J. Intell. Syst. **20** (2011), 33–45
DOI 10.1515/JISYS.2011.003

SiMOR: Single Moving Object Recognition

M. T. Gopala Krishna, M. Ravishankar, D. R. Ramesh Babu and
V. N. Manjunath Aradhya

Abstract. Automatic moving object detection and tracking is very important task in video surveillance applications. In the present work the well known background subtraction model and use of Gaussian Mixture Models (GMM) have been used to implement a robust automated single object tracking system. In this implementation, background subtraction on subtracting consecutive frame-by-frame basis for moving object detection is done. Once the object has been detected it is tracked by employing an efficient GMM technique. After successful completion of tracking, moving object recognition of those objects using well known Principal Component Analysis (PCA), which is used for extracting features and Manhattan based distance metric is used for subsequent classification purpose. The system is capable of handling entry and exit of an object. Such a tracking system is cost effective and can be used as an automated video conferencing system and also has applications like human tracking, vehicles monitoring, and event recognition for video surveillance. The proposed algorithm was tested on standard database on complex environments and the results were satisfactory.

Keywords. Object detection, tracking and recognition, Gaussian mixture model, PCA, surveillance, video conferencing system.

2010 Mathematics Subject Classification. 68U10.

1 Introduction

Moving Object detection and tracking are becoming increasingly recognized as important capabilities in any vision System designed to operate in an uncontrolled environment. Tracking objects can be complex [10] due to loss of information caused by projection of the 3D world on a 2D image, noise in images, complex object motion, non rigid or articulated nature of objects [8], partial and full object occlusions, complex object shapes, scene illumination changes, and real-time processing requirements. Tracking is made simple by imposing constraints on the motion and/or appearance of objects. In our application the number of constraints on the motion and appearance of the object is minimized. The only constraint on the motion of the object is that it should not make sudden change in the direction of motion while moving out of the viewing range of the camera. Unlike other algorithms [4] the present algorithm is capable of handling the entry and exit of an object. Also, unlike [4, 8], no color information is required for tracking an object. There is no major constraint on the appearance of the object though an object which is a little brighter than the background gives better tracking results. Such a tracking system is cost effective and can be used as an automated video conferencing system, and where there is single entry and exit of single moving Vehicles in Commercial Super Malls, Super Markets in the parking zone etc. and also has application as a surveillance tool.

As per the literature survey, many of these algorithms are susceptible to both global and local illumination changes such as shadows and highlights. This will cause the consequent processes of tracking, recognition, and also affect the accuracy and efficiency of the moving object. This problem is the underlying motivation of proposed work. A robust and efficiently computed background subtraction algorithm that is able to cope with the local illumination change problems, such as shadows and highlights, as well as the global illumination changes is considered.

Much research work into moving object recognition has been previously presented. In [3], a Gramian determinant-based method was proposed to detect moving objects between multiple images and to detect changes between color images or any type of multi-spectral images. In [6], a moving object detection method was proposed using global motion estimation and edge information, where the final objects are extracted by combining the contours and moving regions from motion detection. In [2], a method based on background subtraction was proposed to recognize abnormal human behavior in public areas using segmentation of moving objects in real time from images acquired by a fixed color video camera. In [Huawei et al., 2005], a system for moving object detection and shadow extermination by building an adaptive background model was described. In [5], a method for detecting poorly visible moving objects in bad weather was proposed, and was based on measuring a variation on a cross correlation between a short accumulated histogram and along accumulated histogram of detected objects.

In this paper, a novel algorithm for detecting moving objects from a static background scene that contains shading and different illumination is proposed. The proposed method introduces a new Background Subtraction model that helps to detect moving objects from the ordinary background or moving foreground objects. Next, another algorithm is proposed for tracking moving objects in the scene. The proposed method is restricted to indoor and outdoor environments with static background, and the algorithm works fairly well on real image sequences of outdoor scenes. Finally recognizing objects has been done using well known subspace method for extracting feature and nearest neighbor classifier is used for subsequent recognition purpose. This paper is organized as follow: Section 2 discusses about background subtraction process for detecting, GMM for tracking and recognition of moving object. Section 3 discusses the result and discussion. Finally conclusions are drawn at the end.

