

# **An Efficient Hierarchical Approach for Background Subtraction and Shadow Removal using Adaptive GMM and Color Discrimination**

**Kshitij Kumar**

Computer Science and Engineering Department  
Motilal Nehru National Institute of Technology  
Allahabad, India

**Suneeta Agarwal**

Computer Science and Engineering Department  
Motilal Nehru National Institute of Technology  
Allahabad, India

## **ABSTRACT**

This paper presents an efficient approach for moving objects detection and shadow removal from color videos obtained using stationary camera. A background subtraction technique based on modified adaptive GMM has been proposed for detecting moving objects. Speed-up techniques have also been applied to enhance the computational efficiency of the algorithm. Then, a robust algorithm for shadow removal is used to remove cast shadows and ghosts. Foreground is reconstructed using graph cut based cleaning and non-recursive blob finding. Comparative experimental results demonstrate that proposed approach performs better in comparison to other state-of-the-art algorithms.

## **General Terms**

Computer Vision, Security, Video Surveillance.

## **Keywords**

Moving Object Detection; Background Subtraction; GMM; Shadow Removal; Color Discrimination; Graph Cut.

## **1. INTRODUCTION**

Moving object detection from video streams is one of the most active areas of research in computer vision. It is the fundamental step for extracting information in various vision based applications such as traffic monitoring (pedestrian detection, vehicle detection), automated video surveillance (human detection, anomaly detection), and control applications (automated robot bodies, human computer interaction).

Computer vision based motion detection aims to detect, segment, and track pixels which correspond to moving objects. The main challenges for motion detection are, presence of dynamic background (such as shaking leaves and falling water), illumination changes, and noise in video. Due to these challenges, few unwanted objects are also classified as part of foreground, namely, shadows and ghost. As mentioned in [1], a ghost is “a set of connected pixels which do not belong to any real moving object but are detected in motion. Shadow is defined as set of connected background pixels which are detected as moving object due to shadow casted over them by a moving object.” Ghosts and shadows may distort the shape of moving object and lead to misclassification. Hence, both ghosts and shadows are unwanted.

An important assumption in many motion tracking approaches is that the camera remains stationary while capturing the video [2]. This assumption leads to a very broad class of moving object detection techniques, commonly known as - *Background*

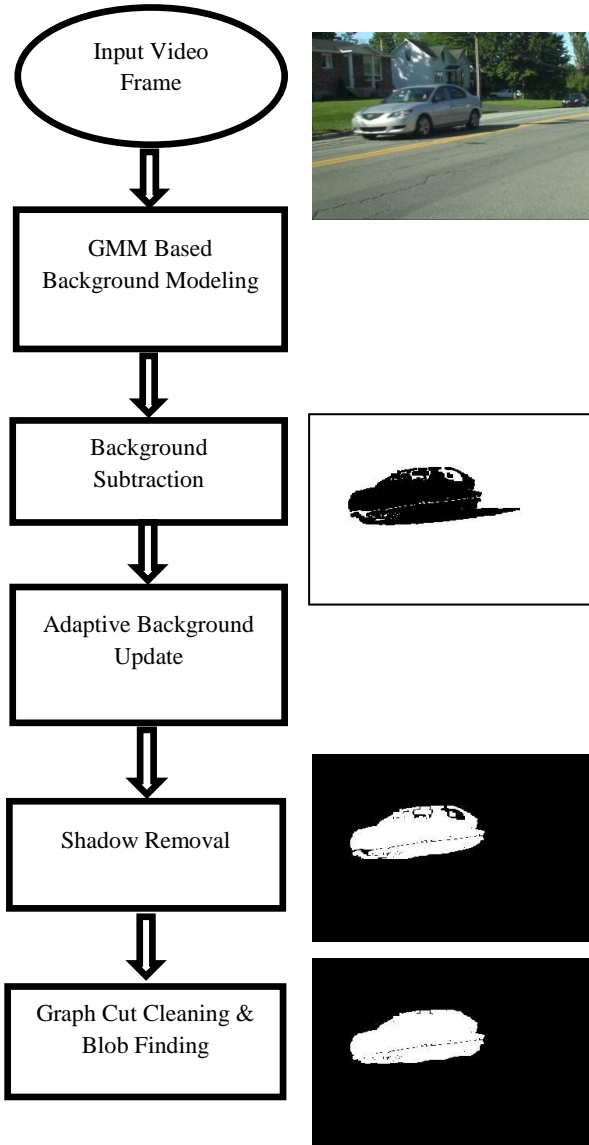
*Subtraction*. The objective of a background subtraction algorithm is to distinguish moving objects (known as *foreground*) from static parts of the scene (called *background*). In most of the cases, background is not already known and needs to be generated automatically by the background subtraction algorithm. When background image is available, moving objects can be obtained by subtracting background image from the current frame. Commonly used background subtraction techniques are - temporal filtering [3], median filtering [4], and single Gaussian model [5].

