Multi-Object Detection Using Modified GMM-Based Background Subtraction Technique

Rohini Chavan, S. R. Gengaje and Shilpa Gaikwad

Abstract Detection of objects is the most important and challenging task in video surveillance system in order to track the object and to determine meaningful and suspicious activities in outdoor environment. In this paper, we have implemented novel approach as modified Gaussian mixture model (GMM) based object detection technique. The object detection performance is improved compared to original GMM by adaptively tuning its parameters to deal with the dynamic changes that occurred in the scene in outdoor environment. Proposed adaptive tuning approach significantly reduces the overload experimentations and minimizes the errors that occurred in empirical tuning traditional GMM technique. The performance of the proposed system is evaluated using open source database consisting of seven video sequences of critical background condition.

1 Introduction

Human requirement of automated detection system in personal, commercial, industrial, and military areas leads to development of video analytics which will make lives easier and enable us to compete with future technologies [1]. On the other hand, it pushes us to analyze the challenges of automated video surveillance scenarios. Humans have an amazing capacity for decision-making but are notoriously poor at maintaining concentration levels. A variety of studies has shown that after 20 min of watching, up to 90% of the information being shown on monitors will be missed. As closed-circuit television (CCTV) culture continues to grow, humans

R. Chavan (✉) · S. Gaikwad
Bharati Vidyapeeth University College of Engineering, Pune, India
e-mail: chavanrohini10@gmail.com

S. Gaikwad
e-mail: spgaikwad@bvucoep.edu.in

S. R. Gengaje
Walchand Institute of Technology, Solapur, India
e-mail: srgengaje@rediffmail.com

D. Pandian et al. (eds.), *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, Lecture Notes in Computational Vision and Biomechanics 30,
https://doi.org/10.1007/978-3-030-00665-5_91

would require to observe feed from hundreds of camera 24 × 7 [2]. It shows requirement of automatic system that analyzes and stores video from 100s of cameras and other sensors, detecting events of interest continuously and browsing of data through sophisticated user interface. It is simply known as video analytics [3, 4].

Recent research in computer vision is giving more stress on developing system for monitoring and detecting humans. It is helpful for people in personal, industrial, commercial, and military areas to develop innovations in video analysis, to compete with future technologies and to accept the challenges in automatic video surveillance system. Video surveillance tries to detect, classify, and track objects over a sequence of images and help to understand and describe the object behavior by human operator. This system monitors sensitive areas such as airport, bank, parking lots, and country borders. The processing framework of an automated video surveillance system includes stages like object detection, object classification, and object tracking. Almost every video surveillance system starts with motion detection. Motion detection aims at segmenting regions of interest corresponding to moving objects from remaining image. Subsequent processes such as object classification and tracking performances are greatly dependent on it. If there is significant fluctuations in color, shape, and texture of moving object, it causes difficulty in handling these objects. A frame of video sequence consists of two groups of pixels. The first group represents foreground objects and second group belongs to background pixels. Different techniques such as frame differencing, adaptive median filtering, and background subtraction are used for extraction of objects from stationary background [5]. The most popular and commonly used approach for detection of foreground objects is background subtraction. The important steps in background subtraction algorithm are background modeling and foreground detection [6]. Background modeling gives reference frame which represents statistical description of entire background scene. The background is modeled to extract interested object from video frames. It is designed with first few frames of video sequence. But, in case of quasi-stationary background such as wavering of trees, flags, and water, it is more challenging to extract exact moving object. In this situation, single model background frame is not enough to accurately detect the moving object but adaptive background modeling technique is used for exact detection of objects from dynamic background [7].

