IJCSC 0973-7391 Vol 6 • Number 2 April - Sep 2015 pp. 309-315 **Impact Factor: 2.5** **Available at** www.csjournals.com

Comparative Analysis of Traditional Methods for Moving Object Detection in Video Sequence

Chirag I Patel[1], Sanjay Garg[2]
[1,2] Computer Science and Engineering Department, Institute of Technology,
Nirma University, Ahmedabad, India.
[1]chirag453@gmail.com

Abstract: Moving object detection is a crucial and key steps in object tracking, object recognition, human action recognition and visual surveillance systems. It is considered as a big challenge for researchers to design such technique which is computationally efficient and consuming less time. This paper provides a systematic comparative analysis of conventional algorithms of moving object detection and performance measures and assesses their effectiveness via suitable parameters.

Keywords: Moving Object Detection, Surveillance system, Gaussian Mixture Model, Background modeling.

Introduction

The detection of moving objects is a fundamental job for computer vision system. The performances of these systems are not sufficient for many applications due to many difficulties in dealing with various constraints like the appearance of object, illumination changes, dynamic background and variations of the environment.

The advances in information technologies in terms of computational cost and time complexity contributed to the use of computer vision to perform several everyday tasks in various application domains, e.g., Smart video surveillance: (measuring traffic flow, pedestrian congestion and athletic performance, compiling consumer demographics in shopping centers and amusement and park), Military security: (patrolling national borders, measuring flow of refugees, monitoring peace treaties. In these applications, the detection of moving objects is a basic task whose main objective is to decide which image pixels belong to moving objects (foreground) and which image pixels belong to background. However, moving object detection algorithms accuracy and efficiency, both depend on complexity of background as well as moving object.

This leads to several problems such as acquisition noise, dynamic backgrounds, climatic conditions, change in illumination, phantom effects, camouflage, occultation etc. Several methods for moving object detection have been proposed in literature [1-9]; apart from that, we have chosen tradition methods from that.

Previous Methods

A. Simple Background Subtraction

In this simple method [10], moving object is approximated by basic approach i.e. by taking absolute difference of consecutive frames. In this approach, video sequence at t-1 time is considered as background frame for the frame at time t. This method is simplest and fast but, it is sensitive to noise and variation in illumination. Simple background subtraction is very quick to adapt to changes in lighting or camera motion. Objects, that stop moving, are no longer detected. Objects, that startup, do not leave behind ghosts. However, frame differencing only detects the leading and trailing edge of a uniformly colored object. As a result very few pixels on the object are labeled, and it is very hard to detect an object moving towards or away from the camera.

Background image B(x,y,t)=I(x,y,t-1) i.e. previous video sequence.
Moving object or foreground F(x,y,t) is

$$F(\mathrm{x},\mathrm{y},\mathrm{t}) = |I(\mathrm{x},\mathrm{y},\mathrm{t}) - \mathrm{B}(\mathrm{x},\mathrm{y},\mathrm{t})|$$
$$= |I(\mathrm{x},\mathrm{y},\mathrm{t}) - I(\mathrm{x},\mathrm{y},\mathrm{t}-1)| > T$$

B. Reference Frame Difference

In this technique [11], moving object detection is performed by taking difference between current frame and background frame. Background frame is taken in absence of moving object. This method is having speed in locating object as well as it is fast. But, this technique is cannot withstand with multi-modal background. It is also depend on threshold, sensitive to noise and illumination invariant. Background subtraction does a reasonable job of extracting the shape of an object, provided the object intensity/color is sufficiently different from the background. Background subtraction is sensitive to changing illumination and unimportant movement of the background (for example, trees blowing in the wind, reflections of sunlight off of cars or water). Background subtraction cannot handle movement of the camera.

In this technique, background image B(x,y,t) = video sequence $I_B(x,y)$, where, moving objects are absence. Foreground objects can be detected as

$$\mathrm{F(x,y,t)}=\left|I(\mathrm{x,y,t})-I_B(\mathrm{x,y})\right|>T$$

C. Averaging Filter

This method average the image over time and creating a background approximation, which is similar to the current static scene except where motion occurs [12]. It is giving batter result compared to simple background subtraction. This technique gives good result with fixed camera with static noise free background. But, this method is threshold dependent and cannot handle the sudden changes of illumination. Reallocation of object is not possible.

