

An adaptive hybrid GMM for multiple human detection in crowd scenario

Karpagavalli P.¹ · Ramprasad A. V.¹

Received: 19 November 2015 / Revised: 29 June 2016 / Accepted: 10 July 2016 /
Published online: 10 August 2016
© Springer Science+Business Media New York 2016

Abstract People gather together for myriad reasons and in an overcrowded region, human detection is a very challenging problem. Automated multiple human detection is one of the most active research fields of computer vision applications. It provides useful information for crowd monitoring and traffic controlling for human safety in public places. The conventional Gaussian mixture model, that has a fixed K values, is not enough for dynamically varying background and further foreground detection is a time consuming process. The automated multiple human detection algorithms are needed to deal with complex background and illumination change conditions of crowd scene. In this paper, an automated multiple human detection method using hybrid adaptive Gaussian mixture model is proposed to handle efficiently the complex background and illumination changes. The proposed hybrid algorithm utilizes spatiotemporal features, adaptive learning control, adaptively changing weights and an adaptive selection with number of K Gaussian components per pixel to withstand in complex background and different lighting conditions. The experimental results and performance measures demonstrate that the proposed hybrid method performs well for crowd scene. By using the proposed adaptive hybrid method, the multiple human detection rate has been improved from 90 % to 95 % and the computational time is reduced from 5secs to 2.5secs.

Keywords Background subtraction · Adaptive Gaussian mixture model · Multiple human detection · Computer vision · Automation

✉ Karpagavalli P.
karpagavallimohan@gmail.com

¹ K.L.N College of Engineering, Anna University, Chennai, India

1 Introduction

In public places, single camera is not sufficient for human detection and monitoring a crowded environment for video surveillance. Real-world applications need multiple-cameras to capture crowd scenes and to prevent the misdetection in a crowd. Video analysis is very important for human safety for multiple human detection, tracking of multiple human, monitoring and traffic controlling in public places such as shopping mall, airport and public meetings. The automatic detection of high density people in a crowd has resulted in number of challenging problems in computer vision applications. In crowded places, the existence of multiple-cameras is necessary to cover many regions, resulting in overlapping during video surveillance. The demand is to process challenging multi-view videos by using computer vision algorithms. Traditional background subtraction algorithm [9, 17] has the advantage of computational speed, but it fails to support multi-view videos in crowded environments.

The central idea behind adaptive background subtraction is to utilize the visual properties of a scene for building an appropriate representation of the background and that can be used to classify a new observation as foreground or background. This model is well suited to a high-density crowd whose background model has a multimodal distribution. Changes in background and lighting conditions cause problems in many of the background modeling methods. When a moving object is still for a short duration, the pixel values are rapidly updated according to the background model by using the Improved Adaptive K Gaussian Mixture Model (IAKGMM). The challenging problems can be prevented by the proposed hybrid adaptive based Gaussian mixture model. The proposed method utilizes the advantages of GMM [18] and Improved Adaptive K GMM (IAKGMM) [29] background subtraction methods. The block diagram of the proposed method is shown in Fig. 1.

The videos are taken from different viewpoints using multiple-cameras to cover large areas. To accurately extract the foreground of each frame, the proposed method hybridises the output of GMM [1, 2, 19, 27, 29] to improved adaptive K Gaussian mixture model output. In this scenario, occlusions occur in owing to walking, running, and standing of people. The proposed algorithm has the following advantages.

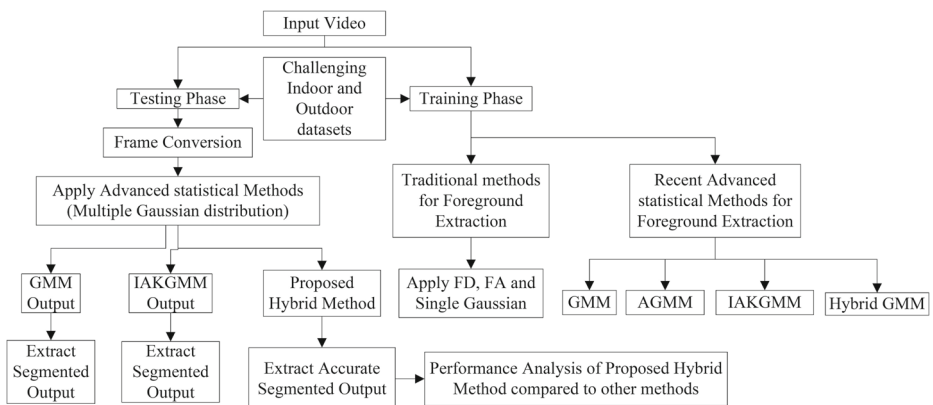


Fig. 1 The proposed multiple human detection framework

- (i) Robust multiple human detection is achieved by using the proposed method for dynamic background, different camera viewpoints, different abnormal conditions, and different illumination conditions in the UMN and PET2009 crowd datasets [5].
- (ii) The proposed method works better in dynamic background and different illumination conditions in a crowd scene.

The paper is organized as follows: Section 2 deals with related works. The proposed hybrid adaptive mixture model is explained in Section 3. In Section 4, experimental results are discussed, and the conclusion is given in Section 5.

2 Related works

The challenging problems, that normally occur, are misdetection in the crowd, occlusion among people, shadow interference, illumination changes and abnormal events in crowd scenes and slow or fast moving objects. The dynamic background for different issues (swaying tree, wind over the water surface, tree branches and over the grass) is explained in the papers [21, 27], but it does not handle illumination changes and multi-view problem in crowd environment. These issues are addressed in the proposed adaptive hybrid GMM method. The advanced statistical models and adaptive based background learning methods are the best methods to handle dynamic background [1, 2, 19] for the crowd video. So, the proposed method utilizes the model. The PETS2009 outdoor dataset for crowd video analysis and surveillance is experimented to evaluate the performance of human detection [5].

The slowly moving foreground objects can be modelled by Eigen background model and the Online Eigen background model is presented without considering the quick and gradual illumination changes in a crowd scene [6, 23]. The modified version of ViBe is used with the new updating policy for modelling the background in a water surface without considering the multi-model distribution technique [24]. The texture based background subtraction is used in [10], and the algorithm does not consider moving or tilting camera videos. The highly dynamic background scenes are not discussed [4, 15] in the detection of moving objects. The highly dynamic background object analysis is discussed [11, 16, 20] which does not support multi-human detection, gradual and sudden illumination conditions.

The multiple head detection is done through homographies method for a multi-camera video [3, 29] but manual initialization is needed for parameter settings. The background subtraction algorithm utilizes the temporal variations of intensity and codebook model is explained in [12, 13, 28] by considering vibrating motion patterns and non-statistical clustering technique. The human-part detection techniques in the crowd environment does not provide good detection rate, if the occlusion is more [26].

The new learning rate control for Gaussian parameters without background color variations is analyzed by D.S. Lee [14]. The adaptive based background subtraction method and model based detection are described in [7, 8] to detect human in the moderate crowd environment and not an environment with a larger crowd. The dynamic background is effectively handled to analyze the various scenes by using an adaptive background subtraction method [22]. The improved adaptive GMM may be used to reduce the processing time [2, 25] and the probabilistic mixture models are quite powerful methods for multiple object motion [9, 18] and The proposed algorithm automatically selects the needed number of Gaussian components per pixel and fully adapts to the scene with less processing time.

The proposed method takes the advantages of adaptive GMM and IAKGMM methods and is applied to challenging crowd video in order to get accurate multi-human detection. It is inferred that the integrated methods, the hybrid models and statistical methods using multiple Gaussians are used to handle the complex background very efficiently. The adaptive based GMM and hybrid models show better results under crowd, swaying trees (dynamic background) and illumination changes conditions (low sunshine, bright sunshine and overcast with shadows). It is found that existing methods are not well suited for some challenging conditions in crowd scenario.