2 Object Detection Using Adaptive Gaussian Mixture Model

2.1 Basic Gaussian Mixture Model

A Gaussian mixture model (GMM) is parametric probability density function presented as a weighted sum of K Gaussian component densities [8, 9]. It is represented by the following equation:

$$P(x_t) = \sum_{i=1}^{k} \left(\omega_{\mathrm{i,t}}\, n(X_t; \mu_{i,t}, \sum i, t)\right) \tag{1}$$

$$\sum_{i=1}^{k} \left(\omega_{\mathrm{i,t}}\right) = 1 \tag{2}$$

where x is a D dimensional data vector and ω is weight of i^{th} Gaussian component. Here, k is the number of Gaussian distributions, t represents time, μ is mean value of the ith Gaussian mixture at time t, and $\sum i, t$ is the covariance matrix. The entire GMM is scaled by mean vectors, covariance matrices, and mixture weights of all component densities. The mean of such mixture is represented by following equation:

$$\mu_t = \sum_{i=1}^{k} \omega_{i,t}\mu_{i,t} \tag{3}$$

There are several variants on the GMM and covariance matrices constrained to be diagonal. The selection of number of components and full or diagonal covariance matrix is often determined by the availability of data for estimating GMM parameters.

Background subtraction object detection technique is popular as it is less complex, simple and easy to implement. It takes the difference between current frame (I_t) and reference frame. The reference frame is denoted by (B_{t-1}). Hence, difference image (D_t) is given by

$$D_t = |B_{t-1} - I_t| \tag{4}$$

Foreground mask (F_t) is given by applying threshold to difference image

$$F_t = 1,\ \text{when}\ D_t > \mathrm{Th}$$
$$F_t = 0,\ \text{when}\ D_t < \mathrm{Th}$$

2.2 GMM Model Initialization and Maintenance

For stationary process pixels, EM algorithm is applicable. K-means algorithm is an alternative to EM [10]. Using K-means approximation, every new pixel value X_t is checked against existing K Gaussian distribution until match is found. A match is given by

$$\mathrm{Sqrt}\left(\left(X_{t+1} - \mu_{i,t}\right)T.\sum\nolimits_{i,t}^{-1}\left(X_{t+1} - \mu_{i,t}\right)\right) < k\sigma_{i,t} \tag{5}$$

where k is constant threshold value which is selected as 2.5. If K distribution is not matched with current pixel value then least probable distribution is replaced with current distribution value as its mean, weight, and variance. Prior weights of K distributions at time t are adjusted as follows:

$$\omega_{k,t} = (1-\alpha)\omega_{k,t} - 1 + \alpha(M_{k,t}) \tag{6}$$

where α is learning rate and $M_{k,t}$ is 1 for model which is matched and 0 for other models. After operating this approximation, weights are again normalized. The μ and σ parameters remain same for unmatched distributions. The parameters of distribution which matches new observations are updated as follows:

$$\mu_{\mathrm{t}} = (1-\rho)\mu_{\mathrm{t}} - 1 + \rho X_t \tag{7}$$

$$\sigma t = (1-\rho)\sigma 2t - 1 + \rho(X_{\mathrm{t}} - \mu t)T(X_t - \mu t) \tag{8}$$

$$\sigma_t = (1-\rho)\sigma_{\mathrm{t-1}} + \rho(X_{\mathrm{t}} - \mu t)^{\mathrm{T}}(X_{\mathrm{t}} - \mu t) \tag{9}$$

where

$$\rho = \alpha . n(X_{\mathrm{t}} | \mu_{\mathrm{k}}, \sigma) \tag{10}$$

One advantage of this technique is that when new thing is added in the model then it will not completely destroy the previous background model but it can update the model.

3 Object Detection Using Adaptive GMM

The system is implemented using two steps such as GMM-based object detection and noise removal using morphological operations. Implementation is done using MATLAB 2014v with the help of computer vision system toolbox. The small detected regions whose area is less than moving object and which are not part of foreground object can be removed using noise removal algorithm. Finally, output binary image is compared with ground truth image for performance evaluation to determine accuracy.

Background modeling is adaptive to accommodate all the changes occurring in the background scene. It is very sensitive to dynamic changes that have occurred in the scene which causes consequent need of adaptation of background as per the variations in background. The research has progressed toward improving robustness and accuracy in background subtraction method for complex background condition like sudden and slow illumination change. A common attribute of BS algorithm is learning rate, threshold, and constant parameter K which can be empirically adjusted to get desired accuracy. However, tuning process for these parameters has been less attentive due to lack of awareness. Stauffer and Grimson [7] suggested that selection of learning rate and threshold value is important among all other parameters. Tuning process for these parameters requires time intense repeated experimentation to achieve optimum results. It is very challenging to set the parameters because it requires understanding of background situation and common setting for different scenarios may not produce accurate result. All these aspects put limitations on effective use of background subtraction algorithm and demand improvement and extension of original GMM.

