

Real time moving object detection for video surveillance based on improved GMM

Shikha Mangal^{1*} and Ashavani Kumar²

Student, Department of Physics, NIT, Kurukshetra, Haryana¹

Professor, Department of Physics, NIT, Kurukshetra, Haryana²

©2017 ACCENTS

Abstract

At present automated surveillance system has become a new trend in the field of security and defence. The moving object detection is one of the fundamental steps for analysis of motion in video surveillance. It provides a way of classification of the pixels into the foreground and background. For a video processing system, the key step is to detect moving object and subtract the background. The mixture of Gaussian (MOG) models is the best suitable for systems having static and complicated background with clutters. It is efficient and robust to illumination changes and camera noise. It reduces the noise from the foreground of the image to a much lower level and causes an efficient detection of objects from a surveillance video that is helpful in many security operations and other applications such as people tracking, traffic monitoring etc. In this paper a new technique is presented to deal with the problem of slow moving objects and to provide fast object detection with robust noise removal and improved background updating. The experimental results show that the proposed method gives better results than the other traditional methods of background subtraction. The dataset that we are using here are CAVIAR an indoor sample video and one other standard outdoor video dataset.

Keywords

Moving object detection, Gaussian mixture model, Video surveillance, Background subtraction, Morphological filtering.

1.Introduction

A static camera observing a scene is a common case in video surveillance applications [1-3]. The video surveillance systems have been using for a long time to monitor security sensitive areas as banks, highways, crowded public places, departmental stores, borders, airports and railway stations. Automated surveillance systems have great importance in the field of security [4-8]. The advancement in the computing power, the high speed network infrastructure and available large capacity storage devices has made multi sensor video surveillance systems [4] much cheaper and feasible. The task of reliably detecting and tracking moving objects in surveillance video, which forms a basis for higher level intelligence applications has not fully solved yet, and has many issues involved requiring solutions [5]. Tracking, segmentation and detection of moving objects is the fundamental steps in computer vision tasks[6]. For this technique, our primary requirement is to segment them from the background and track them within different frames while each frames having its own correct identity.

These advanced objects detection algorithms not only benefits to video surveillance systems but also benefits other application domain such as video compression, human machine interfacing, virtual reality and video editing. A static camera observing a scene in a video surveillance application, but detecting moving objects is an essential step towards high level image processing. The task of moving object detection is associated with some of the problems:1.Complexity of background due to waving tree leaves and river water motion etc. 2.Changes in appearances of the objects in viewpoint and illumination 3.Environmental changes 4.Noise related issues 5.Partial occlusions of the target objects by other objects. The most typical method for moving object detection is Background subtraction, but how to efficiently model and updating the background becomes the challenging aspects of such approaches, improved GMM is used for this purpose here, which is faster than traditional GMM and efficient also. The first part of this paper contains introduction to the video surveillance system and its architecture is also discussed here, then in second part, the related previous work and literature is surveyed. Then the MOG models are explained in third part and in forth part, the results of presented

*Author for correspondence

technique are discussed in detail and then conclusion and future work proposal are also given at last.

2.Related previous work

Every single video processing application has its own set of needs but all of them need primary task of moving objects detection from background, so detecting foreground regions that corresponds to moving objects such as people, vehicles and animals in video is the basic step of every computer vision system because it provides a point of attention on specific regions for detection and simplifies the processing in number of significant steps of identification. Dynamic changes in natural environmental scenes such as weather changes, repetitive motions cause clutters (tree leaves motion) and sudden illumination changes are some of the difficulties that encounter. Most of the techniques used for moving object detection are background modeling or background subtraction, frame differencing, temporal differencing, statistical methods and direct segmentation on each frame. Description of these above defined common techniques is given here:

a) Background subtraction: The background subtraction technique is mainly used with systems having static background or static camera arrangement. The basic idea of this technique is to detect objects in motion by subtracting the current image pixel by pixel from the background image. Reference image is obtained by averaging images over time taking initial few frames. Those pixels having difference above some threshold value are declared as foreground pixel. To improve the quality of foreground region detection or to remove noise, some post processing operations such as dilation or morphological erosion are used. The reference background image is updated with respect to time to adjust with dynamic changes in scenes. Although there are much variations exist in background subtraction technique, but in [7], a pixel at location (x, y) in current frame is considered as foreground if:

$$|A_t(x, y) - B_t(x, y)| > T \quad (1)$$

Here T is a predefined threshold value. The background image is updated with following equation

$$B_{t+1} = \alpha A_t + (1 - \alpha)B_t \quad (2)$$

After classification of all foreground pixels, morphological opening and closing operation is used to eliminate small size regions. This technique is sensitive towards illumination changes and dynamic changes in background.