The proposed work resolves the challenges in complex background and different illumination circumstances and reduces the misdetection problem. The proposed approach hybridizes automatic multiple human detection under various challenging conditions by handling occlusion using multi-camera datasets. The proposed hybrid method takes the advantages of GMM and Improved adaptive K GMM to work under complex background and different illumination conditions.

3 Proposed adaptive hybrid Gaussian mixture model

Multi-target detection is important in real-world applications especially in a crowded area. In a crowd video, increasing complexity [6] depends upon the number of targets of detecting multiple human. Single-view camera is not sufficient for detecting multiple targets, because it does not cover larger areas. To detect multiple humans accurately, multiple-cameras are needed to avoid misdetection and misclassification. The proposed method is applied to multiple-cameras datasets (multi-view) with different changing illuminations datasets [5] and thus it works well for dynamic background and illumination changes.

3.1 Proposed multiple human detection

The observed pixels are updated and modeled by a mixture of K Gaussian distribution based on mean and variance values in a crowd video for multiple human detection. The hybrid adaptive Gaussian mixture model withstands for sudden, gradual illumination changes and clutter background [21, 27]. The history of the observed pixels is described first by K Gaussian distribution, and then, the probability of these pixels can be estimated. From the observation, $\hat{\mu}_m$ is mean values of the i^{th} Gaussian distribution, $\hat{\sigma}_m^2$ is standard deviation of the i^{th} Gaussian distribution and co-variance matrix can be estimated, where $\hat{\Gamma}_m$ is the weight estimation of the i^{th} Gaussian distribution in the mixture model at time t and co-variance matrix is the variance matrix of the i^{th} Gaussian distribution. The Gaussian probability density function f can be defined by the following equation

$$f(\tilde{a}^{(t)}, \hat{\mu}_m, \hat{\sigma}_m^2) = \frac{1}{\hat{\sigma}_m^2} \sqrt{2\pi} e^{-\frac{[\tilde{a}^{(t)} - \hat{\mu}_m]^2}{2\hat{\sigma}_m^2}} \quad (1)$$

Where

- $\hat{\mu}_m$ = Mean of expectation of the pixel value
- $\hat{\sigma}_m^2$ = Standard deviation of the pixel value
- $\tilde{a}^{(t)}$ = pixel value

In a crowd video, the pixel values of each frame at time t in RGB (Red, Green, Blue) color space are denoted by $\tilde{a}^{(t)}$. Background subtraction is based on pixel wise classification method. In order to classify the background and foreground pixels, the decision rule is defined by

$$Decision = \frac{P(BG/\tilde{a}^{(t)})}{P(FG/\tilde{a}^{(t)})} = \frac{P(BG/\tilde{a}^{(t)})P(BG)}{P(FG/\tilde{a}^{(t)})P(FG)} \tag{2}$$

where P is the probability of the observed pixels of the temporal frames, BG is the background, and FG is the foreground of the crowd video. By using the above decision rule, the moving or static pixels are classified from the observed pixel values. Human detection can be carried out by the proposed hybrid method using a threshold equation to classify the background pixels from observed values:

$$\left\{ \begin{array}{l} background, \text{ if } (P(BG)/\tilde{a}^{(t)}) < b_{threshold} \end{array} \right. \tag{3}$$

$$\left\{ \begin{array}{l} foreground, \text{ if } (P(BG)/\tilde{a}^{(t)}) > b_{threshold} \end{array} \right. \tag{4}$$

If the observed pixel values are lower than $b_{threshold}$, those pixels shall be treated as background pixels. The remaining pixel values are treated as foreground pixels. Initially, it is assumed that $b_{threshold}$ is 50, and this value varies adaptively depending upon the observed pixels. The proposed method hybridises the GMM and improved adaptive K GMM background subtraction methods [14, 26] to handle the complex background, different view-points, and illumination changes in crowd videos in a better way. In order to adapt to these changes, the weight values of the Gaussian distribution of each pixel can be updated in the training set by adding weights to the new sample pixels and by rejecting all other old training pixels. Consider the time period T , the observed sample pixels at t are denoted by

$$\psi_T = \{\tilde{a}^{(t)} \dots \tilde{a}^{(t-T)}\} \tag{5}$$

During training phase, in order to adapt to scene changes, the new samples are added and old samples are discarded. Consider a reasonable total time period T and the observed time period is t , the observed pixels are classified from the history of the previous pixels. Based on the weights values of the Gaussian distribution, some of the values belong to foreground pixels and the remaining pixels belong to background. The estimated model can be defined by the K Gaussian distribution components of the IAKGMM

$$P(\tilde{a}^t / \psi_T, BG + FG) = \sum_{m=1}^K \hat{\Pi}_m v(\tilde{a}^t; \hat{\mu}_m, \hat{\sigma}_m^2; I) \tag{6}$$

where I is the identity matrix in the proper dimension, $\hat{\Pi}_m$ represents the weight values of each pixel, K characterizes the number of Gaussian distributions, $\hat{\mu}_m$ represents the mean of the Gaussian distribution, and $\hat{\sigma}_m^2$ characterize the variance of the Gaussian distribution.

3.2 Parameter selection processes

For each frame, the Gaussian values of each pixel are learnt and then updated. The parameter selection process includes the following two processes

1. Learning process.
2. Update Process.

The match matrix A is constructed in the parameter selection process [14]. It is a temporary matrix, consists of a history of the pixel values during the learning process. Steps involved in pixel learning and updating process.

1. STEP1: Construction of A matrix.

Consider A matrix with dimension [750x946]. The matrix A has a mixture of Gaussian weight values.

$$A = \begin{pmatrix} W_{1,1} & W_{1,2} & \dots & W_{1,N} \\ W_{2,1} & W_{2,2} & \dots & W_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ W_{M,N} & W_{M,N} & \dots & W_{M,N} \end{pmatrix} \tag{7}$$

Where the size $M = 750(\text{row})$ and $N = 946(\text{column})$

Where $W_{1,1}, W_{2,2}, \dots, W_{M,N}$ are the Mixture of Gaussian weight values.

2. STEP2: Model a pixel process.

The Gaussian distribution weight values are initially modelled according to pixel observation and then based on the learning process, Gaussian weight values are updated.

3. STEP3: Classify the background pixels.

If the observed pixels match with any one of the existing Gaussian distribution and have low variance, the weight values of the match matrix A are not updated. These pixels are classified as background pixels.

4. STEP4: Classify the foreground pixels.

If the observed pixels do not match with the existing Gaussian distribution and have high variance, then the weight values of the match matrix A are updated. These pixels are classified as foreground pixels.

5. STEP5: Handle a dynamic scene based on Learning and update process.

In order to handle the dynamic variations of each observed pixel, the number of K Gaussian distribution is adaptively selected. The proposed adaptive hybrid GMM withstands quick and gradual illumination changes handling through match matrix A . Assign the Gaussian components $K=1, 2, 3, 4, 5$ to maintain the variety of BG, FG, swaying trees, quick illumination and gradual illumination changes respectively. Hence, the proposed adaptive hybrid GMM is robust to adaptive changes in a crowd scene.

6. STEP6: Adaptation to Multimodal distribution.

Smaller values of threshold T and Gaussian component $K = 1 \text{ or } 2$ are considered for uni-modal distribution. But, adaptive threshold (high values) and adaptive Gaussian components are needed in order to handle multimodal distribution.