Gaussian Mixture Model (GMM) was introduced by *Friedman and Russell* [6]. The most widely accepted version of GMM is a refinement over original algorithm for real time tracking proposed by *Stauffer and Grimson* [7]. GMM is based on the assumption that background is predominantly present in the scene, and hence it can be modeled using Gaussian distributions. GMM can deal with illumination changes in real time video sequences. *Zivkovic* [8] proposed an adaptive GMM algorithm which use per pixel adaptation process and is more computationally efficient.

Many methods for removing shadows have been proposed in the recent literature [9][10][11][12]. *Horprasert et. al.* [9] used a statistical color model (in RGB color space) and the ideas of brightness distortion and color distortion to develop a method that is invariant to illumination changes. In [12], *Cucchiara et. al.* proposed a method based on hue, saturation and motion for detecting shadows. The shadow regions are defined by a small variation of the hue component, and a diminution of the luminance and saturation components. *Cucchiara et al.* [13] provide an excellent survey on shadow detection.

Main contributions of this paper are (i) an efficient background model based on *Zivkovic's* recursive Adaptive GMM, (ii) an efficient solution to deal with shadows using brightness and color discrimination, and (iii) a foreground reconstruction process based upon graph cut and blob finding. Comparative experimental results of proposed moving object detection system with shadow removal are also presented using well known change-detection dataset.

The paper is structured as follows: In next section adaptive GMM based approach for background subtraction is described. Section 3 describes shadow removal and foreground reconstruction approach. Speed-up techniques for efficiency enhancement are discussed in Section 4. Section 5 provides a comparative evaluation of proposed approach. Section 6 concludes the paper.



**Figure 1: Block diagram of proposed hierarchical approach**

## 2. ADAPTIVE GMM BASED BACKGROUND SUBTRACTION

Background subtraction can be expressed in terms of three basic operations:

- **Background Model Generation:** In this phase, a statistical model is prepared and a background image is generated using this model.
- **Background Subtraction:** In this phase, a foreground mask is generated for every frame. This step is simply performed by comparing the background image from the current frame. Foreground mask is a binary mask where background = 0 and foreground = 1.
- **Background Model Update:** In this phase, parameters of the background model are updated for the generation of next background image.

Proposed approach for Background Subtraction is based on Zivkovic's Adaptive GMM [8]. Adaptive GMM can efficiently handle the complex background situations like illumination changes and dynamic background.

## 2.1 Background Model Generation

In real world scenarios, background image is not always known and it is to be generated by the system. Objective of this phase is to prepare an appropriate model which could be used for generating background image. The fundamental assumption made is that the camera remains fixed while capturing the motion. In GMM based approach each pixel is modeled as a mixture of K Gaussian distributions. The recent history of a pixel  $X$  at any time instant  $t$  can be written as  $\{X_1, X_2, \dots, X_t\}$ . The probability of observing the current pixel value  $X_t$  in next frame can be written as:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \eta_i(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

where K is the number of Gaussian distributions,  $\omega_{i,t}$  is an estimate of the weight of the  $i^{th}$  Gaussian,  $\mu$  is the mean,  $\Sigma_{i,t}$  is the covariance matrix of the  $i^{th}$  Gaussian, and  $\eta$  is a Gaussian probability density function,