Recent years, researchers are focused on developing innovative technology to improve performance of IVS in terms of accuracy, speed, and complexity. To design novel approach for GMM parameter tuning based on extraction of statistical features and map with GMM training parameters [11]. Learning rate parameter is very important which determines the rate of change of background. Large amount of experimentation is required to set the value of learning rate for exact detection of foreground object. It is required to develop the system which tunes the parameter automatically for satisfactory performance of GMM.

GMM modeling is able to handle multimodal background scene. Performance of GMM-based background subtraction is decided by pixel-wise comparison of ground truth and actual foreground mask. Performance of the system is evaluated with the help of primary metrics such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) and secondary metrics like sensitivity, accuracy, miss rate, recall, and precision. Precision reflects false detection rate and recall gives accuracy of detection. Precision and recall are the two important measures in order to estimate detection algorithm systematically and quantitatively [12, 13].

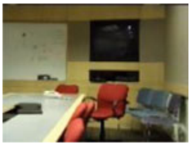
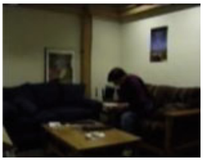
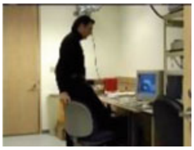
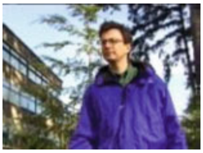
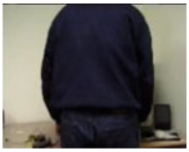
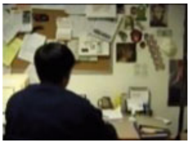
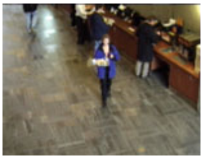
$$\text{Precision}(\%) = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100 \tag{11}$$

Our proposed system brings innovation in original GMM-based object detection system through tuning and adaptation of important parameters such as number of component, learning rate, and threshold.

3.1 Video Database

Wallflower database is open source database [9]. It includes seven sets of video sequence with different critical situations in background. Video frames of size 160 × 120 pixel, sampled at 4 Hz. Data set provider also gives one ground truth image and text file having description of all video sequences. Ground truth is binary image representing foreground mask of specific frame in video sequence. Table shows all test sequences along with their ground truth (Table 1).

Table 1 Wallflower dataset of seven different video sequences along with ground truth

Sr. No.	Name	Test sequence	Ground truth
1	Moved object (MO)		
2	Time of day (TOD)		
3	Light switch (LS)		
4	Waving tree (WT)		
5	Camouflage (C)		
6	Foreground aperture (FA)		
7	Bootstrap (B)		

3.2 Experimental Setup

The main focus of research is based on appropriate selection of GMM training parameters like K, α, and T. The selection of K Gaussian component value is function of complexity of background scene. If the background is simple and unimodal, the value of K must be selected as 1 or 2. For complex multimodal background, value of K is more than 2 and less than 5 so as to improve the accuracy in detection process. Various pairs of α and T are evaluated on Wallflower dataset. After lots of experimentation best pair of α and T is identified based on performance analysis on various Wallflower videos [9]. Parameter initialization, training, and testing are the three important steps for object detection process.

3.3 Parameter Initialization

Object detection system includes various GMM parameters like number of training frames, initial variance, and training parameters (K, α, and T). They are initialized as follows:

Number of training frames: 200 (given by data set provider),
Number of component: 4,
Initial variance: 0.006, and
Threshold: Adjusting value empirically (0.5, 0.6, 0.7, 0.8, 0.9).

4 Experimental Results

GMM-based object detection system is evaluated using various settings of α and T for each sequence. After this experimentation, for all videos, appropriate setting of α and T is decided based on lowest value of total error. Performance metrics are calculated for each sequence by comparing detected mask with ground truth.