The weight values of match matrix A can be updated based on recent observed pixels during learning and matching process. Each value in A represents whether the K^{th} Gaussian of each pixel in each frame matches or not. If the Gaussian values for any of the pixels match with a value of 1, and the value of each pixel within a standard deviation in the range from 2 to 2.5 of the K^{th} Gaussian distribution, these pixels are classified as foreground pixels. If there is a mismatch, these pixels are classified as background pixels. But any one of the pixel from the frame has an almost matches with Gaussian distribution from its mixture. The closest distribution of a matching Gaussian is selected from its mixture [4, 20]. For a new sample \tilde{a}_t at time t , the recursive update equation is as follows:

$$\hat{\Pi}_m \leftarrow \hat{\Pi}_m + \alpha(o_m^{(t)} - \hat{\Pi}_m) \tag{8}$$

$$\hat{\mu}_m \leftarrow \hat{\mu}_m + o_m^{(t)}(\alpha/\hat{\Pi}_m)\bar{\delta}_m \tag{9}$$

$$\hat{\sigma}_m^2 \leftarrow \hat{\sigma}_m^2 + o_m^{(t)}(\alpha/\hat{\Pi}_m)\bar{\delta}_m^T\bar{\delta}_m - \hat{\sigma}_m^2 \tag{10}$$

where α is a variable ($\alpha = 1/T$) and the learning parameter, $\bar{\delta}_m = \tilde{a}^{(t)} - \hat{\mu}_m$, where $\bar{\delta}_m$ represents standard deviation of the selection process. $\bar{\delta}_m^T$ is calculated over T . When none of the K Gaussian matches the current pixel, it implies that the pixel value is not close to any of the distributions in the mixture. Assume that $\tilde{a}^{(t)}$ is a foreground pixel and for each K Gaussian in the mixture, the following parameters are calculated:

$$\eta(\tilde{a}^{(t)}, \hat{\mu}_{1,t}, \sum K, t) \tag{11}$$

where $\hat{\mu}_m$ is the mean value, and $\sum K, t$ is the covariance matrix of the K^{th} Gaussian in which the covariance matrices are diagonal. If the above said condition is satisfied, the mean is set to $\tilde{a}^{(t)}$, the variance is set high and the weight is set low. Over the time, the distribution will slowly gain a higher weight with a low variance value. Therefore, $\tilde{a}^{(t)}$ belongs to part of the background.

The flow diagram of the proposed algorithm is shown in Fig. 2. It explains the adaptive learning and parameter update process per pixel of the proposed method. Finally, the foreground is classified from the background for a given video (challenging conditions).

3.3 Adaptively select K Gaussian Components

The weight $\hat{\Pi}_m$ is the m^{th} component of the improved adaptive K GMM. It describes the pixels that belong to the m^{th} component of the multimodal distribution. The probabilities of the observed pixels are determined from the samples of the m^{th} component pixel of moving objects. The number of data samples is d and each sample corresponds to any one of the components of the adaptive GMM. This can be defined by

$$n_m = \sum_{i=1}^d o_m^{(i)} \tag{12}$$

where n_m is the number of samples from the m^{th} component, and o_m is the new sample from the newly observed pixels. If the components are close with the largest $\hat{\Pi}_m$, then $o_m^{(i)}$ is set to 1 and the others are set to 0. The adaptive weight can be defined by the following equation,

$$\hat{\Pi}_m = \frac{1}{K}(\sum_{i=1}^t o_m^{(i)} - c) \tag{13}$$

The final improved adaptive update equation is as follows:

$$\hat{\Pi}_m \leftarrow \hat{\Pi}_m + \alpha(o_m^{(t)} - \hat{\Pi}_m) - \alpha c_T \tag{14}$$

where α is the learning parameter of the Gaussian process for observed pixels, and c is the coefficient of the Gaussian distribution. The update process starts with the improved adaptive K GMM with one component of the first sample, and later new components are added to the current sample values for all frames.

3.4 Proposed algorithm

- Initially $K = 1$, Learning parameter $\alpha = 0.01$, $b_{threshold} = 50$, mean = $\hat{\mu}_m$, variance = $\hat{\sigma}_m^2$, co-variance = $\sum K, t$, temporary matrix A , $\hat{\Pi}_m$ is the Gaussian weight value, standard deviation = 2 to 2.5.

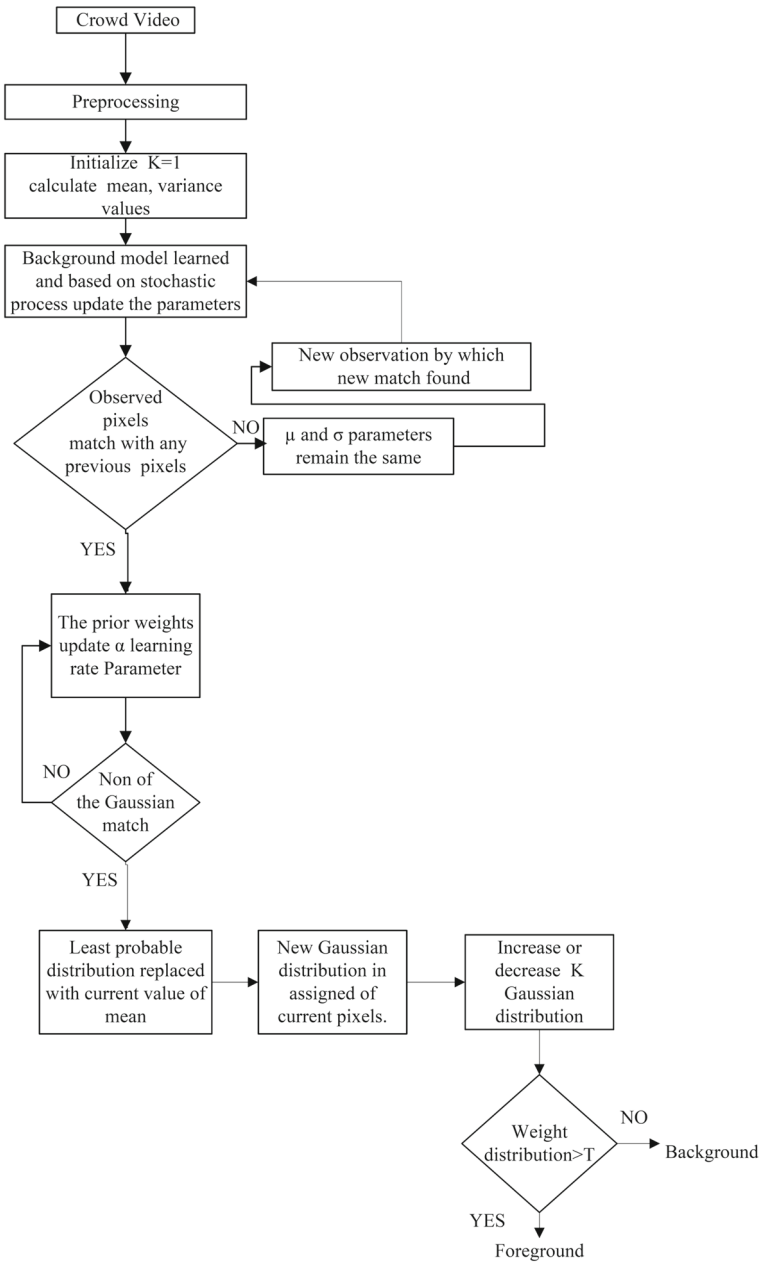


Fig. 2 Flow diagram of adaptive learning and parameter update process of the proposed method

2. Read Input crowd video. Where N = number of frames, K = Number of Gaussian components.
3. Set $b_{threshold}$ value, start learning process, x = Rows and y = columns of each frame, consider time period = T , observed time = t .

4. Pre-process of noise video and Start GMM with learning process for a given crowd video and Construct temporary matrix A .
5. Calculate mean = $\hat{\mu}_m$, variance = $\hat{\sigma}_m^2$, then calculate co-variance = $\sum K, t$.
6. Apply decision rule- $\frac{P(BG/\hat{a}^{(t)})P(BG)}{P(FG/\hat{a}^{(t)})P(FG)}$, then adaptively adjust threshold based on observed pixels and Choose $b_{threshold}$ higher value, it handle multi-model distribution.
7. During matching process, current pixels are matched to observed pixels. If match founds, IAKGMM and Stochastic update process.
8. For multi-model distribution, the threshold, learning rate, weight values and needed number of K Gaussian components are automatically and adaptively selected.
9. The prior weight $\hat{\Pi}_m$, mean $\hat{\mu}_m$, standard deviation $\hat{\sigma}_m$ are the K^{th} distributions of the Gaussian are adaptively updated.
10. If the current pixel does not match with none of the K^{th} distribution, then $\hat{\mu}_m, \hat{\sigma}_m$ of the K^{th} distributions values are remains the same. After that, new adaptive Gaussian distribution is made.
11. Weight $\hat{\Pi}_m$ is smaller, then $b_{threshold}$ value is changed while the remaining weights are renormalized.
12. GMM handles gradual illumination changes and clutter background. IAKGMM utilizes the stochastic approximation procedure to update the parameters with optimal number of mixture components.