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t - \mu_{i,t})^T \Sigma^{-1} (X_t - \mu_{i,t})} \quad (2)$$

GMM based approaches uses an important assumption that RGB color components are independent and have the same variances. Hence covariance matrix is assumed to be of the form,

$$\Sigma_k = \sigma_k^2 I \quad (3)$$

where  $\sigma_k$  is n-dimensional standard deviation,  $I$  is identity matrix. Parameters of the GMM *i.e.* mean, variance, and weight as follows:

$$\mu_{i,t} = \frac{\sum_{k=1}^t X_k}{t} \quad (4)$$

$$\sigma_i^2 = \frac{1}{t} \sum_{k=1}^t (X_k - \mu_{i,t})^2 \quad (5)$$

$$\omega_{i,t} = \text{Frame Sequence Number} \quad (6)$$

After the initialization, initial moving object detection can be made.

## 2.2 Background Subtraction

For classifying a pixel as foreground or background, Gaussian distributions of every pixel are ordered by  $\omega/\sigma$  ratio in descending order. It's obvious that background pixels occur more frequently than foreground pixels and its intensity value remains constant. So, if a pixel matches any of the first B distributions it is classified as background pixel otherwise foreground pixel,

$$B = \arg \min_b \left( \sum_{i=1}^b \omega_{i,t} > T \right) \quad (7)$$

For each pixel, first step is to identify the  $i^{th}$  Gaussian distribution whose mean is closest to  $X_i$ . The Gaussian distribution is declared matched if Mahalanobi's distance,

$$\sqrt{((X_{t+1} - \mu_{i,t})^T \cdot \Sigma_{i,t}^{-1} \cdot (X_{t+1} - \mu_{i,t}))} < k\sigma_{i,t} \quad (8)$$

where  $k$  defines a small deviation positive threshold. There may be two cases:

**Case 1:** Pixel value is matched with one of  $K$  Gaussian distributions. In this case if pixel value is within threshold, we classify pixel as background, otherwise as part of foreground.

**Case 2:** Pixel value doesn't match with any of the Gaussians. In this case, the pixel is classified as foreground.

If a pixel is classified as background pixel then its color value will be used in next frame. If a pixel is classified as foreground pixel then mean of the Gaussian distribution with lowest variance and largest weight is chosen as background pixel value. Result of background subtraction phase is a binary mask where background = 0 and foreground=1.

### 2.3 Background Model Update

In order to make next detection, parameters of GMM need to be updated. Using the equation (8), two cases can occur:

**Case 1:**  $X_i$  is matched with one of the  $K$  Gaussians.

For matched components, weight is increased, mean is brought closer to current pixel value and variance is decreased to make particular distribution more relevant. These updates are given by following equations respectively:

$$\omega_{i,t+1} = (1 - \alpha)\omega_{i,t} + \alpha \quad (9)$$

where  $\alpha$  is a constant learning rate.

$$\mu_{i,t+1} = (1 - \rho)\mu_{i,t} + \rho X_{t+1} \quad (10)$$

$$\sigma_{i,t+1}^2 = (1 - \rho)\sigma_{i,t}^2 + \rho(X_{t+1} - \mu_{i,t+1}) \cdot (X_{t+1} - \mu_{i,t+1})^T \quad (11)$$

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

Gaussian parameters  $\mu$  and  $\sigma$  remains unchanged for unmatched components and only the weight is decreased as:

$$\omega_{j,t+1} = (1 - \alpha)\omega_{j,t} \quad (12)$$

**Case 2:**  $X_i$  doesn't match with any of the  $K$  Gaussians.

The parameters of least probable distribution  $k$  are updated as follows:

$$\omega_{k,t+1} = \text{Low Prior Weight} \quad (13)$$

$$\mu_{k,t+1} = X_{t+1} \quad (14)$$

$$\sigma_{k,t+1}^2 = \text{Large Initial Variance} \quad (15)$$

After updating the parameters, foreground detection can be made in subsequent frames.