If video sequence have total N images, background image can be formed using following equation:

$$B(\mathrm{x,y})=\frac{\mathrm{I(x,y,1)+I(x,y,2)+\ldots I(x,y,N-1)+I(x,y,N)}}{N}$$

Foreground objects can be detected as

$$\mathrm{F(x,y,t)}=\left|I(\mathrm{x,y,t})-I_B(\mathrm{x,y})\right|>T$$

D. Moving Average Filter

In this method, background is build using averaging of current and past n frames [13]. This method is fast and getting good result with fixed camera with static free background. It can handle changes of lighting condition as well as it adopt changes faster than averaging filter. But, this method cannot withstand with moving background and it is threshold dependent.

Suppose, video sequence have total N frames, and background B(x,y,t) can be approximated by using window, having n (where, n<N) number of video sequences,

$$B(\mathrm{x,y,t})=\frac{I(\mathrm{x,y,t})+I(\mathrm{x,y,t-1})+\ldots+I(\mathrm{x,y,t-n})}{n+1}$$

Foreground objects can be detected as

$$\mathrm{F(x,y,t)}=\left|I(\mathrm{x,y,t})-I_B(\mathrm{x,y})\right|>T$$

E. Temporal Median Filter

Background model is designed by calculating median at each pixel location of all frames in the buffer [14]. Batter results are getting by this method compared to simple background subtraction but not good result compared with averaging filter. Also, this method assume that pixel stays in the background for more than half of the frames in the buffer. This method cannot withstand with moving object.

If video sequence have total N images, background image can be formed using following equation:

$$B(\mathrm{x,y})=\mathrm{median}\ \ \mathrm{I(x,y,1)+I(x,y,2)\ldots I(x,y,N))}$$

F. Minimum, Maximum and Maximum inter-frame difference

This technique designs background model made up of minimum, maximum and a maximum of consecutive frame difference [14]. This method is relatively fast compared to averaging and medianmethod. But, this method cannot handle sudden changes of illumination as well as it is non-reliable for noisy sequence.Each pixel is first classified as either a background or a foreground pixel using the background model. Giving the minimum (M), maximum (N) and the largest inter-frame absolute difference (D) images that represent the background scene model, pixel x from image I is a foreground pixel if:

$$|M(\mathrm{x,y}) - \mathrm{I}(\mathrm{x,y})| > D(\mathrm{x,y})$$

or

$$|\mathrm{N}(\mathrm{x,y}) - \mathrm{I}(\mathrm{x,y})| > D(\mathrm{x,y})$$

G. Linear Predictive Filter (LPF)

This technique estimate current background by applying predictive filter on the pixels in the buffer [15]. In this technique, future value of pixel are estimated as a linear function of previous samples. Pixels, whose prediction error is several times worse than the expected error are classified as foreground pixels. The background model in this method can adopt to both sudden and gradual changes in the background. This method is having batter result in compared to reference frame difference, averaging filter, moving average filter and median filter. But, this method is time consuming and difficult to apply in real time. It cannot withstand with moving average.

For a given pixel, the linear prediction of its next value in time is

$$s_t = -\sum_{k=1}^{p} a_k s_{t-k}$$

Where, $s_t$is the predicted value of the pixel at frame t, the $s_{t-k}$is a past value of the pixel, and the a_k are the predictioncoefficients. The filter uses p past values to make itsprediction. The expected squared prediction error, E[e_t^2], is

$$E\left[e_t^2\right] = E\left[s_t^2\right] + \sum_{k=1}^{p} a_k E\ \ s_t s_{t-k}$$

The $a_k$are computed from the sample covariance values of the s_n.

H. One Gaussian Model

Background modeling by single image need an exhaustively fixed background without any noise and artifacts. In this, background model with probability distribution function learned over a set of training frames [16]. To account for illumination changes mean and covariance of each pixel should be interactively updated. This method is reliable over noisy sequences. It can adopt to slow changes in the scene (for example, gradual changes of illumination). It cannot withstand with multimodal background.

