma_i}$ and the order of Gaussians following the ratio. This ordering depends upon the background pixels which corresponds to a high weight and a weak variance and foreground which corresponds to a low weight and a high variance. The foreground is considered for incoming new frame at instance $t + 1$, a match test is performed, to match the incoming pixel to the Gaussian distribution based on the Mahalanobis distance.

$$sqrt\left((X_{t+1} - \mu_{i,t})^T . \sum_{i,t}^{-1} (X_{t+1} - \mu_{i,t})\right) \le k\sigma_{i,t} \tag{5}$$

Where k is a constant established by experimentation whose value equals to 2.5. At this step the binary mask of the foreground frame $F_g\left(X_{t+1}\right)$ is extracted based on two conditions as shown in Fig. 1;

Condition 1: Pixel Matches with one of the K Gaussians. In this case, if the Gaussian distribution is identified as a background, the pixel is classified as background or else classified as foreground.

Condition 2: No match with any of the K Gaussians. In this case, the pixel is classified as foreground. Furthermore the extracted binary mask is filtered using a median filter for reducing noise.

Meanwhile the Phase texture features of the frame X at the time instance $t + 1$ is extracted using Phase Congruency Edge detector [19, 20] for every incoming frame. The phase information is extracted via banks of Gabor wavelets tuned to various spatial frequencies, instead of Fourier transform, as given by,

$$PC(X_{t+1}) = \frac{\sum_n W(X_{t+1}) \left[A_n(X_(t+1)\alpha_1 - |\alpha_2| - T_{PC} \right]}{\sum_n A_n(X_{t+1}) + \epsilon} \quad (6)$$

where α_1 denotes, $cos\left(\phi_n\left(X_{t+1}\right)\bar{\phi}\left(t+1\right)\right)$ the term is a factor that weights for frequency spread (congruency over many frequencies is more significant than congruency over a few frequencies) and α_2 denotes, $sin\left(\phi_n\left(X_{t+1}\right)\bar{\phi}\left(t+1\right)\right)$. A small constant ϵ is included to prevent division by zero. The energy values which surpass T_{PC}, the estimated noise effect, are included to the result. The notations within $'[]'$ highlights that this enclosed quantity remains the same while its value is positive, and zero otherwise. For accurate foreground object extraction, we take the intersection of each pixel for the binary mask of the matrices, $F_g\left(X_{t+1}\right)$ and $PC(X_{t+1})$.

$$F = \sum_{i=1}^{H}\sum_{j=1}^{W} F_g\left(X_{t+1}\right)\bigcap PC(X_{t+1}) \quad (7)$$

Where H and W denotes the height and the width of the image frame
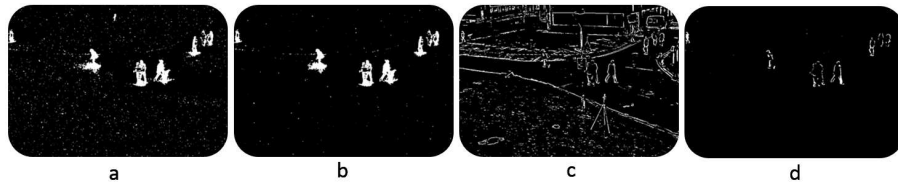
Fig. 1. Process of extracting the binary silhouette of the foreground objects. (a) represent the binary silhouette of the foreground for GMM. (b) the foreground after noise suppression. (c), Phase Congruency edge texture image. (d) extracted foreground using intersection of each pixel for the binary mask.

Finally, a set of morphological operations are performed on the extracted foreground object silhouette. Here we use these operations to enhance the boundary connectivity of the silhouette and to region fill the enclosed boundary of the objects. The process is initiated with a morphological opening operation which performs erosion followed by dilation using a 2 by 2 structured matrix as derived in (Eq. 8). Following this step the edges of the silhouette gets connected, this results on an accurate reconstruction of the boundaries of the silhouette.

