ISSN:0975-9646

Paresh M. Tank et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3787-3792

# A Fast Moving Object Detection Technique In Video Surveillance System

Paresh M. Tank, Darshak G. Thakore

*, Computer Engineering Department,*

*BVM Engineering College, VV Nagar-388120, India.*

***Abstract*— Nowadays automated surveillance system has become a trend in field of security. Video processing algorithms are utilized to implement these systems. For any video processing system, first task is to detect moving object or subtract a background. In Computer Vision, many techniques are available for detection of the moving object, but Mixture of Gaussian (MoG) models [1] is best suited for system having static and complex background with clutters. MoG technique is more accurate but has a larger time complexity which is unrealistic for real time processing. In this paper, we present a fast technique to extract moving object from background using MoG model and Haar wavelet. In this technique, before applying the MoG we down sample each video frame to acceptable resolution using Haar wavelet decomposition. Selection of wavelet decomposition level depends on original video resolution. The technique has been implemented and tested on videos of PETS [2] and CAVIAR [11] databases. For PETS sample, Original video sample frames having resolution 768 X 576 are down sampled to resolution 192 X 144 using level three Haar wavelet decomposition. Then MoG model is applied to subtract the background. Our result shows this technique is able to detect all moving objects from video in presence of complex background and clutters. We observed that this technique works almost three times faster than using only MoG model without sacrificing the quality of results.**

**In the first part of paper, introduction to visual surveillance system and its architecture have been discussed. In second part, related work in literature is surveyed. Haar wavelet decomposition and MoG models are explained in third part. In forth part, results of presented technique are thoroughly discussed and at last concluding and future work remarks are given.**

***Keywords*— Video surveillance, Background subtraction, Mixture of Gaussian model, Wavelet decomposition**

## I. INTRODUCTION

Video surveillance has been in use to monitor security sensitive areas such as banks, department stores, highways, crowded public places and borders. The advancement in computing power, availability of large-capacity storage devices and high speed network infrastructure has made multi sensor video surveillance systems [12] cheaper and feasible.

Automated surveillance systems are of critical importance for the field of security. The task of reliably detecting and tracking moving objects in surveillance video, which forms basis for higher level intelligence applications, has not fully solved and has many open questions [13].

Segmentation, detection and tracking of multiple objects of a known class are fundamental and difficult problems in computer vision [14]. For this task, we need to detect the objects of interest first and segment them from the background and then track them across different frames while maintaining the correct identities.

The use of smart object detection algorithms is not limited to video surveillance only. Other application domains also benefit from the advances in the research on these algorithms. Some examples are virtual reality, video compression, human machine interface, augmented reality, video editing and multimedia databases.

The principle sources of difficulties in the task of moving object detection are: 1) changes in appearance of the objects with viewpoint, illumination and articulation 2) partial occlusions of the target objects by other objects 3) Complexity of the background that is presence of waving tree leaves, waving of river water etc (4) environment changes.

## II. RELATED WORK

Each video processing application has different requirements but all needs primary and crucial task of moving objects detection from background. Thus, detecting regions that correspond to moving objects such as people, vehicles and animals in video is the first basic step of almost every vision system since it provides a focus of attention and simplifies the processing on subsequent steps of classification and activity identification. Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving due to wind), motion detection is a difficult problem to process.

Most widely used techniques for moving object detection are background modelling, statistical methods, temporal differencing and direct segmentation on each frame, whose descriptions are given below.

### Temporal Differencing

Temporal differencing technique is also known as frame differencing method that detects moving object by doing difference of consecutive frames (two or three) in a video sequence. This method is adaptive and robust to dynamic scene changes; however it generally fails in detecting whole relevant pixels of some types of moving objects. For the homogeneously coloured region of the object the frame differencing algorithm fail in extracting all pixels of the

object's moving region [4]. Also, this method fails to detect non moving objects in the scene. Additional methods need to be used in order to detect stationary objects. Lipton [3] et al. presented a two-frame differencing scheme where the pixels that satisfy the following equation are treated as foreground.

$$|I_t(x,y) - I_{t-1}(x,y)| > \tau$$

To remove problem of two frame differencing in some cases, three frame differencing can be used. Collins et al. developed a mixed method that utilizes three-frame differencing with an adaptive background subtraction model [4]. The hybrid algorithm successfully segments moving objects in video without side effects.