4 Experimental results

4.1 UMN, Crowd datasets

The UMN and crowd outdoor datasets are used to analyze the proposed method. These datasets have a crowd of people with a sudden abnormal event or activity and the movement of people in different directions with occlusion in the video. To obtain accurate human detected output in a sudden and gradual illumination conditions, the proposed hybrid method of multiple human detection is applied to the crowd dataset. Initially, α value is set as 0.01 and then, it is changed adaptively (ie. 0.01 to 0.05) according to different locations and scene types in order to handle the dynamic background very effectively. In the complex background, the K Gaussian components are adaptively selected based on the multiple distributions in a video by using the proposed work.

4.2 PETS2009 datasets

Many computer vision applications need standard datasets for object recognition and video analysis. This can be fulfilled by the PETS2009 crowd datasets [5]. The PETS2009 datasets consist of three datasets recorded at the White Knights Campus, University of Reading, and UK. The datasets consist of people detection and density estimation (Dataset S1), people counting and tracking (Dataset S2), and flow analysis and event recognition (Dataset S3). The proposed work is evaluated and analyzed in PETS2009 datasets with all combinations of the above mentioned datasets. The performance of the proposed method is measured and analyzed using the Receiver Output Characteristics (ROCs), the Mean Absolute Error (MAE) and Mean Relative Error (MRE). For the ROC analysis,

$$Precision = TP / (TP + FP) \quad (15)$$

$$Recall = TP / (TP + FN) \quad (16)$$

$$F - Measure = 2(Precision.Recall)/(Precision + Recall) \quad (17)$$

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (18)$$

where TP is True Positive when the foreground pixels are classified correctly as foreground,

FP is False Positive when the background pixels are classified incorrectly as foreground,

TN is True Negative when the background pixels are classified correctly as background,

FN is False Negative when the foreground pixels are classified incorrectly as background.

The validity of the proposed method is tested by using Mean Absolute Error (MAE) and Mean Relative Error (MRE). They are defined as

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^N |G(i)_m - T(i)_m| \quad (19)$$

$$MRE = \frac{1}{N} \cdot \sum_{i=1}^N \frac{|G(i)_m - T(i)_m|}{T(i)_m} \quad (20)$$

where N is the number of frames of the test sequence, and $G(i)_m$ and $T(i)_m$ are the guessed and true number of persons in the i^{th} frame for the multimodal distribution, respectively. The efficiency of the proposed system has been evaluated by carrying out extensive works on simulation of the algorithm using UMN, crowd and PETS2009 datasets. It is assumed that PETS2009 and UMN, crowd datasets are recorded according to the camera view-point and by the way people move on the ground plane in the real world. The proposed method is processed with crowd video at 24 frames per second. The multiple human detections are obtained by using frame differencing algorithm, hybrid frame difference to GMM algorithm, GMM and the proposed hybrid algorithm.

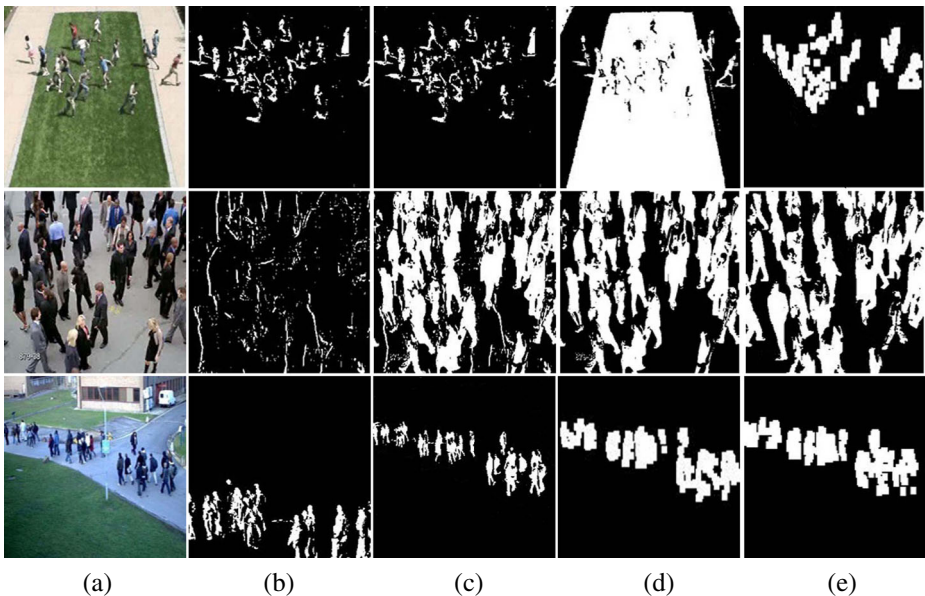


Fig. 3 Results on real sequences of three dynamic scenes: (a) input frames; (b) frame differencing outputs; (c) hybrid outputs of frame differencing with GMM; (d) outputs of GMM; (e) outputs of the proposed method

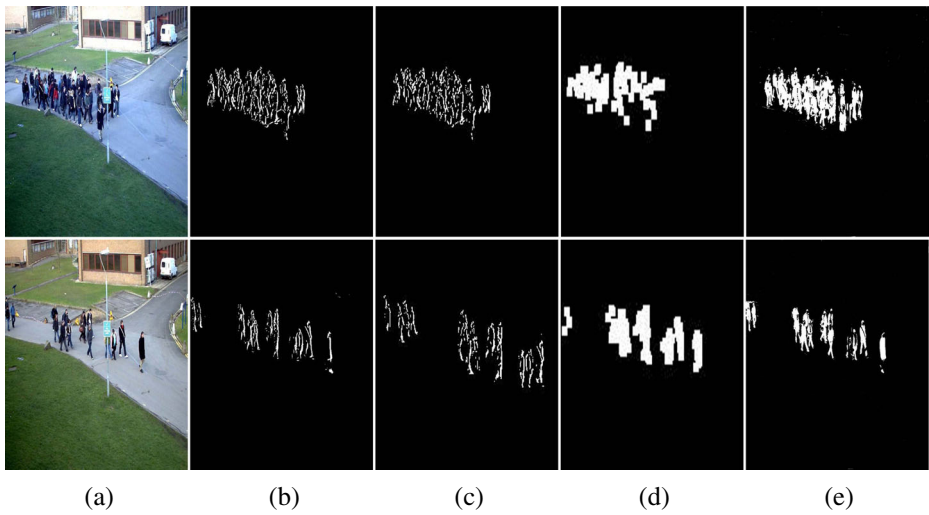


Fig. 4 Results on real sequences of two dynamic scenes: (a) input frames; (b) frame differencing outputs; (c) hybrid outputs of frame differencing with GMM; (d) outputs of GMM; (e) outputs of the proposed method