Then the object silhouettes is completed by performing region filling operation as shown in (Eq. 9), where the results on each of these steps are illustrated on Fig. 2.

$$\gamma_B\left(F\right) = \bigcup_f \{B_x | B_x \subseteq F\} \tag{8}$$

$$F_k = \left(X_{k-1} \oplus B_x\right) \bigcap A^c k = 1, 2, 3 \tag{9}$$

where B_x is the structured matrix and A^c is the matrix which contains the filled set and its boundary.
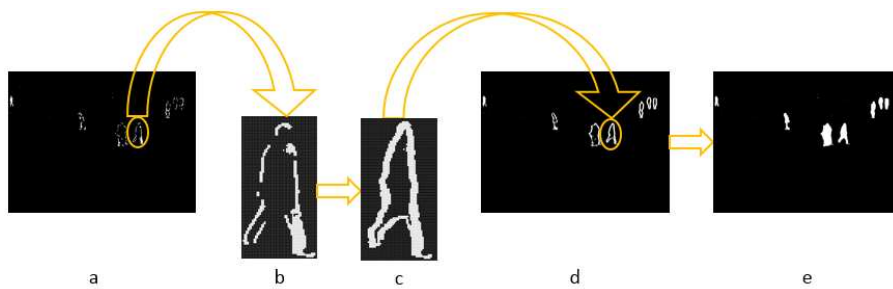

Fig. 2. Steps of the morphological operations performed on the extracted foreground object silhouette. (a) represent the original foreground silhouette extracted using the proposed method; (b) is the zoomed view of the silhouette of the object; (c) represent the boundary reconstruction of the 2 by 2 structure matrix; (d) is the full view of the boundary reconstructed image; (e) the extracted foreground silhouette after region filling.

4 Experiment Results

The experiments and analysis were carried out qualitatively and quantitatively, where the proposed model along with 5 other model in literature (See Table. 1). The binary silhouettes of the foreground extraction were taken as a final result to determine the efficiency of the models. The results were evaluated using precision recall criteria, this access the ability, to extract the true positive and eliminate the true negative detection of the implemented models.

The developed models were implemented for five different sequences from two most popular crowd surveillance databases for dynamic environments, i.e. PETS2010 [21] and OTCBVS [22]. The description of these sequences are tabulated in Table. 2. The significance of these data sets apart from testing the models on different crowd behaviors, is to challenge the accuracy of segmentation, i.e. detecting small objects from pixel range of 10×25 to 20×75, which
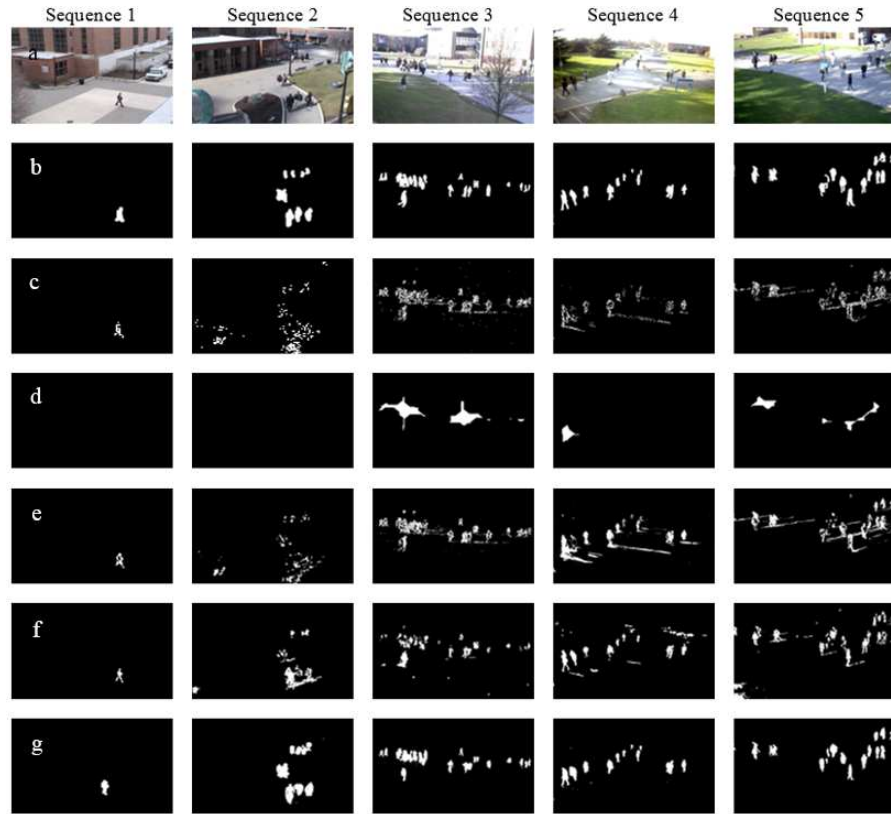
Fig. 3. The sample results of the binary silhouettes of the extracted foreground, from the different background modeling methods. The first row (a) shows the original image, (b) shows the ground truth of the extracted binary silhouettes of the original image, (c) shows the extracted binary silhouettes of GMM1, (d) shows the extracted binary silhouettes of GMM2, (e) shows the extracted binary silhouettes of GMM3 and (f) shows the extracted binary silhouettes of GMM4 followed by (g) which shows the extracted binary silhouettes of proposed method.