2 Proposed Method

In this section, the concept of background subtraction, Gaussian Mixture Model (GMM) and Principal Component Analysis (PCA) are briefly discussed.

2.1 Background Subtraction Process

To achieve moving object detection, there are many methods or models which are widely used by researchers. In which, most widely used is background subtraction method. Background subtraction model works efficiently due to surveillance cameras which are fixed and these are monitored toward a fixed direction. The process involves by taking the difference between the consecutive frames. The output obtained using the background subtraction model will contain only the moving object without any background, where subtraction value is compared with threshold which is empirically fixed. Figure 1 shows the object detection after background subtraction and it is noticed that the frame contains unwanted.

2.2 Gaussian Mixture Models (GMM) based Approach

GMM is a simple linear superposition of Gaussian components, aimed at providing a richer class of density than a single Gaussian [1]. The formulation of Gaussian mixtures will provide us with a deeper insight into this important distribution and

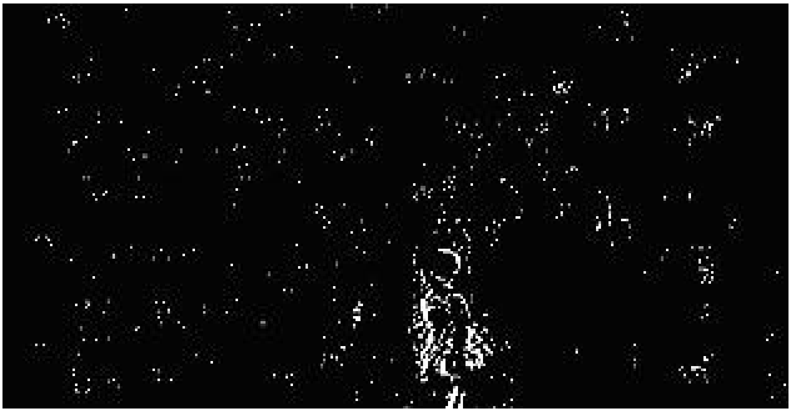

Figure 1. Object detection after background subtraction process with noise.

Figure 2. Resultant frame after applying median filter.

will serve to motivate the expectation-maximization algorithm. A distribution can be written as a linear superposition of Gaussian in the form:

$$P(x) = \sum_{k=1}^{K} \pi_k \eta(x/\mu_k, \Sigma_k) \tag{1}$$

which is called a mixture-of-Gaussians, where $\eta(x/\mu_k, \Sigma_k)$ is the multivariate Gaussian Distribution of the form:

$$\eta(x/\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-1/2(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)}. \tag{2}$$

Each Gaussian density $\eta(x/\mu_k, \Sigma_k)$ is called a component of the mixture and has its own mean μ_k and covariance Σ_k. The parameter π_k in Equation (1) is called mixing coefficient. If we integrate both sides of Equation (1) w.r.t x, both $P(x)$ and the individual Gaussian components are normalized, we obtain $\sum_{k=1}^{K} \pi_k = 1$. Also, the requirement that $P(x) \geq 0$, together with $\eta(x/\mu_k, \Sigma_k) \geq 0$, implies $\pi_k \geq 0 \; \forall k$. Combining this with Equation (1) we obtain $0 \leq \pi_k \leq 1$.

From the sum and product rules, the marginal density is given by

$$P(x) = \sum_{k=1}^{K} P(x)P(x/K) \tag{3}$$

which is equivalent to Equation (1) in which we can view $\pi_k = P(k)$ as the prior probability of picking the k^{th} component, and the density $\eta(x/\mu_k, \Sigma_k) = P(x/K)$ as the probability of x conditioned on k.