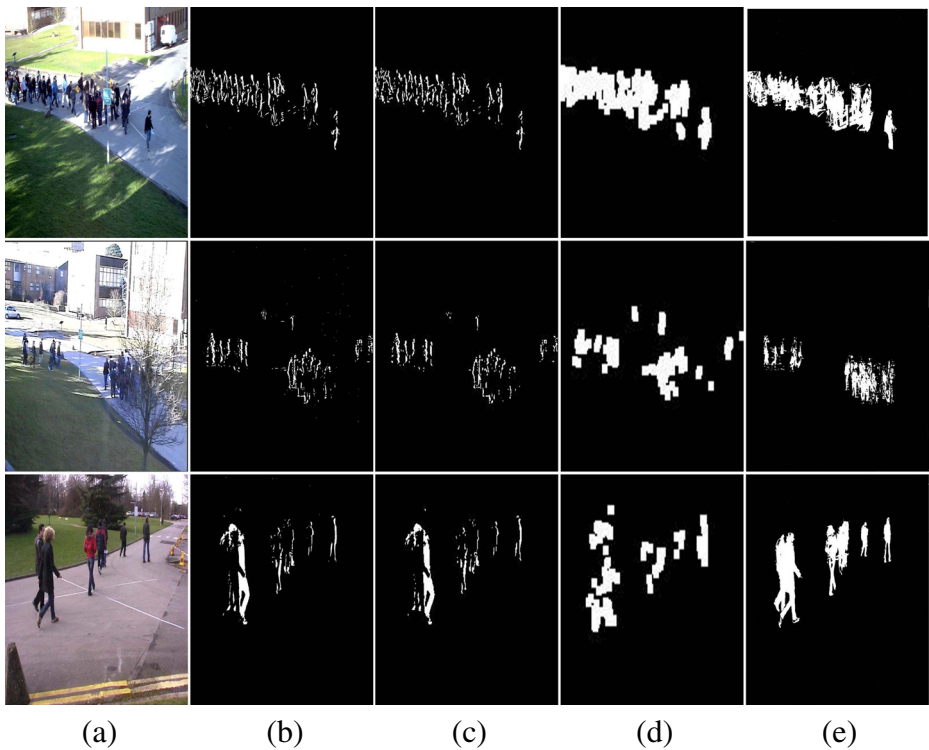


Fig. 5 Results on real sequences of three dynamic scenes: (a) input frames; (b) frame differencing outputs; (c) hybrid outputs of frame differencing with GMM; (d) outputs of GMM; (e) outputs of the proposed method

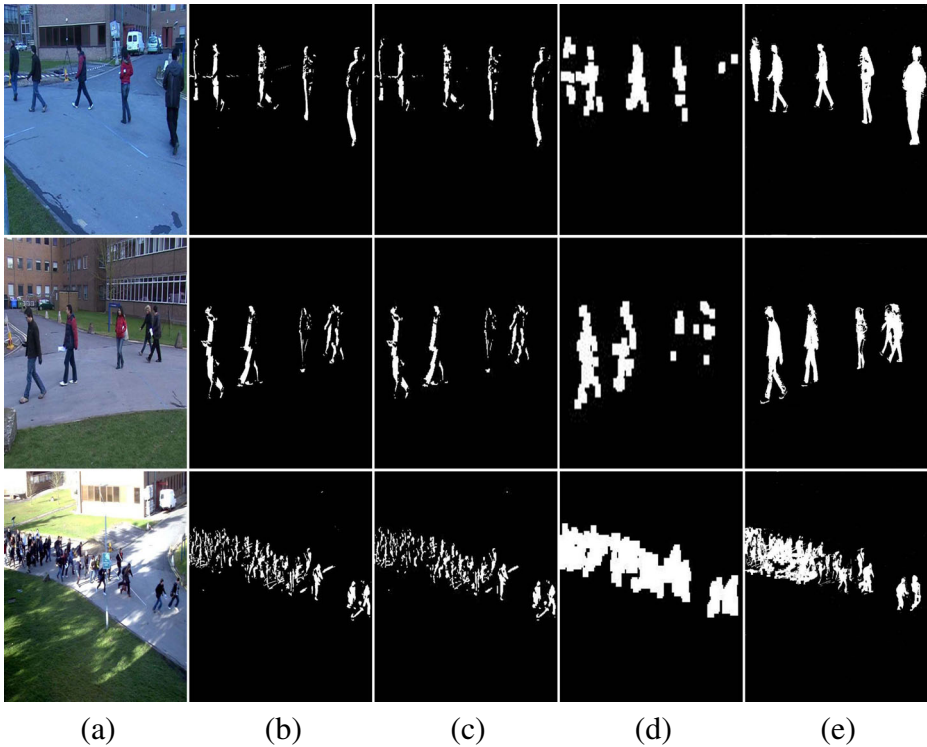


Fig. 6 Results on real sequences of three dynamic scenes: (a) input frames; (b) frame differencing outputs; (c) hybrid outputs of frame differencing with GMM; (d) outputs of GMM; (e) outputs of the proposed method

Figures 3, 4, 5 and 6 show the input frame, frame differencing, hybrid frame differencing to GMM, GMM and the proposed hybrid adaptive mixture model.

Table 1 lists the challenging characteristics of twelve sequences of the PETS2009 datasets [5], UMN and crowd datasets. Table 2 shows the parameter settings of the proposed work compared to other existing algorithms based on threshold values, learning rate and number of Gaussian components used to update the dynamic background. From Table 2, it is observed that the proposed work has adaptively changed threshold according to the observed pixels. From the same table, it is demonstrated that the proposed method works well in a crowded situation.

In Table 3, the precision and recall values for the different methods are compared with the proposed method. In this analysis, the precision and recall values of the proposed method are found to have higher values than the other techniques. Table 4 summarizes the performance of the proposed method analyzed on the basis of the MAE, MRE and accuracy. From the values of MAR, MRE and accuracy, it is shown that the proposed method has minimal error with improved accuracy. Tables 5 and 6 presents the F-measure and computational time against all the other methods compared to the proposed method. Figure 7a-b shows the graph of the proposed method based on the true positive and false positive rates compared with the existing methods. It shows the accurate detection of the foreground in a crowd video. Figure 7c shows the comparison graph of the proposed method based on detection rate. It shows that detection rate of the proposed method is improved compared to the

Table 1 Relevant challenging characteristics of ten sequences from PETS 2009 datasets and two sequences from crowd datasets used to evaluate the performance of the proposed method. where FL = Frame Length, IF = Input Frame, MC = Medium Crowd, VLC = Very Low Crowd, HC = High Crowd, OV = Overcast, BS = Bright Sunshine, LS = Less Sunshine

Videos [16]	Challenges	FL	IF	Conditions	People(in crowd video)
S1.L1.13-57(1)	Fluttering tape	221	95	MC, OV	31
S1.L3.14-17(1)	Fluttering tape	91	73	MC, BS, shadows	40
S3.MF.12-43(6)	Lack of natural lighting	107	77	VLC, LS	7
S3.MF.12-43(5)	Shadows	107	97	VLC, OV	7
S3.MF.14-13(1)	Fluttering tape	231	91	HC, BS, shadows	42
S3.MF.12-43(1)	Fluttering tape	108	77	VLC, OV	7
S3.MF.14-13(1)	Fluttering tape	231	91	MC, BS, shadows	42
S3.MF.14-37(3)	Tree partially occluding	108	23	MC, BS, shadows	27
S3.MF.12-43(7)	Vehicle and tree moving	107	91	VLC, OVt	7
S3.MF.14-52(4)	Reflective windows	92	84	HC, BS, shadows	27
S1.L2.14-06(1)	Fluttering tape	201	102	HC, OV	37
S1.L1.13-59(1)	Fluttering tape	241	112	MC, OV	31
UMN dataset	Dynamic background	120	87	BS, shadows	15
Crowd2 dataset	Dynamic background	135	36	LS	30

existing methods [29]. Figure 8a-b shows the graph of the proposed method based on the detection rate and error rate. From Fig. 8a-b, it is proved that the proposed approach always achieves low false positive rate (less than 1 percentage) with high detection accuracy for all α values.

4.3 Adaptation over lighting changes and scene changes

The proposed hybrid method utilizes the property of adaptive adjustment of mixture weight through adaptive learning rate parameter α for different locations of pixels and different dynamic background changes. The GMM [17, 18, 29] algorithm is tuned to stand for quick changes in background and IAKGMM [2, 22] algorithm tolerates the shadow problem, quick and sudden lighting changes through adaptive control of learning rates. The visual qualities of the proposed method shown in Figs. 3e, 4e, 5e and 6e are plausible for different challenging conditions in a crowd video compared to other [18, 27] methods.