**Background Subtraction**

Background subtraction technique is mainly used when system having static background means system having fixed camera system. This technique detects moving object by subtracting the current image pixel-by-pixel from a reference background image. Reference image is created by averaging images over time using first few frames. The pixels with the difference above some threshold value are declared as foreground pixel. To improve quality of detected foreground regions (to remove noise), some post processing operation such as morphological erosion and dilation can be used. The reference background image is updated over time to adjust with dynamic scene changes. There are many variational approaches exist in literature for background subtraction.

In work of Heikkila and Silven [5], a pixel at location (x, y) in the current image is marked as foreground if

$$|I_t(x,y) - B_t(x,y)| > \tau$$

Condition is satisfied where T is a predefined threshold. The background image is updated by the use of following equation.

$$B_{t+1} = \propto I_t + (1-\propto)B_t$$

After classifying all foreground pixels, morphological closing and opening operation are used to eliminate the small-sized regions. This technique is sensitive to dynamic changes i.e. when stationary objects uncover the background or sudden illumination changes occur.

**Mixture of Gaussian**

Background subtraction was a powerful technique but it had been only successful in indoor environments. In 1998, Stauffer and Grimson [1] developed a technique which represents each pixel by a mixture of Gaussians (MoG) and updates each pixel with new Gaussians during run-time. This background subtraction technique has become successful in indoor as well as outdoor environments.

In this technique, the values of each pixel are calculated as a mixture of Gaussians usually three to five Gaussians are used. Using the variance of each of the Gaussians of the mixture, the background and foreground pixel are classified. Pixel that do not match to the background distributions are classified as foreground pixel. This technique is very robust and has been used in the background subtraction for many computer vision works. In a real time indoor and outdoor tracker, Stauffer and Grimson have used an adaptive Gaussian mixture model in order to evaluate and determine the background of an image [1]. The result has proven that this technique is able to deals with lighting changes, repetitive motions from clutter.

**Eigen Background**

N. Oliver et al. [6] proposed an eigenspace model for moving object detection. In this method, dimensionality of the space constructed from sample images is reduced by using Principal Component Analysis (PCA).

Their principle idea is that after the reduction of space using PCA, the reduced space will represent only the static parts of the scene. So if an image is projected on this space, it gives all moving objects. This model has some limitations. It is not able to model dynamic scenes totally. So, it is not preferred for system having outdoor surveillance requirements.

**Non Parametric Kernel Density Estimation**

Kernel density estimation is another important technique in background subtraction. Kernel density estimation is a nonparametric technique for density estimation in which a known density function (the kernel) is averaged across the observed data points (pixels) to create an approximation to density. In [7], Elgammal et al. has presented kernel density estimation technique. Their result shows it can handle situations where the scene background is not completely static and contains small motions such as moving tree branches and illumination changes. It is prune to false detections that arise due to small camera displacements.

**Direct Segmentation on Frame**

In addition to various techniques explained above, we can directly utilize some versatile image segmentation techniques available in literature like Mean shift and Normalized cut.

Mean Shift is a powerful and versatile non parametric iterative algorithm that can be used for lot of purposes like finding modes, clustering, analysis of multi modal feature space and to delineate arbitrary shaped contour in it etc. Mean Shift was introduced in Fukunaga and Hostetler [9] and has been extended to be applicable in other fields of Computer Vision. If the input is a set of points then Mean shift considers them as sampled from the underlying probability density function. If dense regions (or clusters) are present in the feature space, then they correspond to the mode (or local maxima) of the probability density function [8].