Suppose, mean and variance of collected samples from past video sequences are defined as μ and k. One gaussian model is defined as:

$$\Pr(\mathrm{O}) = \frac{1}{(2\pi)^{\frac{m}{2}} |K|^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(\mathrm{O}-\mu)^T K^{-1}(\mathrm{O}-\mu)\right]}$$

I. Gaussian Mixture Model (GMM)

To account for background containing dynamic textures (waves on water, trees shaken by wind) multimodal background is used [16]. This method is illumination invariant. It is giving good results when background is moving. This type of background model can adopt both sudden and gradual changes in the background. Due to high dynamic background, which cause large changes in the background model, background model fails. But, accuracy depend on how well background model is designed also there isn't neighborhood concept in designing background model. This method is computationally very expensive.

The GMM $G\ t\ =\ C_i\ t\ _{i=1^m}$, is a finite set of clusters of size m, where a cluster at the t[th] instant is given by,

J. $C_i\ t\ =\ \mu_i\ t\ ,\delta_i\ t\ ,\pi_i\ t$ (1)

Where, μ_i(t), δ_i(t) and π_i(t) are the respective mean vector, co-variance matrix and the mixing parameter of C_i(t) at the t^{th} instant.

Initialization:

The GMM is initialized with a single Cluster $C_1(1) = (X_1; \delta_{init}; 1:0)$ where, X_1 is the data vector at t = 1 and δ_{init} is the initial co-variance matrix whose values are assigned from the domain knowledge.
Update:
In this sub-section we deduce the equations for updating the GMM G(t-1) learned till the (t-1)th instant to G(t) with the current data vector X_t. We consider the data vector to be belonging to the cluster C_j(t-1), if $X_t - \mu_j \; t-1^{\,T} \delta_j \; t-1^{\,-1} \; X_t - \mu_j \; t-1 \;\; {}_i \; n\lambda$ Where, λ is a user defined threshold and n is the dimension of the data vector (X € R^n). Now, we consider the following cases. In the first case, we assume that $\exists_j : X_t \in C_j \; t-1$. Let N_i (t) be the number of data vectors that has been assigned to C_i(t) till the tth instant. Thus, we have

$$\pi_i(t) = \frac{N_i \; t}{t} \tag{2}$$

$$\pi_i \; t \; = \frac{t-1 \;\; \pi_i \;\; t-1 \; + \delta \;\; i-j}{t} \tag{3}$$

$$\pi_i \; t \; = \; 1-\alpha_t \;\; \pi_i \;\; t-1 \; + \alpha_t \delta \;\; i-j \tag{4}$$

Where $\alpha_t = \frac{1}{t}$,

$\delta \;\; i-j$ is Kronecker's delta. Now, we update the mean and co-variance in C_j(t-1) only. To update the mean, we proceed as follows.

$$\mu_i \; t \; = \frac{1}{N_i \; t} \sum_{X \in C_j \; t} X \tag{5}$$

$$\mu_i \; t \; = \frac{N_j \;\; t-1 \;\; \mu_j \;\; t-1 \; + X_t}{t \pi_j \;\; t} \tag{6}$$

$$\mu_i \; t \; = (1-\beta_j \;\; t \;)\mu_i \;\; t-1 \; + \beta_j \;\; t \;\; X_t \tag{7}$$

Where, $\beta_j \;\; t \; = \frac{\alpha_t}{\pi_i \;\; t}$

Similarly, we can update the co-variance matrix. From definition, we can compute the co-variance matrix at the tth instant as,

$$\delta_j^2 \;\; t \; = \frac{1}{N_j \;\; t} \sum_{X \in C_j \; t} \; X - \mu_j \;\; t \;\;\;\; X - \mu_j \;\; t \;\;^{T} \tag{8}$$

$$\delta_j^2 \;\; t \; = \frac{1}{N_j \;\; t} \sum_{X \in C_j \; t} \; XX^T - \mu_j \;\; t \;\; \mu_j \;\; t \;\;^{T} \tag{9}$$

Now, further manipulating, by substituting the update rule for μ_j(t), it can be shown that the updated co-variance matrix is given by,

$$\delta_j^2(t) = (1-\beta_j(t))\left(\delta_j^2(t-1) + \beta_j(t)(X_i-\mu_j(t-1))(X_i-\mu_j(t-1))^T\right) \quad (10)$$

In the second case, it may happen that $\exists_j : X_j \in C_j(t-1)$. In such cases, we initialize a new cluster $C_k(t) = (X_t, \delta_{init}^2, \alpha_t)$. If G(t-1) contains less than m clusters, then we add $C_k(t)$ to it. Otherwise, $C_k(t)$ replace the cluster with lowest weight. More so in these particular cases, the mixing parameters of all other clusters are penalized $\pi_i(t) = (1-\alpha_i)\pi_i(t-1), i \neq k$.