The input video of Fig. 3a has dynamic background, complex motion of human and abnormal activities with crowd environment and the proposed hybrid method works well

Table 2 Parameter settings of BGS algorithms. Where FD = Frame Differencing, HFDGMM = Hybrid Frame Differencing with GMM, GMM = Gaussian Mixture Model

Methods	Settings
FD	$b_{threshold}=15$
HFDGMM [21]	$b_{threshold}=16, \alpha = 0.02, n=5$
GMM [12, 27]	$b_{threshold}=15, \alpha = 0.01, n=3$
Proposed	$b_{threshold}=50, 20, 30, 35, 55$ etc., , $\alpha = 0.01$ to $0.25, n=5$

Table 3 Precision and recall values for different datasets analyzed by the proposed method and compared with existing methods. where FD = Frame Differencing, PRE = Precision Value, REC = Recall Value

Videos Datasets	FD		HFDGMM		GMM		Proposed	
	PRE	REC	PRE	REC	PRE	REC	PRE	REC
S1.L1.13-57(1)	74	61	76	70	89	78	97	87
S1.L3.14-17(1)	82	70	83	72	84	79	97	83
S3.MF.12-43(6)	81	65	82	69	84	81	99	79
S3.MF.12-43(5)	72	84	74	80	89	82	96	89
S3.MF.14-13(1)	73	69	75	71	86	80	95	86
S3.MF.12-43(1)	76	82	77	84	87	79	93	84
S3.MF.14-13(1)	74	67	76	70	90	78	90	82
S3.MF.14-37(3)	73	82	74	84	84	70	95	78
S3.MF.12-43(7)	68	82	70	81	87	71	92	80
S3.MF.14-52(4)	70.5	72	72	78	88	79	96	84
S1.L2.14-06(1)	74	56	76	63	91	82	93	87
S1.L1.13-59(1)	67	81	70	83	87	76	94	75
UMN	76	85	78	80	87	81	97	83
Crowd2	70	55	75	62	86	85	93	96

in that conditions. It is observed that, the quick and accurate multiple human detection is obtained in Fig. 3e by using the proposed method and it also prevent the abnormal activities in crowd scene compared to other methods. This method proves to be very useful to crowd monitoring and controlling crowd scenario (eg. Shopping malls, public meeting etc.,).

Table 4 Performance measures using the MAE and MRE for the proposed method. where IF = Input Frame, MAE = Mean Absolute Error, MRE = Mean Relative Error

Videos	IF	MAE	MRE	Accuracy (%)
UMN	87	11.36	40.01	86
Crowd	36	8.09	30.45	91
S1.L1.13-57(1)	95	1.32	6.7	88
S1.L3.14-17(1)	73	1.05	8.01	92
S3.MF.12-43(6)	77	6.87	14.55	87
S3.MF.12-43(5)	97	17.75	32.13	85
S3.MF.14-13(1)	91	7.7	17.05	93
S3.MF.12-43(1)	77	9.01	60.67	89
S3.MF.14-13(1)	91	18.67	52.07	92
S3.MF.14-37(3)	23	0.65	13.07	84
S3.MF.12-43(7)	91	8.34	23.09	87
S3.MF.14-52(4)	84	9.66	32.12	90
S1.L2.14-06(1)	102	5.67	19.05	94
S1.L1.13-59(1)	112	1.4	10.71	93

Table 5 Performance measures using F- Measure (FM) for the proposed method

Video sequences	Input Frame	FD	HFDGMM	GMM	Proposed
UMN	87	80.25	78.98	83.89	89.45
Crowd2	36	61.6	67.88	85.49	94.47
S1.L1.13-57(1)	95	66.87	72.87	83.14	91.73
S1.L3.14-17(1)	73	75.52	77	81	89.46
S3.MF.12-43(6)	77	72.12	74.94	82.47	87.87
S3.MF.12-43(5)	97	77.54	76.88	85.36	92.37
S3.MF.14-13(1)	91	70.94	72.94	82.89	90.27
S3.MF.12-43(1)	77	78.89	80.35	82.8	88.27
S3.MF.14-13(1)	91	70.32	72.87	83.57	85.81
S3.MF.14-37(3)	23	77.23	78.68	76.37	85.66
S3.MF.12-43(7)	91	74.35	75.09	78.18	85.58
S3.MF.14-52(4)	84	71.24	74.88	83.26	89.9
S1.L2.14-06(1)	102	63.75	68.89	86.26	90
S1.L1.13-59(1)	112	73.33	75.94	81.12	83.43

Figures 3a, 4a and 5a are walking sequence input videos that are obtained from PETS2009 datasets with sparse, dense and medium crowd with different views and illumination conditions (low sunshine, bright sunshine and overcast). The proposed method generates accurately segmented multiple human and are shown in Figs. 3e, 4e and 5e. Figure 5a, the PETS2009 crowd formation input video has multiple flow gathering, waving trees and overcast conditions. The proposed method produces an accurate segmented result as shown in Fig. 5e compared with other techniques. Figures 5a and 6a shown very low crowd with different illumination conditions and the proposed method has accurately detected the multiple human. They are shown in Figs. 5e and 6e. While Fig. 6a PETS2009

Table 6 Performance measures using computational time (in secs) for the proposed method

Video sequences	Input Frame	FD	HFDGMM	GMM	Proposed
UMN	87	5.7	4.5	3.4	2.03
Crowd2	36	5.2	4.01	3.1	2.02
S1.L1.13-57(1)	95	5	4	3.01	2
S1.L3.14-17(1)	73	5.3	4.3	3.03	2.4
S3.MF.12-43(6)	77	4.9	4.06	3.13	2.32
S3.MF.12-43(5)	97	4.67	4.25	3.45	2.21
S3.MF.14-13(1)	91	5.32	4.43	3.23	2.4
S3.MF.12-43(1)	77	5.02	4.65	3.63	2.17
S3.MF.14-13(1)	91	4.73	4.23	3.34	2.43
S3.MF.14-37(3)	23	4.63	4.12	3.2	2.5
S3.MF.12-43(7)	91	4.75	4.21	3.25	2.24
S3.MF.14-52(4)	84	4.82	4.45	3.56	2.47
S1.L2.14-06(1)	102	4.55	4	3.42	2.48
S1.L1.13-59(1)	112	5	4.77	3.61	2.41

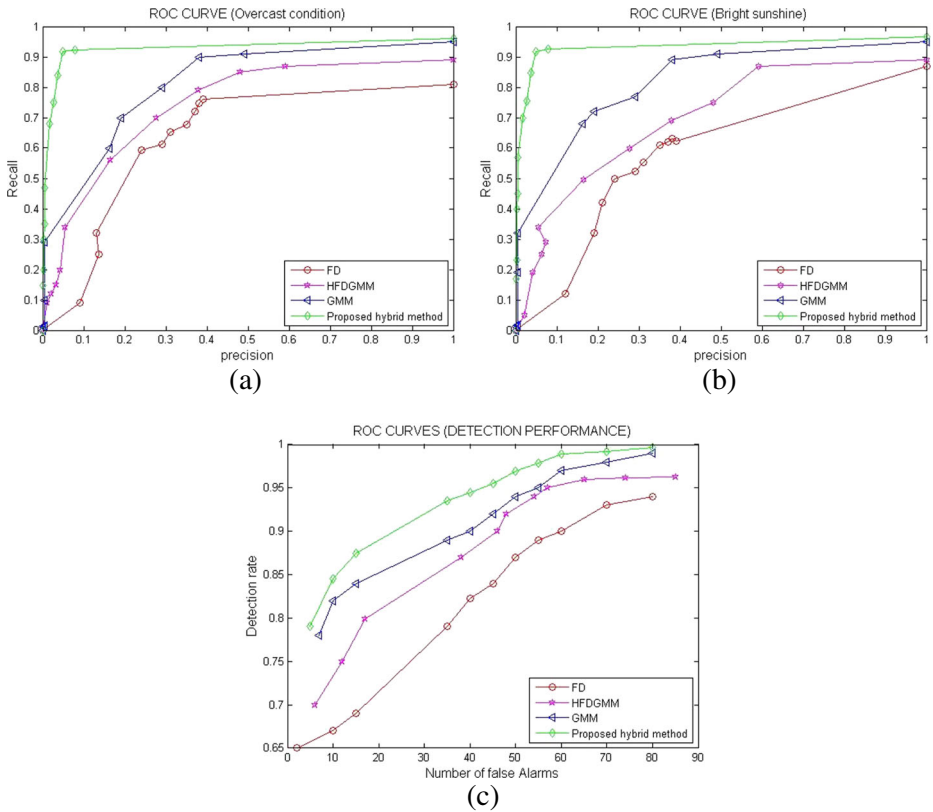


Fig. 7 (a, b and c) show the Quantitative comparisons of precision versus recall and detection rate proposed method to all the other methods

running sequence input video has a dense crowd with overcast conditions, the proposed method produces a good segmented result compared to other techniques and it is shown in Fig. 6e.