From Bayes' theorem the posterior probabilities $P(K/x)$, which are also known as responsibilities, are given by:

$$\gamma(Z_k) \equiv P(K/x) \tag{4}$$

$$= \frac{P(x)P(x/K)}{\sum_l P(l)P(x/l)} \tag{5}$$

$$= \frac{\pi_k \eta(x/\mu_k, \Sigma_k)}{\sum_l \eta(x/\mu_l, \Sigma_l)} \tag{6}$$

The form of the Gaussian mixture distribution is governed by the parameters π, μ and Σ, where we have used the notation $\pi \equiv \pi_1, \pi_2, \pi_3, \ldots, \pi_K$, $\mu \equiv \mu_1, \mu_2, \mu_3, \ldots, \mu_K$ and $\Sigma \equiv \Sigma_1, \Sigma_2, \Sigma_3, \ldots, \Sigma_K$. We now adapt an iterative algorithm, known as Expectation Maximization (EM) algorithm, to estimate the values of Σ, μ and π. We first choose some initial values for these parameters by running K-means clustering algorithm. Then we alternate between two steps known as Expectation (E) Step and the Maximization (M) step to update the values of these parameters until convergence criteria is reached. The EM algorithm for GMM2 can be summarized as follows:

Algorithm for EM:

1. Initialize the parameters μ_k, Σ_k and π_k by running K-means clustering algorithm and Evaluate the log of the likelihood function using Equation (7)

2. E Step Evaluate the responsibilities using Equation (6) with current parameter values.

3. M Step Re-estimate the parameters using the current responsibilities:

$$\mu_k^{\text{new}} = \frac{1}{N} \sum_{n=1}^{N} \gamma(Z_{nk}) u_n$$

$$\Sigma_k^{\text{new}} = \frac{1}{N} \sum_{n=1}^{N} \gamma(Z_{nk})(A)(A)^T$$

$$\pi_k^{\text{new}} = \frac{N_k}{N}.$$

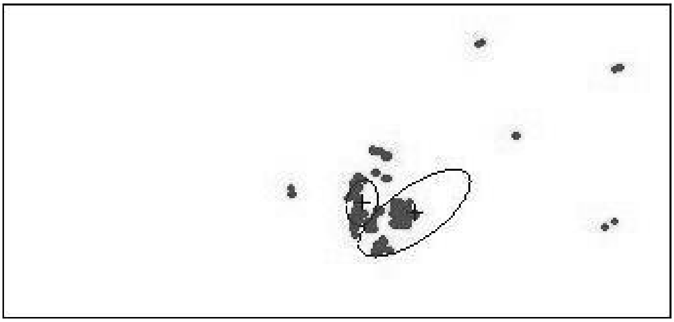

Figure 3. Illustration of Gaussian Mixture Models for Input image sequence of Figure 2.

4. Evaluate the log likelihood:

$$\ln P(X/\mu, \Sigma, \pi) = \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_k \eta(x_n/\mu_k, \Sigma_k) \tag{7}$$

and check for convergence of log likelihood. If the convergence criterion is not Satisfied iterate from step 2.

Using the obtained means of k clusters μ_k, $\forall k = 1, \ldots, K$, only 2 mixtures of Gaussians in order to track the object is used. Figure 3 depicts the mean points obtained for the input image sequence using mixture-of-Gaussians. In order to take the positional information of the moving object, the average of two mean points obtained from mixture model is computed. This information is then used to track (rectangle) the moving object.