Comparison Analysis and Performance Evaluation

A. Database

Database use in this work is Kinect database implemented for foreground segmentation. It contains nine single person sequences, recorded with a kinect device, to show depth and color camouflage situations that are prone to errors in color-depth scenarios.

B. Performance Evaluation

Different algorithms have been proposed for moving object detection in variety of applications, as a primary step towards surveillance system.

True Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) are used to measure the comparison performance in terms of quantitative forms. Precision and Recall are considered in this work and these parameters are calculated as:

$$Recall = \frac{TP}{TP+FP} \text{ and } Precision = \frac{TP}{TP+FN}$$

In this work, value of recall and precision for various moving object detection method is calculated on Kinect dataset and shown in Table 1 and Table 2.

Table 1. Comparison of Recall of various approaches.

Method	Recall
Simple Background Subtraction:	0.3904
Reference frame difference	0.4039
Moving Average Filter	0.5903
Temporal Median Filter	0.6038
Minimum, Maximum and Maximum inter-frame difference:	0.6325
LPF	0.7436
One Gaussian	0.7683
GMM	0.9657

Table 2. Comparison of Precision of various approaches.

Method	Precision
Simple Background Subtraction:	0.3907
Reference frame difference	0.4021
Moving Average Filter	0.5208
Temporal Median Filter	0.5399
Minimum, Maximum and Maximum inter-frame difference:	0.6021
LPF	0.6493
One Gaussian	0.7439
GMM	0.9897

DOI: 10.090592/IJCSC.2015.633

IJCSC 0973-7391 Vol 6 • Number 2 April - Sep 2015 pp. 309-315 **Impact Factor: 2.5** **Available at** www.csjournals.com

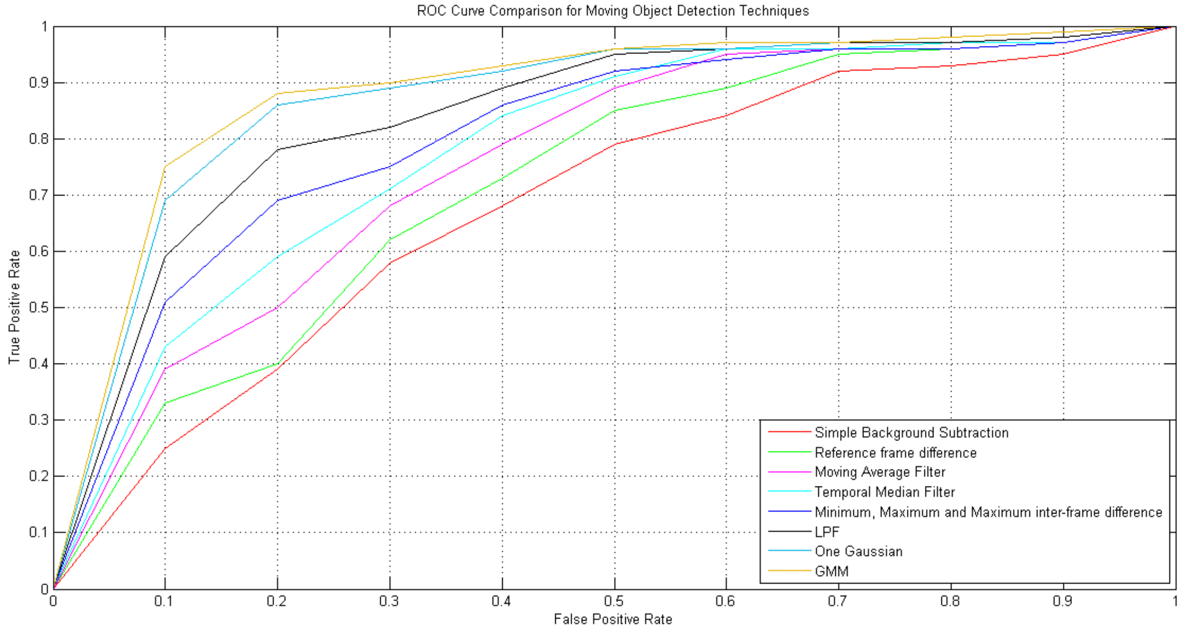

Figure 1 ROC curve comparison for Moving Object Detection

Actually, moving object detection result is binary detection problem hence, the performance measures are used as misdetection rate, false alarm rate and receiver operating characteristic (ROC). ROC defines the performance of moving object detection as binary detection problem. It is graphical representation of true positive rate vs false positive rate. Figure 1 shows the ROC curve for various moving object detection techniques.