b) Direct segmentation on frame: Some versatile image segmentation techniques are available in literature such as Mean Shift and Normalized cut for moving object detection. Mean shift is an efficient and multipurpose non parametric iterative algorithm that can be used for various purposes like clustering, analysis of multimodal feature space and finding modes and to locate arbitrary shaped contour in it etc. This mean shift method was introduced in [9] and has been extended to be applicable in other fields of computer vision. The mean shift considers a set of points of input as sampled from the used probability density function, if clusters or dense regions are present in the feature space, then they corresponds 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 the first step, smoothing of discontinuities in feature space is done. These step smoothers the image without disturbing edges of image. In second step, all the clusters close to each other in some known parameters are grouped together and form the segment in image. Normalized cut is yet another versatile method, which is suitable for large problems of image segmentation. In this technique image segmentation problem is treated as graph partitioning problem. Various methods available for graph partitioning, but their limitations [10], proposed a modified cost function and normalized cut method to partition a graph [11]. Cut cost was computed by them as a fraction of the total edge connections to all the nodes in the graph and called this disassociation measure normalized cut (Ncut).

c) Eigen background: An Eigen space model for moving object detection was proposed by [12].In this method, the dimensions of the space created from sample image is reduced using principle component analysis (PCA).After reduced space that significant space will represent only the static parts of the scene. So when an image is projected at this space, it gives all the moving objects. Some limitations are there as it is unable for dynamic scene modelling so it is not preferred for outdoor surveillance requirements.

d) Non parametric kernel density estimation: It is an important technique of statistical approach for background subtraction. It is a non-parametric technique for density estimation where a known density function (kernel) is averaged over the observed data pixels to create an approximate density. A kernel density approximation technique has presented in [13].Their results shows that it can handle small dynamic background changes and

illumination changes. This system is robust to false detections which arise from small camera displacements.

e) Traditional mixture of Gaussian with improvements: Reliable and efficient moving object detection is characterized by some of the important features as high precision, accuracy in shape detection and reactivity to changes in time, flexibility and efficiency in various indoor and outdoor scenarios to provide object detection in real time. Background subtraction is an effective technique for moving object detection but it is limited only to indoor environments. A technique was developed in 1998 which represents each pixel by a mixture of Gaussian (MOG) and updates each pixel with new Gaussian during run-time [1]. This is called mixture of Gaussian or GMM method of background subtraction and was found successful both in indoor and outdoor environments. In this technique, the each pixel value is calculated as a mixture of Gaussians, generally three to five Gaussians are used. On the basis of variance of each of the Gaussians, the foreground and background pixels are classified. This technique is highly robust and used widely in background subtraction in many computer tasks. An adaptive Gaussian Mixture Model has used in real time indoor and outdoor tracking in order to evaluate and determine the background of an image [14].

If pixels are X_1, \dots, X_t and K Gaussian distributions are there with (k is 3 to 5) where X_i ($i=1, 2, \dots, t$) are measurements of (R,G,B) values at each pixel. The probability of observing the current pixel value X_t , where $X_t = (X_t^r, X_t^g, X_t^b)^T$ is

$$P(X_t) = \sum_{k=1}^t w_{k,t} \eta(X_t, \mu_{k,t}, \Sigma_{k,t}) \quad (3)$$

$$\eta(X_t, \mu_{k,t}, \Sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma_{k,t}|^{\frac{1}{2}}} e^{\{-\frac{1}{2}(X_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (X_t - \mu_{k,t})\}} \quad (4)$$

Where $w_{k,t}, \mu_{k,t}, \Sigma_{k,t}$ are weight parameters, mean vector and covariance matrix of the k th Gaussian mode $\sum_{k,t} \sigma_k^2 \cdot I$. This assumes that red green and blue color channels are independent and have the same variance. Different Gaussians are considered to describe different colour clusters, and weight parameter $w_{k,t}$ denotes the probability that this colour cluster occurs at time t . At each pixel, the k Gaussian are always ordered by the value w/σ . By ordering, the most likely background distributions gravitate towards the bottom and finally replaced by new distributions. Then for each pixel the first B_t distributions among k gaussians are chosen as current background model, where B_t is time varying value

determine by the weight parameter $w_{k,t}$ and a user defined threshold $T \in (0,1)$. It is given as :

$$B_t = \text{argmin}(\sum_{k=1}^b w_{k,t} > T) \quad (5)$$

Given a new frame, new measurement X_{t+1} of each pixel is checked against k gaussian until a match is found. where a match is defined as a pixel value within 2.5 standard deviation of the mean value of gaussian. The first matched gaussian updates all of its parameters with the following on line k means approximations while the others only update their weight parameters and keep their mean vectors and variances unchanged.