4.4 Quantitative evaluation

Figure 7a-b presents the Receiver Operating Characteristics curve of people in overcast and people in bright sunshine in a crowd sequence. Figure 7c shows that the proposed method has high detection rate compared to other methods. From the figure, a noticeable improvement is achieved to detect multiple human in a crowd video. From the quantitative analyzation [14, 18] view, the proposed hybrid method is analyzed using PETS2009 (benchmark) and UMN, crowd datasets which have a complex background and challenging sequences.

The statistical graph is obtained for different values of α to all the compared methods. It is shown in Fig. 8. In this work, the detection rates and error rates of the proposed work are compared under different α (learning parameter) values with accuracy and false positives rates for multiple human detection. Results in Fig. 8 shows that, the proposed method achieves low false positive rates while maintaining high detection accuracy for all α values.

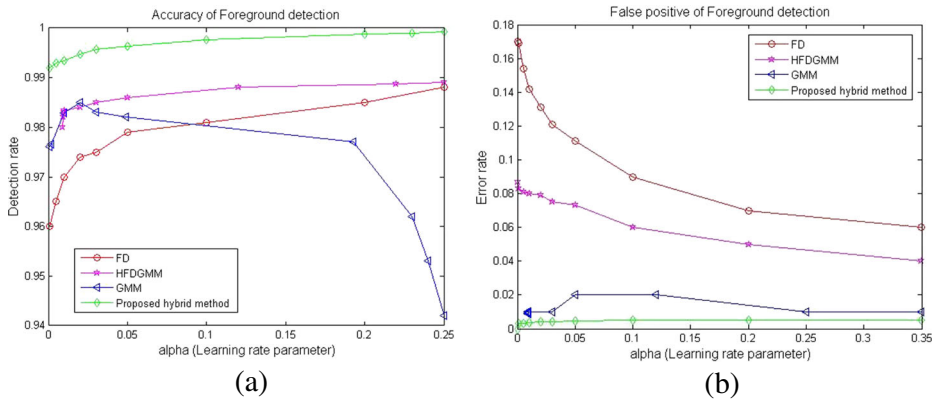


Fig. 8 Quantitative comparisons under different α values (a) comparisons of accuracy of foreground detection in multiple foreground detection. The detection level is of 95 % (b) comparisons of false positives rates in multiple foreground detection. The false positive level of is less than 1 % (Proposed hybrid GMM with IAKGMM)

The false positive rate measures the accuracy of the detection of multiple objects in a crowd scene. The false alarm rate measures the degree of false detection. The performances of the proposed methods are shown in Fig. 8a-b and it proves that the proposed method detects accurate multiple human in a crowd scene for different challenging conditions.

From the above figure, it is shown that the proposed hybrid method outperforms the traditional GMM and other existing methods [14, 18, 29] in terms of detection ratio, false alarm rate and accuracy level for detecting multiple human in a crowd scene. The proposed hybrid method performs noticeably when there is a quick, sudden illumination conditions and crowd condition. It indicates that the hybrid method has a significant performance for complex background and illumination changes in terms of false positive rates. From the above simulation results, it is observed that a static background needs only less number of Gaussian components whereas the adaptive and non-static background (complex) needs to be modeled by the mixture of multiple Gaussian components.

4.5 Inference

Automatic detection of multiple human in a crowded environment is very important in consumer applications like video analysis for human safety in public areas. So, the proposed work achieves an automatic multiple human detection in crowd videos using multi-camera surveillance datasets. Pixel-based performance metrics are provided using the PETS2009, UMN and crowd datasets.

The state-of-art evaluation tools are: precision (PRE), recall (REC), F-Measure (FM), Mean Absolute Error (MAE), Mean Relative Error (MRE) and accuracy. The proposed work has the advantages of different challenging conditions such as illumination changes and complex background. The proposed hybrid algorithm is adaptable to different conditions through adaptive thresholding, parameter settings, adaptive learning and update process for each background pixel as well as foreground pixel.

Adaptive and robust detection can be achieved based on adaptive mixture of Gaussian weights, and automatic selection of needed number of Gaussian components for dynamic background. From the simulation results, it is noted that the proposed frame work handles the complex background effectively using adaptive learning rates and insensitivity to

gradual and sudden illumination changes in the background. So, this is sensitive to slow variations and fast variations of motion in the video, thereby enabling to detect abnormal changes in foreground for surveillance applications.

The proposed hybrid method works well in a crowded environment. The performance analysis of the proposed background subtraction technique enables efficient foreground extraction in the crowd environment. The above techniques are analyzed with different indoor and outdoor datasets. The ROC curve is used to analyze the performance of the proposed method which gives the performance of the accurate foreground detection rate. True Positive defines how many correct positive samples occur between all positive samples available during the test. False Positive defines how many incorrect positive samples occur between all negative samples available during the test. From the ROC analysis, more TP samples give the accurate multiple foreground detection and more TN samples give the accurate background detection. From the visual superiority of Figs. 3–6, it is inferred that the proposed multiple human detection method works well in a crowded scene for dynamic background and different challenging conditions (lighting conditions, illumination changes and different viewpoints). The proposed method has less computational complexity with high speed compared to other methods in a multiple human detection.

From Table 3, the precision and recall values of the proposed method are found to have 23 %, 21 % and 8 % higher value than FD, HFDGMM and GMM techniques, respectively. It is proved that the true positive samples are more than the false positive samples and further the visual excellence of the result shows that the performance of the proposed method is better in crowd video. From the results in Table 4, the MAR, MRE and accuracy are more than 85 % most of the time with minimum error rates for the proposed method. In Tables 5 and 6, the proposed method obtains the best F-measure (FM) and reduction of computational time against all the other methods respectively. From the same table, it is evident that the proposed algorithm outperforms all the other algorithms for the given crowd scenes [6, 10, 12] with less computational time. From Fig. 7a-c, it can also be inferred that, the performance of the proposed method efficiently works in crowd videos. From Fig. 8a-b it is proved that the proposed approach always achieves low false positive rate (less than 1 percentage) with high detection accuracy for all α values. From the above observations, the proposed method have the following advantages,

- (i) The proposed hybrid algorithm have advantages of adaptive adjustment of learning rates and also adaptively select the needed number of K components and thresholding of each frame at different positions and properties of observed pixels.
- (ii) The advantages of the proposed method are quick updating of each pixel and selecting the needed number of Gaussian components for complex background [20]. The proposed hybrid algorithm is automatic and completely adapts to the scene.

5 Conclusion and future work

This work presents an accurate and automated multiple people detection framework in a crowd using a hybrid adaptive background subtraction method. The proposed work hybridises the concept of GMM approach with IAKGMM approach into a single framework. The proposed hybrid algorithm adapts to both sudden, gradual illumination as well as dynamic background changes. The proposed algorithm is compared with the state-of-the-art human detection methods in a crowd video for visual surveillance applications and it works better in dynamic background and different illumination changes.