Comparison, shown in form of tables and graphical form, shows that GMM is far superior to all convention techniques.

Conclusion

Moving object detection is a preprocessing and the ultimate task in surveillance system. In this proposed comparison study, we have studied various moving object detection techniques. The goal of this work is to provide a better understood of performances of moving object detection technique in video surveillance systems via comparative analysis using tabular and graphical evaluation. Comparison is evaluated using Precision, Recall and in terms of ROC curve. Among all traditional moving object detection techniques Gaussian mixture model gives better performance.

Reference

1. Greiffenhagen M., Ramesh V., Comaniciu D., Niemann H., Statistical modelling and performance characterization of a real-time dual camera surveillance system. *Proc. Int. Conf. Comp Vision PatRecog*, 2000; 2; 335-342.
2. Costantini R., Ramponi G., Bracamonte J., *et al.* Countering illumination variations in a video surveillance environment. *In Proc.of SPIE, Electronic Imaging Conf.*, San Jose, USA, 2001; 4304: 85-97.
3. J. Guo, E.S. Chng, R. Deepu, "Foreground motion detection by difference-based spatial temporal entropy image", In Proceeding of IEEE region 10 Conference, pp.379-382, 2004.
4. Magee DR. "Tracking multiple vehicles using foreground, background, and motion models". In Proceedings of the *Statistical Methodsin Video Processing Workshop*, (Copenhagen, Denmark), June 2002; 7-12.
5. Cucchiara R, Grana C, Piccardi M, Prati A. Detecting moving objects, ghosts and shadows in video streams. IEEE Trans. Patt Anal Mach Intell, PAMI 2003; 25(10): 1337-1342.
6. Mech R, Wollborn M. A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera. Signal Process 1998; 66(2): 203-217.
7. Cavallaro A, Ebrahimi T. Video object extraction based on adaptive background and statistical change detection. In Proc. of SPIE Visual Comm Image Process (VCIP) 2001; 465-475.
8. Rittscher J., Kato J., Joga S., Blake A. A probabilistic background model for tracking. *In Proc. 6th Eur. Conf. CompVision*, 2000; 2: 336-350.
9. Migdal J., Eric W., Grimson L. Background Subtraction Using Markov Thresholds, wacv-motion. *IEEE Workshop on Motion andVideo Computing (WACV/MOTION'05)* 2005; 2: 58-65.
10. Loza, A., Mihaylova, L., Canagarajah, N., and Bull, D., "Structural Similarity-Based Object Tracking in Video Sequences", In: The 9-th International Conference on Information Fusion, 10-13 July 2006, Florence, Italy, 2006.

11. M.A.T. Ho, Y. Yamada, Y. Umetani, "An HMM-based temporal difference learning with model updating capability for visual tracking of human communicational behaviors", In Proceeding of 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp.163-168, 2002.
12. Y. Zhen, C. Yanping, "A real-time motion detection algorithm for traffic monitoring systems based on consecutive temporal difference", In Proceeding of 7th Asian Control Conference, Hong Kong, pp. 1594-1599, 2009.
13. Yilmaz, A., Javed, O., Shah, M., "Object tracking: A survey", ACM Computing Surveys, Vol. 38, No. 4, Article No. 13 (Dec. 2006), 45 pages.
14. Benezeth, Y., Jodoin, P.M., Emile, B., Laurent, H., and Rosenberger, C., "Review and evaluation of commonly-implemented background subtraction algorithms", International Conference on Pattern Recognition (ICPR 2008), 19th Publication Date: 8-11 Dec. 2008, pp. 1-4, 2008.
15. Cheung, S.-C., Kamath, C. "Robust techniques for background subtraction in urban traffic video" Video Communications and Image Processing, SPIE Electronic Imaging, San Jose, January 2004, UCRL-JC-153846-ABS, UCRL-CONF-200706, 2004.
16. Fuentes, L., Velastin, S., "From tracking to advanced surveillance", In Proceedings of IEEE International Conference on Image Processing, Barcelona, Spain, 2003.