Mean shift can be used to segment images or frames of video. Mean shift segments image in two steps [8]. In first step discontinuities preserving smoothing are done in feature space using mean shift mode seeking procedure. This first step smooths the image without perturbing edges in image. In second step, all the clusters residing close to each another in some known parameters are grouped together and form the segment in image.

Normalized cut is a versatile technique, which is suitable for large class of problem related to image segmentation. In this technique, image segmentation problem treated as graph partitioning problem. To partition a graph, various techniques are available with varying demerits. Shi and Malik [10] proposed a modified cost function and normalized cut to partition a graph into two. They computed the cut cost as a fraction of the total edge connections to all the nodes in the graph and called this disassociation measure the normalized cut (Ncut).

## III. Our Approach

Distinguishing foreground objects from the stationary background is a significant and difficult research problem. Most visual surveillance system's first step is to detect foreground objects. This creates a focus of attention for higher processing levels such as tracking, classification and behaviour understanding. It also reduces computation time considerably since only pixels belonging to foreground objects need to be dealt with. Short and long term dynamic scene changes such as repetitive motions (e. g. waiving tree leaves), light reflectance, shadows, camera noise and sudden illumination variations make reliable and fast object detection difficult. Hence, it is important to pay necessary attention to object detection step to have reliable, robust and fast visual surveillance system

Object detection module has been implemented in three steps as shown in Figure 1. First pre-processing step checks resolution of the input video frame. If frame resolution is too large then algorithm would down sample the size of the frame using Haar wavelet decomposition. Main purpose of this pre-processing is to reduce the frame size for the reduction of time complexity for the detection process.

In second step, moving objects detection is done using adaptive mixture of Gaussian techniques [1]. This step requires tuning of various parameters so that it works according to the needs. This procedure has been implemented for color as well as gray videos.

In last step, Bounding box on each moving object blobs are applied. Which gives size and location of blobs in frame coordinates. Mixture of Gaussian techniques works as explained in remaining section of this part.

Stauffer and Grimson [1] presented a novel adaptive online background mixture model that can robustly deal with lighting changes, repetitive motions, clutter, introducing or removing objects from the scene and slowly moving objects. Their motivation was that a unimodal background model could not handle image acquisition noise, light changes and multiple surfaces for a particular pixel at the same time. Thus, they used a mixture of Gaussian distributions to represent each pixel in the model.

Due to its promising features, we implemented and integrated this method in our visual surveillance system for detecting moving object in video.

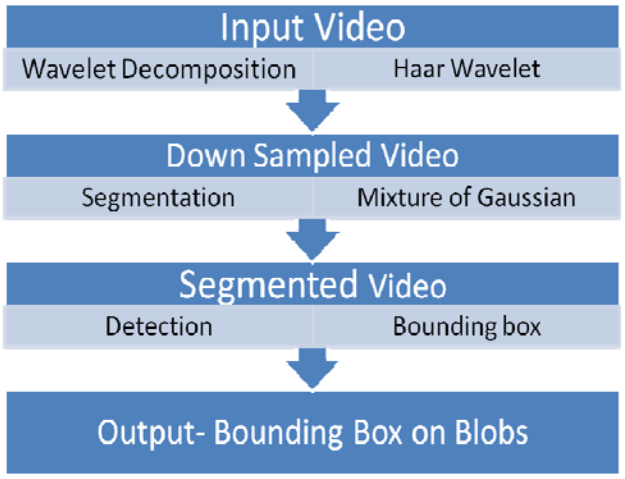


Figure 1 Moving Object Detection Architecture

In this model, the values of an individual pixel (e. g. scalars for gray values or vectors for color images) over time is considered as a "pixel process" and the recent history of each pixel in frame,

$\{X_1 \ldots\ldots X_t\}$, is modelled by a mixture of $K$ Gaussian distributions. The probability of observing current pixel value is given by following equation [1].