Results are as follows (Fig. 1):

GMM-based background subtraction technique gives best overall detection performance at $\alpha = 0.001$ and $T = 0.9$. These parameter settings improve the accuracy of foreground mask which is almost matching with ground truth. Learning rate and threshold have enough power to tune object detection performance. Best overall performance setting has less probability to give best result at individual level. Best individual performance may be obtained by different settings of parameters for some of the sequences. Performance analysis for best α and T can be done at pixel level. Empirically selection of higher threshold value gives merging of foreground objects with background. It leads to increase in false negative and decrease in true positive. For faster changing background, empirical selection of lower value of α is too low to adapt such background changes. Empirical setting of threshold value, $T = 0.9$, is

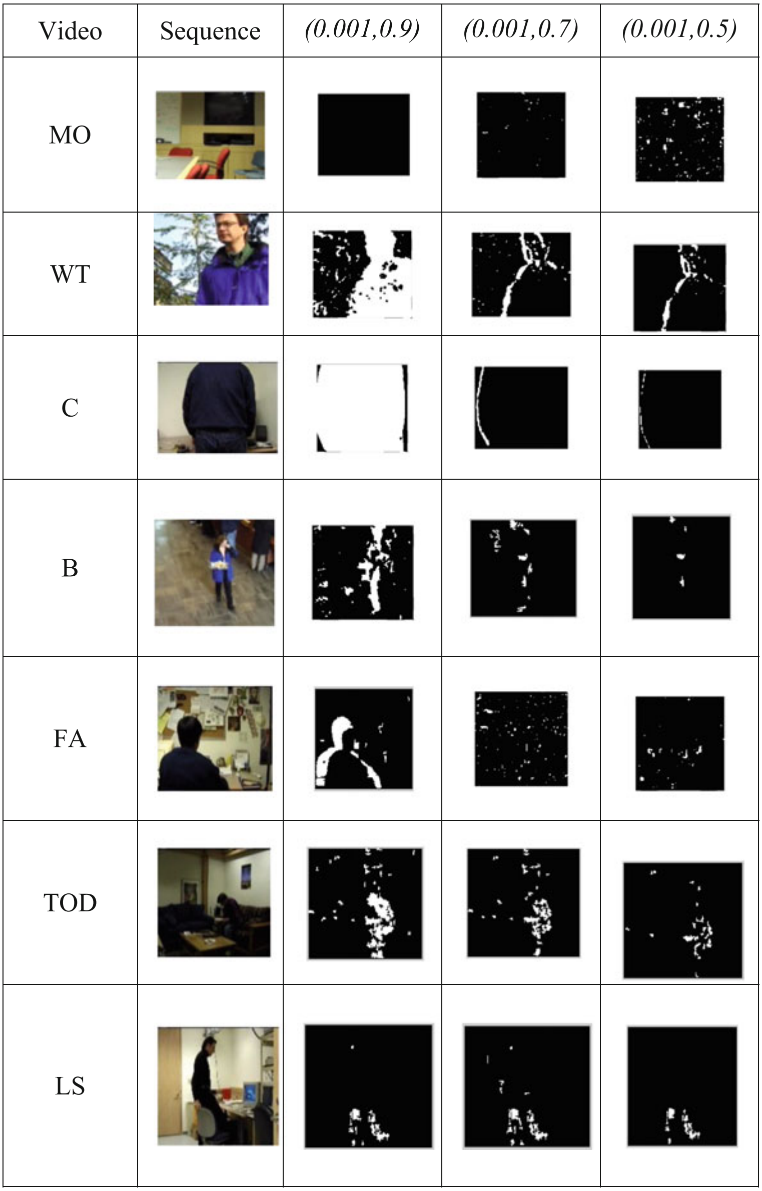


Fig. 1 Foreground mask obtained using GMM for different values of α and T

being so high that all foreground pixels are merged into the background. It gives an increase in false negative and decrease in true positive pixels. Same way, empirical selection of learning rate, $\alpha = 0.001$, is being too low for rapid changing background. Thus, misclassification is higher and accuracy is lower for those video sequences in