challenges the accuracy of segmentation. The non-synchronized nature of the data set adds the effect of sudden and gradual illumination variances continuously after every 30 to 40 frames. To summarize, the selected sequences are to yield the challenges such as, bootstrapping (BS), sudden illumination (SI), gradual illumination (GI), camouflage (CF), waving trees (WT) and shadows (S).

Table 1. Description of the implemented models in Fig. 3.

Model	Methodology	Related publication
GMM1	GMM using multiple Gaussians by variable K Gaussian distribution	[13, 14]
GMM2	GMM using phase texture information and distance transform	[15]
GMM3	GMM using RTDENN	[16]
GMM4	Using recursive GMM along with a spatial temporal filter	[17]
Proposed Method	Using a combination of GMM and PC edge features as temporal filter for each pixel	

Table 2. Description of the datasets.

Image sequence	Database	No. of frames	Size	Challenges	Description
Sequence 1	OTCBVS	1507	320×240	BS,SI,GI	Dataset 03: OSU Color-Thermal Database
Sequence 2	OTCBVS	1054	320×240	GI,BS,S	Dataset 03: OSU Color-Thermal Database
Sequence 3	PETS2010	841	768×576	BS,GI,CF	Data set of S0, City Center, view 4
Sequence 4	PETS2010	841	768×576	SI,GI, WT, S	Data set of S0, City Center, view 2
Sequence 5	PETS2010	841	768×576	SI,GI, WT, S	Data set of S0, City Center, view 1