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \sum_{i,t})$$

Where $\omega_{i,t}$ is an estimate of the weight (what portion of the data is accounted for this Gaussian) of the $i^{th}$ Gaussian ($G_{i,t}$) in the mixture at time $t$, $\mu_{i,t}$ is the mean value of $G_{i,t}$ and $\sum_{i,t}$ is the covariance matrix of $G_{i,t}$ and $\eta$ is a Gaussian probability density function and it is given by equation [1].

$$\eta(X_t, \mu, \sum) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum|^{\frac{1}{2}}} e^{(-\frac{1}{2}(X_t-\mu_t)^T \sum^{-1} (X_t-\mu_t))}$$

Decision on $K$ depends on the available memory and computational power. Generally value of $K$ is taken between 3 and 5 for complex background [1]. Also, the covariance matrix is assumed to be of the following form for computational efficiency of the algorithm.

$$\sum_{k,t} = \alpha_k^2 I$$

This assumes that red, green, blue color components are independent and have the same variance.

The procedure for detecting foreground pixels is as follows. At the beginning of the system, the $K$ Gaussian distributions for a pixel are initialized with predefined mean, high variance and low prior weight. When a new pixel is observed in the

image sequence, to determine its type, its RGB vector is checked against the *K* Gaussians, until a match is found. A match is defined as a pixel value within (=2.5) standard deviation of a distribution. Next, the prior weights of the *K* distributions at time *t*, $W_{k,t}$ are updated by following equations [1].

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

Where α is the learning rate and $M_{k,t}$ is 1 for the matching Gaussian distribution and 0 for the remaining distributions. After this step the prior weights of the distributions are normalized and the parameters of the matching Gaussian are updated with the new observation using following equation [1].

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho(X_t)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t)$$

Where

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k)$$

If no match is found for the new observed pixel, the Gaussian distribution with the least probability is replace with a new distribution with the current pixel value as its mean value, an initially high variance and low prior weight. In order to detect the type (foreground or background) of the new pixel, the *K* Gaussian distributions are sorted by the value of $w/\sigma$. This ordered list of distributions reflect the most probable backgrounds from top to bottom since background pixel processes make the corresponding Gaussian distribution have larger prior weight and less variance. Then the first B distributions are chosen as the background model, where

$$B = argmin_b\left(\sum_{k=1}^{b}\omega_k > T\right)$$

And *T* is the minimum portion of the pixel data that should be accounted for by the background. If a small value is chosen for *T*, the background is generally unimodal. The accumulated pixels define the background Gaussian distribution whereas scattered pixels are classified as foreground.

## IV. RESULTS

We have successfully applied algorithm to standard surveillance video of CAVIAR [11] and PETS [2] database. Our result shows that the proposed algorithm is able to detect all moving object in presence of clutter in background and illumination change in outdoor as well as indoor environments.

Original video sample frames having resolution 768 X 576 are down sampled to 192 X 144 resolution using level three Haar wavelet decomposition. Then MoG model is applied to subtract the background. Our result shows that this technique is able to detect all moving object from the video even in presence of complex background and clutters. We have observed that this technique works almost three times faster than using only MoG model.

The applied moving object detection algorithm makes use of the mixture of the Gaussian techniques. In this module before applying the MOG, algorithm checks resolution of the input video frame. If the frame resolution is too large then the algorithm will down sample the size of the frame using Haar wavelet decomposition. Main purpose of this pre-processing is to reduce the frame size so that the detection process time decreases.

Bounding boxes are placed around the detected moving object blobs. Some false detection is eliminated by considering size of the blobs. Other false detection further can be eliminated during track initialization in tracking module.

Figure 2 in next page shows results for PETS [2] database sample of camera3 for three frames. This video samples are of outdoor environment having complex background. On the top row in the original frame we can visualize that there is presence of waving tree leaves, moving vehicle, stationary vehicle in parking and multiple humans. Despite of all these complexities, our implemented module has detected all moving objects successfully. On the middle frame, a human coming out of car and he is not in motion is also detected.

In this, there is some false noise detection. Elimination of this noisy false detection can be done in tracking module. Because all this false detection are present only for two to three consecutive frames and all actual objects persist in each frame until it goes out of the camera view. So in tracking module we can initialize the track of objects which has continuous presence in the detection.