## 2.4 Adaptive Update

In practice, the illumination in the scene could change gradually (daytime or weather conditions in an outdoor scene) or suddenly (switching light in an outdoor scene). A new object could be brought into the scene or a present object removed from it. In order to adapt to these changes, background model is updated by adding new samples and discarding old ones. We choose a reasonable time period  $T$  and at time  $t$ , we have history  $\{X_t, X_{t-1}, \dots, X_{t-T}\}$ . After each period of  $T$ , the Gaussian probability distribution function given by equation (2) is recomputed for every pixel. Thus proposed approach is able to adapt to illumination changes and scene dynamics.

## 3. SHADOW REMOVAL AND FOREGROUND RECONSTRUCTION

Background subtraction using adaptive GMM is susceptible to shadows which are often detected as foreground. Shadows included in foreground distort the shape of moving objects and leads to false positives in next frames. So, a shadow removal method is used to increase the accuracy of our model.

Shadow removal method presented here is based on *Horprasert's* computational color model [9]. The discrimination between expected RGB color of a pixel  $i$ ,  $E_i = [\mu_R(i), \mu_G(i), \mu_B(i)]$  in the background image, and present RGB color value in current image,  $X_i = [X_R(i), X_G(i), X_B(i)]$  is measured. This discrimination is done by decomposing in two parts: brightness distortion and color distortion.

Brightness distortion of a pixel  $i$ , denoted by  $\lambda_i$ , represents the fraction of remaining brightness with respect to expected value. This can be obtained by minimizing equation (16).

$$\phi(\lambda_i) = (X_i - \lambda_i E_i)^2 \quad (16)$$

Color distortion is defined as the orthogonal distance between observed color value and the expected color value. The color distortion of  $i^{th}$  pixel is given by:

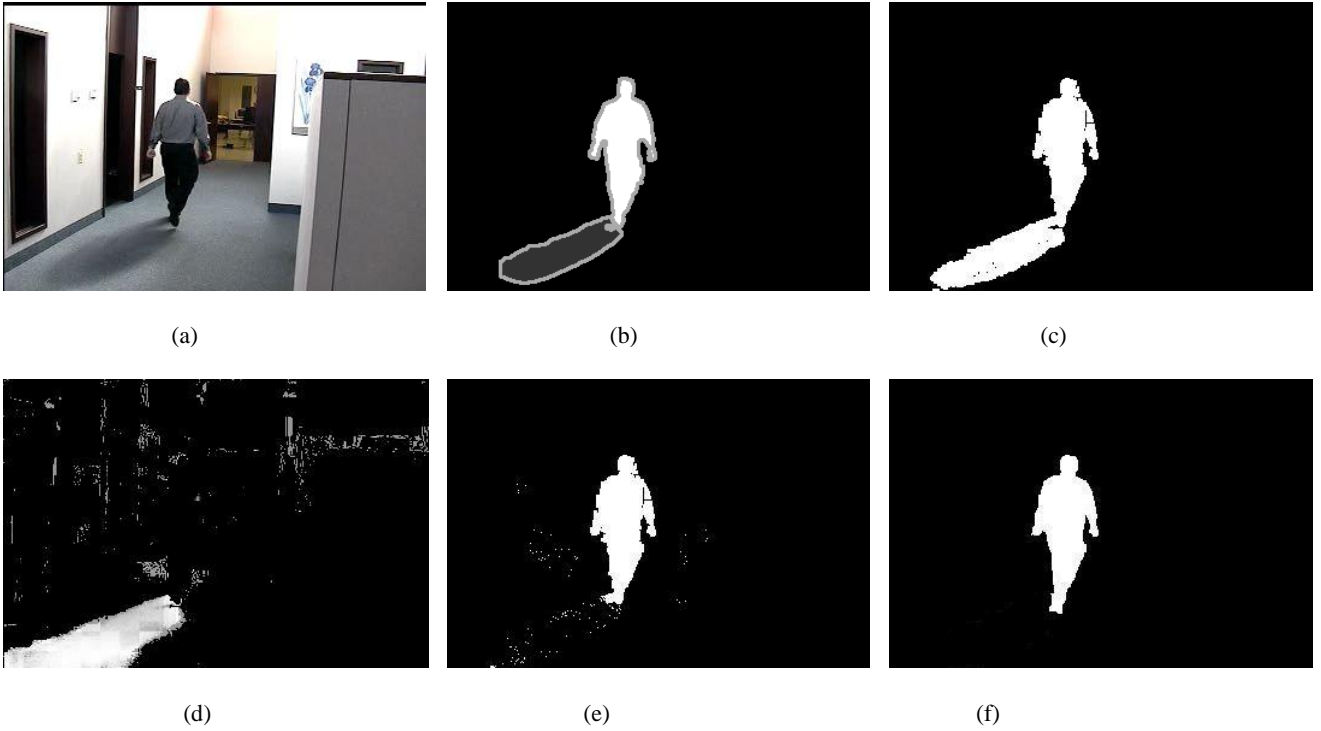
$$CD_i = \|X_i - \lambda_i E_i\| \quad (17)$$