The sample results of the binary silhouettes are shown in Fig. 3. Here the results of each model is arranged in row wise manner with respect to each sequence. The performance of each model was evaluated with respect to the ground truth which is shown in the second row. Most of the models were able to provide an accurate object detection and segregation. However GMM2 model which was based on phase texture information and distance transform failed to detect and segregate the silhouettes of the objects in each sequence. This was due to the
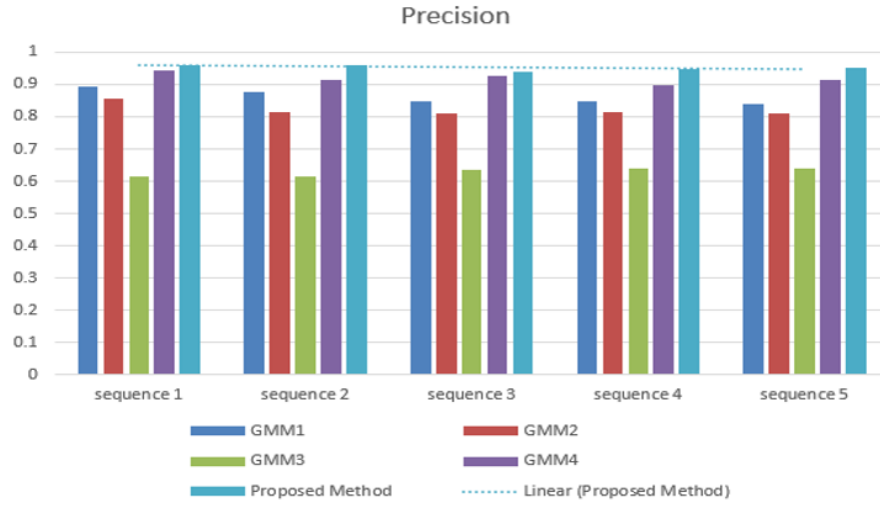
Fig. 4. Quantified Precision results obtained for five different algorithms and five sequences tested.
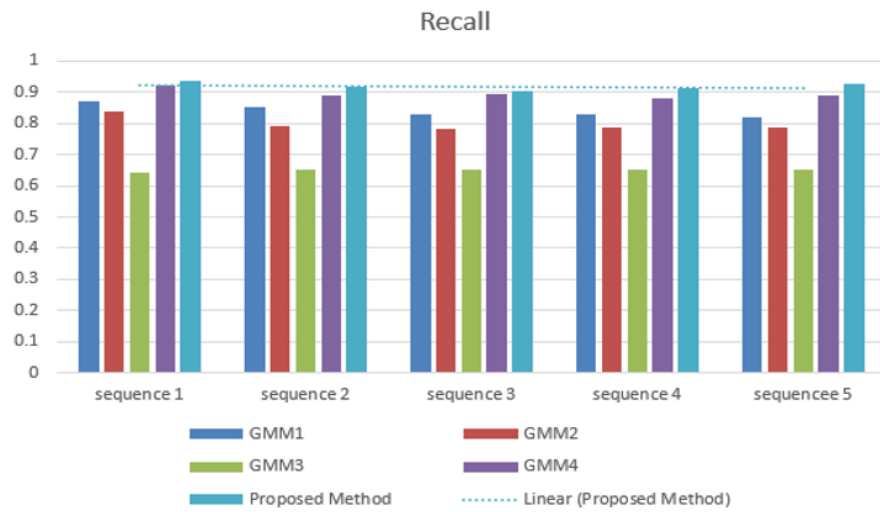

Fig. 5. Quantified Recall results obtained for five different algorithms and five sequences tested.

small object size in crowd surveillance data and the parameters used in the distance transform measures. Meanwhile the other developed models were able to eliminate challenges such as BS, WT, CF and GI.

The effect of shadows were compensated by GMM4 and proposed method, while GMM1 and GMM3 suffered heavily by extracting false foreground in sequence 4 and 5. The effect of SI was the greatest challenge to overcome, where all the models apart from GMM2 and Proposed method, suffered heavily by extracting false foreground mask due to the sudden intensity variation. The GMM2 and Proposed method were clearly able to overcome this issue since these models used phase texture features, which are invariant to illumination changes. However our proposed method was clearly able to overcome all of the above mention issues, and was able to accurately segment the foreground objects. Furthermore, the Proposed model outclassed all the other developed models and resulted in a higher precision recall criteria (See Fig. 4 and 5).

5   Conclusion

This paper proposed a background modelisation incorporating GMM and PC to adapt to unconstrained environment conditions. The main motivation of this work was to develop a robust background modeling method to extract foreground objects with accurate segregation while retaining the object silhouette. The proposed method was evaluated with four other background modeling methods.

The performance of the models was assessed qualitatively and quantitatively using precision and recall criteria. The results presented demonstrate the superiority of the proposed method in terms of accuracy for the background/foreground extraction and was able to efficiently segregate the individual objects in crowds. Moreover, the model efficiently compensated the challenging effect of sudden illumination and presence of shadows by incorporating the phase texture information along with the pixel gradient information. Future works will focus on using foreground silhouette characteristic of the objects for improving behavior learning crowd analytic algorithms such as Kalman filter and Optical flow.