Table 2 Performance evaluation of proposed system on Wallflower dataset

Video	True positive (TP)	True negative (TN)	False positive (FP)	False negative (FN)	Precision (%)	Accuracy (%)
MO	0	19,200	0	0	100	100
WT	12,558	766	306	4570	90	80.34
C	3399	1039	770	8231	81.53	83.68
B	2546	15,927	295	432	89.61	85.49
TOD	125	17,762	0	1313	90.16	91.43
FA	3754	13,718	745	983	83.44	84.69
LS	718	12,670	3359	2453	69.72	77.87

which sudden change of illumination is occurring. This experimentation also suggests that various settings of (α and T) may result in more improved performance for different video sequences than fixed selection of (α and T) (Table 2).

5 Conclusion

Proposed research emphasizes on proper tuning of important GMM parameter leading to improvement in the performance accuracy of GMM-based object detection system. GMM parameters mainly include number of mixture component (K), learning rate (α), and threshold (T). We have implemented two approaches for tuning these parameters such as traditional empirical tuning and automated adaptive tuning based on background dynamics. Traditional empirical tuning method is implemented using different settings of α and T, while K is kept constant to high value for complex scene. After large number of experimentation, appropriate pair of α and T is selected based on low performance error.

Proposed adaptive tuning method involves adaptation of α and keeping T and K constant to appropriate value. Unique EIR concept is used to extract background dynamics for current frame. Learning rate is tuned depending on EIR. This modified approach improves the result of GMM compared to the original GMM. This result strongly emphasizes the strength of learning rate adaptation. Performances of GMM with these tuning methods are evaluated based on foreground mask obtained using GMM and ground truth image in database. The analysis of performance can be done using primary metrics such as TP, TN, FP, and FN as well as secondary metrics like precision and accuracy. Our proposed system performance is compared with traditional empirical method and other existing techniques. Our research is implemented on MATLAB 2014 platform. Different functions from MATLAB computer vision toolbox are used for implementation of algorithm.

References

1. Hsieh J-W, Yu S-H, Chen Y-S (2006) An automatic traffic Surveillance system for vehicle tracking and classification. IEEE Trans Intell Transp Syst 7
2. Chauhan AK, Krishan P (2013) Moving object tracking using Gaussian mixture model and optical flow. Int J Adv Res Comput Sci Softw Eng 3
3. Picardi M (2004) Background subtraction technique: a review. In: IEEE international conference on systems, man and cybermetics
4. Viola P, Jones M (2005) Detecting Pedestrians using patterns of motion and appearance. Int J Comput Vision 63(2):153–161
5. Bo W, Nevatia R (2007) Detection and tracking of multiple partially occluded Humans by Baysian combination of edgelet based part detector. Int J Comput Vision 75(2):247–266
6. Ran Y, Weiss I (2007) Pedestrian Detecting via periodic motion analysis. Int J Comput Vision 71(2):143–160
7. Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real time tracking. In: International conference on computer vision and pattern recognition 2
8. Cucchiaria R, Grana C, Piccardi M, Prati A (2003) Detecting moving objects, ghosts and shadows in video streams. IEEE Trans PAMI 25(10):1337–1342
9. Toyama K, Krumm J, Brumitt B (1999) Wallflower: Principles and practice of background maintenance. In: International conference of computer vision, pp 255–261
10. Zhang, LZ, Hou Z, Wang H, Tan M (2005) An adaptive mixture Gaussian background model with online background reconstruction and motion segmentation. ICIT, pp 23–27
11. White B, Shah M (2007) Automatically tuning background subtraction parameters using Particle swarm optimization. In: IEEE international conference on multimedia and Expo, China, pp 1826–1829
12. Harville M, Gordon G, Woodfill J (2001) Foreground segmentation using adaptive mixture models in color and depth. In: Proceeding of the IEEE workshop on detection and recognition of events in Video, Canada
13. Elgammal A, Harwood D, Davis L (2000) Non parametric model for background subtraction. In: European conference on computer vision, pp 751–767