2.3 Principal Component Analysis (PCA)

The main objective of subspace based approach is to project the data of objects onto a dimensionally reduced space where the actual recognition will be carried out. Turk and Pentland in 1991 first explored the (PCA) [9] for face recognition and used the PCA projected components as the features. More formally, it can be shown as follows: Let there be N object images $(A_1, A_2, \ldots, A_N)$ constituting the training set denoted by m by n matrices. Now the average matrix $\overline{A}$ of all training samples has to be calculated and then subtracted from the original objects A_i and the result is stored in Φ_i:

$$\overline{A} = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{8}$$

$$\Phi_i = A_i, \ldots, \overline{A}. \tag{9}$$

In the next step, the covariance matrix C is calculated as follows:

$$C = \frac{1}{N}\sum_{i=1}^{N} \Phi_i \Phi_i^T. \tag{10}$$

Now the eigenvalues λ_i $(i = 1, \ldots, s)$ and the corresponding eigenvectors U_i $(i = 1, \ldots, s)$ are calculated. Out of s eigenvectors, d of them corresponding to largest eigenvalues are used for feature extraction. The higher the eigenvalue, the more characteristic features of an object does the particular eigenvector describe. Finally, feature extraction is done as follows:

$$F^k = U^T(A_k - \overline{A}) \quad \forall k = 1, \ldots, N. \tag{11}$$

Given a test object image I, use Equation (5) to get the feature matrix $F = U^T \times I$, then a nearest neighbor classifier can be used for subsequent classification. In this work a well known Manhattan distance metric is used for classification purpose and it is defined as: $\mathrm{dist}(F, F_k) = \|F - F_k\|$.

3 Results and Discussion

In this section, the proposed method with a dataset containing seven standard benchmark suites of indoor and outdoor video sequences is experimentally evaluated. We also created our own dataset under different environmental conditions. All our experiments are carried out on a PC machine with P IV, 2.0 GHz CPU and 1 GB RAM memory under Matlab 7.0 platform.

Information about the above said dataset is depicted in Figure 4. The sequences reported in Figure 4, where both indoor and outdoor sequences are present are used in our proposed experiment study. The sample video sequences show an intensity at brighter light conditions, terrace video sequence was recorded in clouding conditions which show at very low light condition and corridor video sequence was recorded in two different light conditions. The campus sequence is a noisy sequence from outdoor campus site where vehicles and students are walking around the campus. The two indoor sequences report two laboratory rooms in two different perspectives and light conditions. In the laboratory sequence, besides walking people, movement of chair is also considered.

	Ball on Ground	Ball Hanging	Sample	Terrace	Corridor
					
Sequence type	indoor	indoor	outdoor	outdoor	outdoor
Sequence length	72	50	80	350	1100
Image size	320x240	320x240	320x240	360x240	320x240
Shadow strength	medium	high	low	very low	low
Object class	Ball	Ball	people	people	people
Object size	small	small	medium	medium	Medium
Noise level	Very High	Very High	High	High	Low

	Campus I	Laboratory I	Laboratory II	Intelligent room
				
Sequence type	outdoor	outdoor	indoor	indoor
Sequence length	2230	70	987	900
Image size	320x240	320x240	320x240	320x240
Shadow strength	medium	high	Very Low	Low
Object class	Vehicles/P	People	people/other	people
Object size	large	medium	medium	Medium
Noise level	High	medium	low	medium

Figure 4. Information about the dataset.

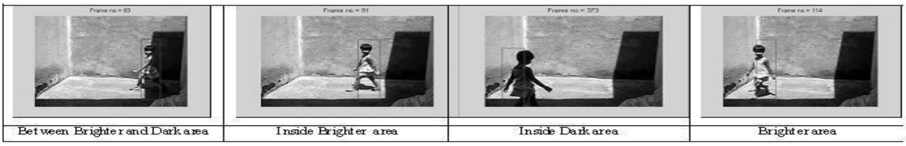

Figure 5. Video under different lighting condition.

Figure 6. Video under occlusion.

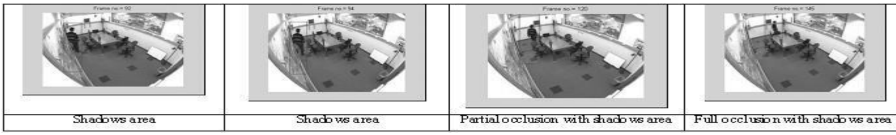

Figure 7. Video under shadow and occlusion.

Figure 8. With noisy environment.