Figure 3 shows results for CAVIAR [11] database sample for indoor video. This video sample has large corridor area with multiple humans leaving and entering in the view simultaneously. In the original frame (top row) we can visualize that there is varying illumination in the corridor and also some people are present at the end of the corridor and between the pillars of the corridor. Despite of all these complexities, our module has successfully detected all moving objects present in the sample.

In this figure, we can observe that the shadows of objects are also detected as a part of moving objects. This is because of mixture of Gaussian technique utilizing the motion clue so shadow will be detected as a foreground.

Shadow removal can be taken care of by the standard color based techniques available in the literature.

## V. CONCLUSION

In this work, a smart visual surveillance system based on MoG [1] and Haar wavelet with real-time fast moving object detection capabilities is presented. The state of the art in contemporary work has been thoroughly discussed and the architecture for our system has been designed and implemented. The system operates on both color and gray scale video imagery from a stationary camera. It can handle object detection in indoor and outdoor environments, under changing illumination conditions and complex background with clutters.

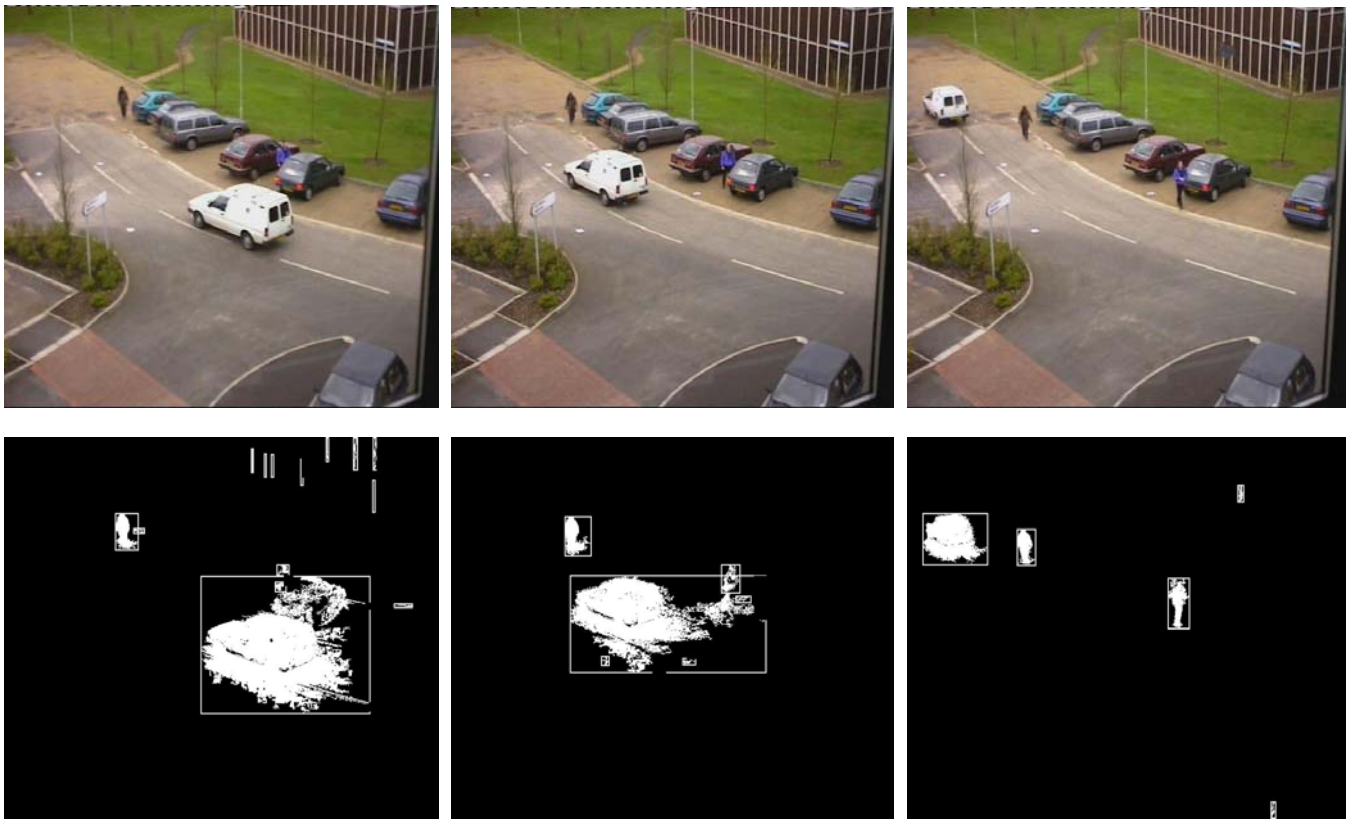

**Figure 2 Top row shows original frames (frame no. 820,840,900) PETS [2] database sample (outdoor). And bottom row shows detected objects.**

**Figure 3 Top row shows original frames (frame no. 150,200,388) CAVIAR [11] database sample (indoor). And bottom row shows detected objects**

Our approach is suitable for any real time video surveillance system especially for fast detection of moving objects in the video frame sequence.

We have successfully applied algorithm to standard surveillance videos of CAVIAR [11] and PETS [2] database. Our result shows that the algorithm is able to detect all moving objects in the presence of clutters in background and illumination change in the outdoor as well as indoor environments.

This implemented module can be applied to any computer vision application for foreground extraction or moving object detection. This module's output can be utilized for the purpose of object detection and tracking.

## VI. FUTURE WORK

This work can be kept continuous for classifying detected objects and tracking trajectories of objects of the interest.