Camera devices have different sensitivities for different color. So, in order to balance weights on the three RGB color channels, the pixel values are scaled and normalized by standard deviation. Normalized color value of a pixel  $i$ , denoted as  $s_i$  is given by:

$$s_i = [\mu_R(i) / \sigma_R(i), \mu_G(i) / \sigma_G(i), \mu_B(i) / \sigma_B(i)] \quad (18)$$

Hence, the brightness distortion and color distortion is given by:

$$\lambda_i = \min \left[ \left( \frac{X_R(i) - \lambda_i \mu_R(i)}{\sigma_R(i)} \right)^2 + \left( \frac{X_G(i) - \lambda_i \mu_G(i)}{\sigma_G(i)} \right)^2 + \left( \frac{X_B(i) - \lambda_i \mu_B(i)}{\sigma_B(i)} \right)^2 \right] \quad (19)$$



**Figure 2. Result of proposed approach (a) Input Frame (b) Ground Truth (c) Background Subtraction (d) Shadow Modeling (e) Shadow Removal (f) Final Foreground after reconstruction**

$$= \frac{\left( \frac{X_R(i)\mu_R(i)}{\sigma_R^2(i)} + \frac{X_G(i)\mu_G(i)}{\sigma_G^2(i)} + \frac{X_B(i)\mu_B(i)}{\sigma_B^2(i)} \right)}{\left( \left[ \frac{\mu_R(i)}{\sigma_R(i)} \right]^2 + \left[ \frac{\mu_G(i)}{\sigma_G(i)} \right]^2 + \left[ \frac{\mu_B(i)}{\sigma_B(i)} \right]^2 \right)} \quad (20)$$

$$CD_i = \sqrt{\left( \frac{X_R(i) - \lambda_i \mu_R(i)}{\sigma_R(i)} \right)^2 + \left( \frac{X_G(i) - \lambda_i \mu_G(i)}{\sigma_G(i)} \right)^2 + \left( \frac{X_B(i) - \lambda_i \mu_B(i)}{\sigma_B(i)} \right)^2} \quad (21)$$

A pixel in the current image may be classified as follows:

$$X_i = \begin{cases} \text{Shadow} & CD_i < \beta_1 \text{ and } \lambda_i < 1 \\ \text{Highlight} & CD_i < \beta_1 \text{ and } \lambda_i > \beta_2 \end{cases} \quad (22)$$

$\beta_1$  is a threshold to determine chromaticity difference between expected and current value. In order to avoid misclassification of a very dark background pixel as shadow, a normalized threshold  $\beta_2 = 1/(1-\epsilon)$  is used.  $\epsilon$  is a lower bound for the normalized brightness distortion.