Figure 9. Cloudy environment.

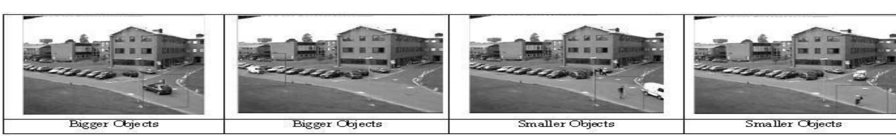

Figure 10. Noisy and complex background.

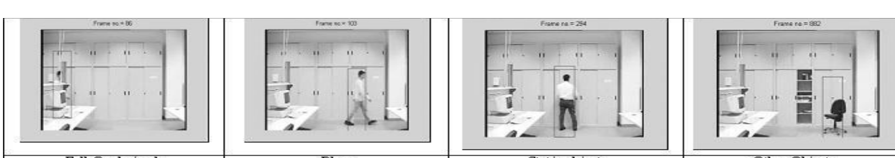

Figure 11. Occlusion in indoor environment.

Figure 12. Recognizing object as pedestrian.

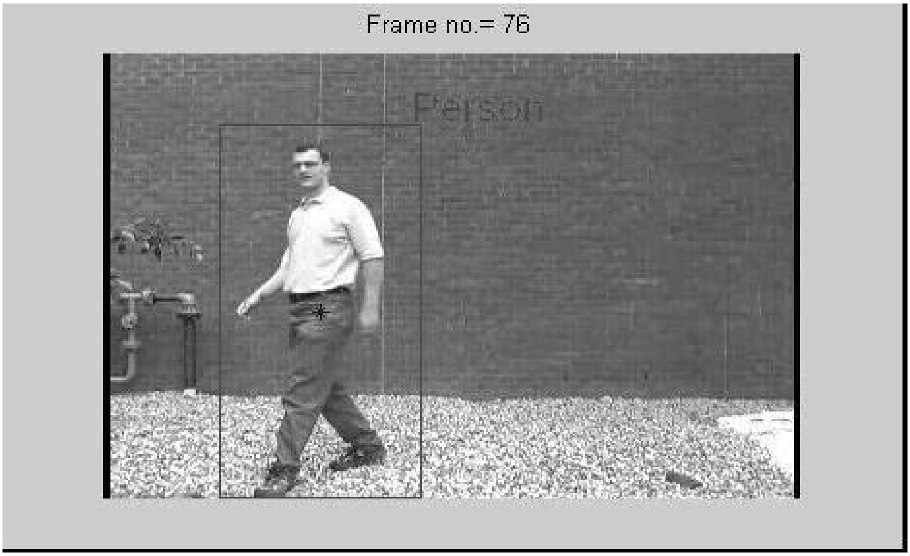

Figure 13. Recognizing object as pedestrian.

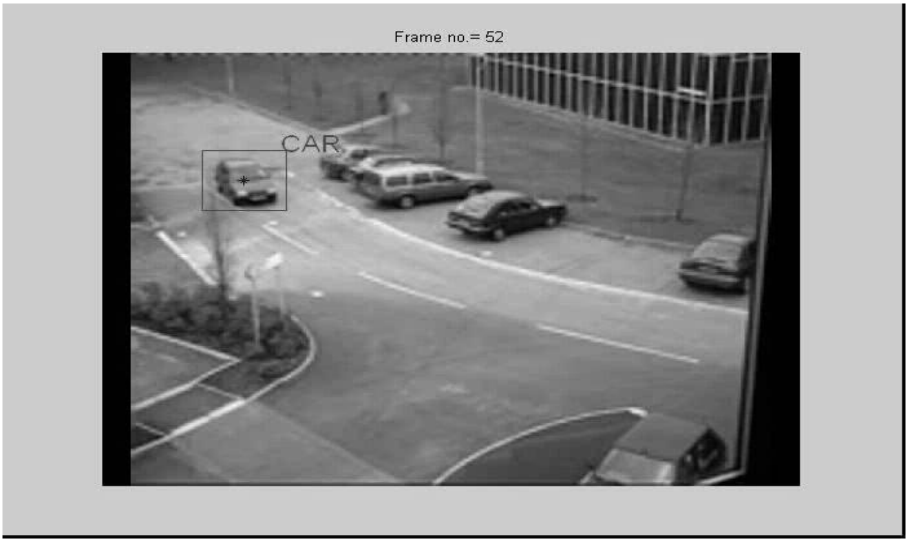

Figure 14. Recognizing object as car.