Classification from all detected moving object in video can be done for human, animals, vehicles or any class for which system is aimed. In addition to classification, tracking of detected class can also be implemented. Some post processing task such as shadow removal can also be considered.

After completion of detection, classification and tracking, this work could be utilized in various video surveillance applications like Visual security and surveillance, Driving assistance system, Human computer interaction, Scene analysis and activity recognition, Event detection, Interpretation of video and logical inference, Video annotation.

## REFERENCES


[1] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In Proc. *Computer Vision and Pattern Recognition*, **2**: 246–252, 1999.

[2] http://www.hitechprojects.com/euprojects/cantata/datasets_cantata/dataset.html

[3] A. J. Lipton, H. Fujiyoshi, and R.S. Patil. Moving target classification and tracking from real-time video. In Proc. *of Workshop Applications of Computer Vision*, pages 129–136, 1998.

[4] R. T. Collins et al. A system for video surveillance and monitoring: VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000.

[5] J. Heikkila and O. Silven. A real-time system for monitoring of cyclists and pedestrians. In Proc. of *Second IEEE Workshop on Visual Surveillance*, pages 74–81, Fort Collins, Colorado, June 1999.

[6] N. M. Oliver, Barbara Rosario, and Alex P. Pentland, a Bayesian Computer Vision System for Modelling Human Interactions. IEEE Transaction on *Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, August 2000.

[7] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, in: *European Conference on Computer Vision*, Dublin, Ireland, June 2000.

[8] D. Comaniciu and Peter Meer, "Mean Shift: A Robust approach towards feature space analysis," IEEE Transactions on *Pattern Analysis and Machine Intelligence*, vol 24, No 5, May 2002.

[9] Fukunaga and Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition", IEEE *Transactions on Information Theory,* vol 21, pp 32-40, 1975.

[10] J Shi and J Malik, "Normalized cut and image segmentation", IEEE Transactions on *Pattern Analysis and Machine Intelligence*, vol.22, NO. 8, IEEE, 2000.

[11] http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/

[12] A theses titled "MOVING OBJECT DETECTION, TRACKING AND CLASSIFICATION FOR SMART VIDEO SURVEILLANCE" by Yigithan Dedeoglu.

[13] A theses titled "Automatic Tracking of Moving Objects in Video for Surveillance Applications" by Manjunath Narayana

[14] http://gradworks.umi.com/33/24/3324996.html