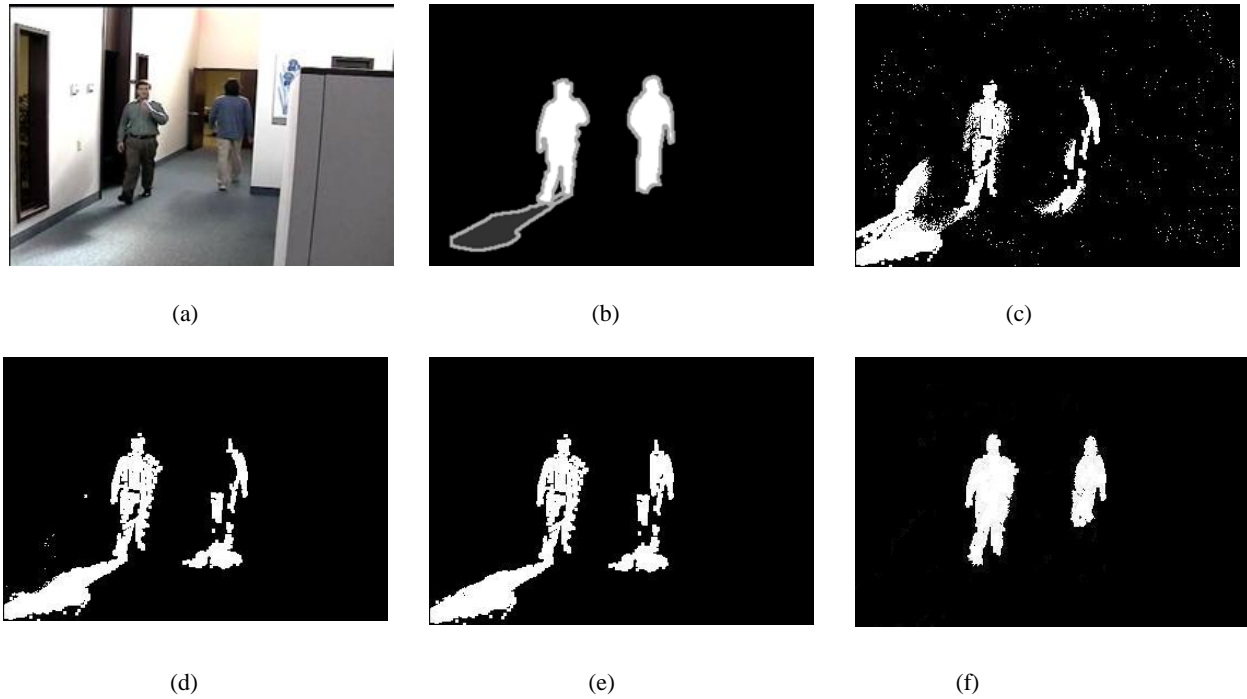
Since shadow removal may deform the shape of an object, a foreground reconstruction process is needed. *Yuri Boykov and Vladimir Kolmogorov's* [14] graph cut method is used for foreground reconstruction. Every frame is formulated as a network graph where each pixel is a node. Edges between the nodes are obtained using d-Connectivity (d=4, 8). Edges are classified in two sets: n-links and t-links. n-links connect neighboring pixels and t-links connect pixel nodes with

terminals. Each edge is assigned a weight typically given by  $w_{i,j} = e^{-\beta(x_i - x_j)^2}$  where  $\beta$  is a parameter and  $x_i, x_j$  are intensity values of nodes. A segmentation problem is formulated as a min-cut max-flow problem. All the nodes corresponding to source are labeled as 1 and to sink as 0. Then, segmentation process groups pixels based on labels assigned to them. Thus reconstruction process attempts to obtain a piecewise constant labeling for given frame. Final foreground result is obtained after the reconstruction process.