Figure 15. Recognizing object as car.

Figures 5 to 11 show the single object tracking under different environmental conditions obtained from the proposed method. From the figures it is ascertained that the proposed method based on GMM performs well under the conditions given above. The main advantage of the proposed method is it clusters the exact moving object even if there is some noise in the image sequence (refer Figure 3). For recognition purpose, we considered two main objects, one is pediatrician and car. We trained varying number of images considering pediatrician and car and for testing purpose, real time video was considered. Figures 12 to 15 show the result of recognizing individual objects. From the results it was clear that performance of the system was satisfactory under real time complex video.

4 Conclusions

A robust and efficient automated single object recognition system is presented. The proposed method is based on the concept of GMM for tracking and PCA for recognition. First background subtraction method is applied for detecting the moving object and then GMM features are used for tracking the objects. The algorithm has experimentally shown to be quite accurate and effective in detecting a single moving object even under bad lighting conditions or occlusions or shadow regions. The system was tested on different standard dataset and demonstrated the effectiveness of the algorithm. Finally recognition of object is also carried using well known PCA technique and performance of the system was satisfactory. Future work focuses on tracking multiple objects at the same time as well as on improving tracker accuracy during camera motion.

Bibliography

[1] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[2] D. Duque, H. Santos and P. Cortez, Moving object detection unaffected by cast shadows, In: Proceedings of IEEE Int. Conf. *Image Processing*, 2005, pages 413–416.

[3] E. Durucan and T. Ebrahimi, Moving object detection between multiple and color images, In: IEEE Conf. *Advanced Video and Signal Based Surveillance*, 2003, pages 243–251.

[4] P. Fieguth and D. Terzoplulos, Color-based tracking of heads and other mobile objects at video frame rates, In: IEEE Conference on *Computer Vision and Pattern Recognition*, 1997, pages 21–27.

[5] W. Huawei, B. Jianhua, L. Cuihua and W. Qiying, A system for moving object detection and shadow extermination, In: IEEE Int. Workshop *VLSI Design and Video Technology*, 2005, pages 117–120.

[6] R. Jinchang, P. Astheimer and D. D. Feng, Real-time moving object detection under complex background, In: IEEE 3rd Int. Symp. *Image and Signal Processing and Analysis*, 2003, pages 662–667.

[7] K. Onoguchi, Moving object detection using a cross correlation between a short accumulated histogram and a long accumulated histogram, In: *IEEE 18th International conference*, 2006.

[8] Y. Richard, D. Xu, G. A. John and S. J. Jesse, Robust real time tracking of non-rigid objects. *workshop on Visual information processing*, In: Pan-Sydney area, 2004, pages 95–98.

[9] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, **3** (1991), 71–86.

[10] A. Yilmaz, O. Javed and M. Shah, Object tracking: *A survey ACM Computing Survey*, **38** (2006), 4–90.

Received October 1, 2010.

Author information

M. T. Gopala Krishna, Department of ISE, DSCE, Bangalore, India.
E-mail: gopalmtm@gmail.com

M. Ravishankar, Department of ISE, DSCE, Bangalore, India.
E-mail: ravishankarmcn@gmail.com

D. R. Ramesh Babu, Department of CSE, DSCE, Bangalore, India.
E-mail: bobrammysore@gmail.com

V. N. Manjunath Aradhya, Department of ISE, DSCE, Bangalore, India.
E-mail: aradhya1980@yahoo.co.in