The performance of the proposed method is evaluated and analyzed by using ROC analysis, MAE and MRE calculations. The proposed method is a hybrid of GMM output and the improved adaptive K GMM output, which produces an efficient multiple people detection in a crowded environment for solving challenges present in the PETS2009 datasets and UMN, crowd datasets. This proposed method is based on the spatial and temporal pixel variations in each and every frame, which are useful in tolerating the effect of challenging factors such as waving trees, crowd scenario, overcast conditions, natural lighting, and bright sunshine (sudden and gradual illumination changes). The proposed algorithm having the properties of adaptive learning control, adaptively changing weights and automatically selects the necessary number of K Gaussian components per pixel in order to handle a complex background as well as illumination changes.

The simulation results and the performance graphs show that the proposed method is flexible and has less computational time. The proposed method is analyzed and examined by the precision, recall, accuracy, f-measure, computational time, MAE, and MRE parameters. It is found that the proposed method exhibits better results compared to the existing methods. By using the proposed method, the multiple human detection rate is improved from 90 % to 95 % and the computational time is reduced from 5secs to 2.5secs than the existing methods in different practical situations. In future, an improved methodology may be adopted to enhance the performance in terms of increasing the detection rate and optimizing the computational time. The present technique can also be improved to handle other challenging conditions such as rain, overlapping and so on. Furthermore, by using FPGA/ARM based hardware implementation, this work may support real-time surveillance systems.

Acknowledgments Research and computing facilities was provided by the K.L.N. College of Engineering, Tamil Nadu, India.

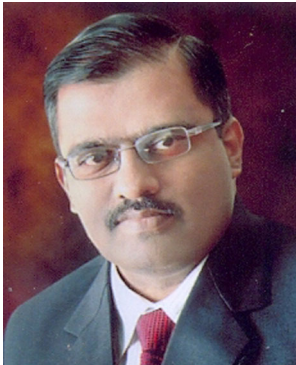
References

1. Bouwmans T (2014) Traditional and Recent approaches in Background Modeling for Foreground Detection: An Overview. *Comput Sci Rev* 11:31–66
2. Bouwmans T, El Baf F, Vachon B (2008) Background Modelling using Mixture of Gaussians for Foreground Detection: A Survey. *Recent Patents Comput Sci* 1(3):219–237
3. Eshel R, Moses Y (2008) Homography based multiple-camera detection and tracking of people in a dense crowd. In: *Proceedings of the conference on computer vision and pattern recognition*
4. Fazli S, Pour HM, Bouzari H (2009) A Novel GMM-Based Motion Segmentation Method for Complex Background. *The 5th IEEE-GCC Conference*
5. Ferryman J, Ellis AL (2014) Performance evaluation of crowd image analysis using the PET 2009 dataset. *Pattern Recognit Lett* 44:3–15
6. Hu Z et al (2011) Selective Eigenbackgrounds method for background subtraction in crowd scenes. *IEEE International Conference on image processing*, pp 3277–3280
7. Karpagavalli P, Ramprasad AV (2013) Estimating of the density of people and counting the number of people in crowded environment for human safety. In: *IEEE international conference on communications and signal processing (ICCSP)*, pp 663–667
8. Karpagavalli P, Ramprasad AV (2013) Human detection and segmentation in the crowd environment by combining APD with HLBD approaches. In: *National conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*, pp 1–4
9. Kentaro T, Krumm J, Brumitt B, Meyers B (1999) Principles and practice of background maintenance. In: *Proceedings of the Seventh IEEE international conference on computer vision (ICCV99)*, 1:255261
10. Kertesz C (2011) Texture-based foreground detection. *Int J Signal Process Image Process Pattern Recognit* 4:51–61
11. Kim Z (2008) Real time object tracking based on dynamic feature grouping with background subtraction. In: *Proceedings of IEEE conference on computer vision pattern recognition*, pp 1–8 (2008)
12. Kim K et al (2005) Real-time foreground-background segmentation using codebook model. *Real-Time Imaging* 11(3):172–185

13. Ko T, Soatto S, Estrin D (2008) Background subtraction on distributions. In: Computer vision ECCV 2008, pp 276289. Springer
14. Lee DS (2005) Effective gaussian mixture learning for video background subtraction. *IEEE Trans Pattern Anal Mach, Intell* 27(5):827–832
15. Li HH, Chuang JH, Liu TL (2011) Regularized background adaptation: A novel learning rate control scheme for gaussian mixture modeling. *IEEE Trans Image Process* 3(20):822–836
16. Mahadevan V, Vasconcelos N (2008) BGS in highly dynamic scenes. In: IEEE conference on computer vision and pattern recognition, pp 1–6
17. Richard J, Radke S, Al-Kofahi O, Roysam B (2005) Image change detection algorithms: a systematic survey. *IEEE Trans Image Process* 4(3):294–307
18. Stauffer C, Grimson WEL (2000) Learning patterns of activity using real-time tracking. *IEEE Trans Pattern Analysis and Mach Intell* 22(8):747–757
19. Sobral A, Vacavant A (2014) A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Comput Vis Image Underst* 122:4–21
20. Suo P, Wang Y (2008) An Improved Adaptive Background Modeling Algorithm Based on Gaussian Mixture Model. In: IEEE Proceedings of ICSP 2008
21. Sun L, Sheng W, Liu Y (2015) Background modeling and its evaluation for complex scenes. *Multimedia Tools Appl* 74:3947–3966
22. Tan R, Huo H, Qian J, Fang T (2006) Traffic video segmentation using adaptive-K Gaussian mixture model. In: The international workshop on intelligent computing (IWICPAS 2006), pp 125–134
23. Tian Y, Wang Y, Hu Z, Huang T (2013) Selective Eigenbackground for background modeling and subtraction in crowded scenes. In: IEEE transactions on circuits and systems for video technology
24. Van Droogenbroeck M, Paquot O (2012) Background Subtraction: Experiments and Improvements for ViBe. *Change Detection Workshop (CDW)*
25. Wang H, Suter D (2006) A novel robust statistical method for background initialization and visual surveillance. In: Computer vision ACCV 2006, pp 328337. Springer
26. Wu B, Nevatia R (2005) Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. In: Proceedings IEEE international conference on computer vision, vol 1, pp 90–97
27. Yao L, Ling M (2014) An improved mixture-of-Gaussians background model with frame difference and blob tracking in video stream. *Sci World J* 424050
28. Zhao L, Davis LS (2005) Closely coupled object detection and segmentation. In: Proceedings IEEE international conference on computer vision, pp 454–461
29. Zivkovic Z (2004) Improved adaptive Gaussian mixture model for background subtraction. In: Proceeding ICPR, 2004



Karpagavalli P. received her B.E. degree in Electronics and Communication Engineering from Madurai Kamaraj University in 1997 and M.E. degree in Communication Systems from Anna University Chennai in 2006. Now she is doing her PhD at Anna University Chennai, India. Currently she is working as an Associate Professor in Electronics and Communication Engineering department at K.L.N. College of Engineering, Tamil Nadu, India. She has published 12 papers in both International and National Conferences. She has published 4 National/International journals. Her research interest includes computer vision, video surveillance and object recognition. She is the life time member of ISTE. Email id: karpagavalli.p@klnce.edu.



Ramprasad A. V. obtained his B.E. at TCE, Madurai and M.E at ACCET, Karaikudi. He received Ph.D from Anna University Chennai. He has vast experience of 18 years in the area of Teaching / Research and Administration. He is working as Principal at K.L.N. College of Engineering, Tamil Nadu, and India. Totally he presented 13 numbers in International conferences and 15 numbers in National conferences. He has published 25 National/International journals. He has reviewer of the referred national / international journals like Journal of optics, international journal of image and graphics. He is the professional society lifetime member of ISTE / FIE / IEEE. He is acted as the chief coordinator for various funded projects from AICTE and CSIR.