#### 4. SPEED-UP TECHNIQUES FOR EFFICIENCY ENHANCEMENT

The algorithms presented in Section 2 and Section 3 are computationally very complex for real time applications. In order to increase the efficiency of algorithm few simplifications to original algorithms are proposed below:

1. Background image is recomputed only after every N frames rather than for each frame. For most of the time, values of every pixel are simply collected and stored for later use. Furthermore, since successive frames tend to be very similar, only every fourth frame is used for statistics collection.
2. In original algorithm, for each pixel of every frame, mean, variance, and weight need to be recalculated. For simplification, it is assumed that the variance is fixed and equal to the variance of image noise. Weight of a Gaussian is taken to be equal to the number of frames. Sum of pixel values are also stored so that mean is simply the sum divided by frame count.
3. The ratios used in (20) are pre-computed for faster calculation. Division operations are replaced by inverse multiplications because multiplication is computationally efficient.



**Figure 3. Comparative Results (a) Input Frame (b) Ground Truth (c) Temporal Difference (d) GMM (e) Adaptive GMM (f) Proposed Approach**

**Table 1. Comparative Evaluation of proposed approach on different statistical parameters**

Technique	Average Re	Average Sp	Average FPR	Average FNR	Average PWC	Average FMeasure	Average Precision
Temporal Difference	0.70	0.969	0.031	0.017	4.35	0.61	0.62
GMM [Stauffer-Grimson]	0.70	0.986	0.014	0.020	3.10	0.66	0.70
Adaptive GMM [Zivkovic]	0.70	0.984	0.016	0.019	3.15	0.66	0.71
Proposed Approach	0.75	0.983	0.009	0.011	2.30	0.67	0.73

## 5. EXPERIMENTAL RESULTS

Publicly available change-detection dataset provided on <http://www.changedetection.net> website [15] is used for experimental purposes and to compare our results with state-of-the-art techniques. The change-detection dataset consists of 31 indoor and outdoor video sequences grouped in following six categories: dynamic background, shadow, camera jitter, baseline, intermittent object motion, and thermal. Resolution of videos varies from 320x240 to 720x576 and length of videos varies from 1000 to 8000 frames. Proposed algorithm is implemented using Intel’s Open Source Computer Vision (*OpenCV*) library. All the experiments were performed in Windows XP environment using CodeBlocks IDE on Intel Core2Duo 1.86GHz processor. Figure 2 shows subjective results obtained using proposed approach. Figure 3 shows comparative results of proposed approach along with other state-of-the-art techniques using change-detection dataset.

Following well known metrics are used to evaluate the performance of proposed algorithm:

$$\text{Recall (Re)} = \frac{TP}{TP + FN} \quad (23)$$

$$\text{Specificity (Sp)} = \frac{TN}{TN + FP} \quad (24)$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN} \quad (25)$$

$$\text{False Negative Rate (FNR)} = \frac{FN}{TN + FP} \quad (26)$$

$$\text{PWC} = 100 \times \frac{FN + FP}{TP + FN + FP + TN} \quad (27)$$

**Table 2. Evaluation of efficiency of proposed approach**

Technique	Running Time
Temporal Difference	20 ms/frame
GMM [Stauffer & Grimson]	38.9 ms/frame
Adaptive GMM [Zivkovic]	47.6 ms/frame
Proposed Approach	27.7 ms/frame

$$\text{Precision (Pr)} = \frac{TP}{TP + FP} \quad (28)$$

$$\text{FMeasure} = 2 \times \frac{\text{Pr.Re}}{\text{Pr} + \text{Re}} \quad (29)$$

Here,

True Positives (TP) = Number of pixels correctly identified as foreground,

False Negatives (FN) = Number of actual foreground pixels incorrectly identified as background

False Positives (FP) = Number of background pixels incorrectly identified as foreground,

True Negatives (TN) = Number of pixels correctly identified as background,

These metrics are evaluated for each video sequence and averaged for each category. Based on above metrics, a comparative evaluation of proposed approach with different state-of-the-art algorithms is shown in Table 1. It can be observed clearly that both GMM and adaptive GMM based approaches suffer with problem of shadows and illumination changes contributing to high false positive rate. Proposed approach removes the problem of shadows and illumination changes. This can be verified by low false positive and false negative rates. Proposed approach improves on percentage of wrong classifications (PWC). A better precision is obtained in comparison to other state-of-the-art techniques.