References

1. Yilmaz, O.J.A., Shah, M.: Object tracking: A survey. ACM Comput. Surv **38** (2006) 45
2. Wang, X.: Intelligent multi-camera video surveillance: A review. Pattern Recognition Letters **34** (2013) 3–19
3. S. Yasir, S.A.: Comparison of stochastic filtering methods for 3d tracking. Pattern Recognition **44** (2011) 2711–2737
4. Morris, B.T., Trivedi, M.M.: A survey of vision-based trajectory learning and analysis for surveillance. Circuits and Systems for Video Technology, IEEE Transactions **18** (2008) 1114–1127
5. Pilet, C.S.J., Fua, P.: Making background subtraction robust to sudden illumination changes. ECCV **4** (2008) 567–580

6. A.Hassan, S.M. Aamir, W.N.I.F.M.T.: Mixture of gaussian based background modelling for crowd tracking using multiple cameras. International Conference On Intelligent and Advanced Systems **5** (2014) 1–4

7. L. Horng-Horng, C.J.H., Tyng-Luh, L.: Regularized background adaptation: A novel learning rate control scheme for gaussian mixture modeling. Image Processing, IEEE Transactions **20** (2011) 822–836

8. Elgammal, D.H.A., Davis, L.: Non- parametric model for background subtraction. Proc. ECCV **20** (2000) 751–767

9. Malathi, T., Bhuyan, M.K.: Multiple camera-based codebooks for object detection under sudden illumination change. International Conference on Communications and Signal Processing (ICCSP) **20** (2013) 310–314

10. Bouwmans, T.: Recent advanced statistical background modeling for foreground detection - a systematic survey. (2011)

11. C. Cuevas, R.M., Garcia, N.: Versatile bayesian classifier for moving object detection by non-parametric background-foreground modeling. 19th IEEE International Conference on Image Processing (ICIP) (2012) 313–316

12. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. International Conference on Computer Vision and Pattern Recognition **2** (1999) 252

13. L. Dawei, X.L., Goodman, E.: Online background learning for illumination-robust foreground detection. International Conference on Control Automation Robotics and Vision (ICARCV) **11** (2010) 1093–1100

14. T. Huang, X. Fang, J.Q., Ikenaga, T.: Adaptively adjusted gaussian mixture models for surveillance applications, in advances in multimedia modeling. S. Boll, Q. Tian, L. Zhang, Z. Zhang, and Y.-P. Chen, Eds., ed: Springer Berlin Heidelberg **5916** (2010) 689–694

15. X. Gengjian, S.J., Li, S.: Background subtraction based on phase feature and distance transform. IEEE 17th International Conference on Image Processing **17** (2012) 3465–3469

16. Manuel Alvar, Andrea Rodriguez-Calvo, A.S.M.A.A.: Mixture of merged gaussian algorithm using rtdenn. Machine Vision and Applications **25** (2014) 1133–1144

17. Zezhi Chen, T.E.: A self-adaptive gaussian mixture model. Computer Vision and Image Understanding **122** (2013) 35–46

18. Z. Zivkovic, F.v.d.H.: Recursive unsupervised learning of fnite mixture models. IEEE PAMI **5** (2004) 651–656

19. Kovesi, P.: Phase congruency detects corners and edges. Proc. VIIth Digital Image Computing: Techniques and Applications **8** (2013) 10–12

20. A.Hassan, S.M. Aamir, W.N.I.F.M.T.N.R.W.: Foreground extraction for real-time crowd analytics in surveillance system. 2014 IEEE 18th International Symposium on Consumer Electronics (ISCE 2014) **18** (2014) 1–2

21. Ferryman, J., Ellis, A.: Pets2010: Dataset and challenge. Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) **7** (2010) 143–150

22. Dataset:, O.: http://www.cse.ohio-state.edu/otcbvs-bench/ (2012)