A simple way to measure complexity of algorithms is processing time. Table 2 shows that proposed algorithm performs faster than GMM and adaptive GMM.

## 6. CONCLUSION

In this paper, an efficient hierarchical system for background subtraction and shadow removal has been proposed. Here, a background model based on modified adaptive GMM is built and parameters of GMM, mean, weight and variance are tuned for better performance. Algorithm further proposes an efficient shadow removal and foreground reconstruction procedure based on color discrimination and graph cut. Then, some speed up techniques for increasing computational efficiency of algorithm has been suggested.

An evaluation framework based on change-detection dataset has been used for measuring the performance of proposed approach. Experimental results are compared with existing state-of-the-art techniques which shows that proposed model can efficiently deal with shadows and illumination changes. Better accuracy and precision has been obtained and used

speed-up techniques improve run time of algorithm. Proposed approach can be explored further to gather the behavioral information of the moving objects.

## 7. REFERENCES

- [1] Bijan Shoushtarian and Helmut E. Bez. 2005. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recogn. Lett.* 26, 1 (January 2005), 5-26.
- [2] Thomas B. Moeslund and Erik Granum. 2001. A survey of computer vision-based human motion capture. *Comput. Vis. Image Underst.* 81, 3 (March 2001), 231-268.
- [3] Polana, R.; Nelson, R., "Low level recognition of human motion (or how to get your man without finding his body parts)," *Motion of Non-Rigid and Articulated Objects, 1994.*, Proceedings of the 1994 IEEE Workshop on , vol., no., pp.77,82, 11-12 Nov 1994.
- [4] Dai Kexue, Li Guohui, Tu Dan, and Yuan Jian, "Prospects and Current Studies on Background Subtraction Techniques for Moving Objects Detection from Surveillance Video". *Journal of Image and Graphics*, July 2006, pp. 919-927.
- [5] Wren, C.; Azarbayejani, A.; Darrell, T.; Pentland, A., "Pfinder: real-time tracking of the human body," *Automatic Face and Gesture Recognition, 1996.*, Proceedings of the Second International Conference on , vol., no., pp.51,56, 14-16 Oct 1996.
- [6] Nir Friedman and Stuart Russell. 1997. Image segmentation in video sequences: a probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence (UAI'97)*, Dan Geiger and Prakash Pundalik Shenoy (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 175-181.
- [7] Stauffer, Chris; Grimson, W. E L, "Adaptive background mixture models for real-time tracking," *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* , vol.2, no., pp.,252 Vol. 2, 1999.
- [8] Zivkovic, Z., "Improved adaptive Gaussian mixture model for background subtraction," *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on* , vol.2, no., pp.28,31 Vol.2, 23-26 Aug. 2004.
- [9] T. Horprasert. D. Harwood and L. S. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," presented at the *IEEE Frame-Rate Applications Workshop*, Kerkyra, Greece, 1999.
- [10] Stephen J. McKenna, Sumer Jabri, Zoran Duric, Azriel Rosenfeld, Harry Wechsler, *Tracking Groups of People*, *Computer Vision and Image Understanding*, Volume 80, Issue 1, October 2000, Pages 42-56.
- [11] J. Stander, R. Mech, and J. Ostermann. 1999. Detection of moving cast shadows for object segmentation. *Trans. Multi.* 1, 1 (March 1999), 65-76.
- [12] Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A., "Detecting moving objects, ghosts, and shadows in video streams," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.25, no.10, pp.1337,1342, Oct. 2003.

- [13] Prati, A.; Mikic, I.; Trivedi, M.M.; Cucchiara, R., "Detecting moving shadows: algorithms and evaluation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.25, no.7, pp.918,923, July 2003.
- [14] Boykov, Y.; Kolmogorov, V., "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.26, no.9, pp.1124,1137, Sept. 2004.
- [15] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, *changedetection.net: A new change detection benchmark dataset*, in *Proc. IEEE Workshop on Change Detection (CDW-12) at CVPR-12, Providence, RI, 16-21 Jun., 2012.*