$$w_{k,t+1} = (1 - \alpha)w_{k,t} + \alpha\mu_{k,t+1} \quad (6)$$

$$\text{(if match, } \mu_{k,t+1} = 1; \text{ otherwise } \mu_{k,t+1} = 0)$$

$$\mu_{k,t+1} = (1 - \rho)\mu_{k,t} + \rho X_t \quad (7)$$

$$\sigma_{k,t+1}^2 = (1 - \rho)\sigma_{k,t}^2 + \rho(X_t - \mu_{k,t})^T (X_t - \mu_{k,t}) \quad (8)$$

$$\rho = \alpha\eta(X_{t+1} | \mu_{k,t}, \sigma_{k,t}) \quad (9)$$

Where α is learning rate and ρ is the learning factor for the current distribution. If no match is found then the last Gaussian is replaced with mean value X_{t+1} , a high variance and low power weight. Finally all pixels in the new frame having no matches to the current background model are flagged as foreground pixels. The results which are obtained using this technique have proven that this technique is able to deal with lighting changes and repetitive motions from clutter. Distinguishing foreground objects from the stationary background is a significant research problem. Detection of foreground is the basic step of most of the visual surveillance systems which allows focus of attention for higher level processing as classification, tracking and behaviour understanding. Computation time also reduced to low level only pixels belonging to foreground are to be processed.

Dynamic scenes such as light reflectance, repetitive motions as waving tree leaves, camera noise, shadows and sudden illumination changes make reliable and fast object detection difficult. So it is necessary to pay attention to object detection to have robust, effective, reliable and fast visual surveillance system. There are following steps which are involved in this technique of improved GMM for moving object detection [15]. First pre-processing step checks input frame resolution if that resolution of frame is large enough than an algorithm will be used to down sample or compress the frame size using haar wavelet decomposition. This pre-processing is done mainly to reduce the frame size for reducing time complexity of the detection process. Second step consists of

detection of moving or foreground objects using adaptive Mixture of Gaussian technique as we explained earlier, this step needs tuning of various parameters together so that it works accordingly. This approach can work for both colour and grey videos. Now in third step, bounding boxes are applied on each moving object blobs, which provides location of blobs and size in frame coordinates.

3. Mean square error and PSNR calculation

Peak signal-to-noise ratio, also known as PSNR, is a term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is mostly used to measure the amount of noise added to the signal during processing. Here moving object is detected and so the resultant foreground detected images will have some kind of noise associated with them, we will calculate PSNR to know the quality of the image so that we can further do processing to remove noise. PSNR is more easily defined by MSE (Mean Squared Error). Given a noise-free $m \times n$ monochrome image I and its noisy approximation K , MSE is defined as:

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (10)$$

The PSNR (in dB) is defined as:

$$PSNR = 10 * \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (11)$$

Here, MAX_I is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255.

4. Results

We have successfully applied our algorithm to one standard surveillance video CAVIAR [11] and other outdoor standard video database. The results which we obtained shows that this proposed algorithm is able to detect all moving objects present in the presence of illumination changes and clutter in background both in indoor and outdoor environments.

Our method allows image down sampling of the original sample video frames if they have high resolution using haar wavelet decomposition and also here morphological filtering is done to remove the noise from the foreground detected and much cleaner foreground will be detected. The MOG method is applied to subtract the background. Our results shows

that the technique is able to detect all moving objects from the video frames even having clutters and complex background problems. This technique is also observed to work 2 times faster than conventional MOG method. The applied technique make use of the mixture of Gaussian technique as well as haar wavelet decomposition. This pre-processing of decomposition is used mainly to reduce the computational time. Bounding boxes are placed around the detected moving object blobs. Elimination of some false detection is done considering size of the blobs and some other false detection can be eliminated during track initialization in tracking modules.

Figure 1 shows results for CAVIAR[11] database sample for indoor video. This video sample consists of large corridor area with some persons leaving and entering the scene simultaneously. We can visualize that there are varying illumination in the original sequence of frames, people are present at different positions as some at the end of the corridor while some other are in between the pillars, but instead of all this complexity our module is detecting all the moving objects present in the view successfully. One more thing to be noticed here is the presence of shadow of objects as a part of moving objects. This is because; the MOG technique utilizing the motion clue that's why shadow is also detected as foreground. By using standard colour based technique this shadow can be removed. *Figure 2* shows moving object detection results for another standard outdoor video dataset[2]. Here real time object detection is done using MOG method and we have some vehicles (cars) are moving in the sample video. This dataset consists of complex outdoor background, where in the top row of frames we can visualize that there are waving tree leaves and moving objects on road. In all these complexities also, our method is working successfully for detecting and removing noise to provide clean foreground detection. Here we see a very little false detection taking place and our method is providing higher efficiency in reduced false detection and noise removal.





Figure 1 Shows (a) original frame (b) Foreground Detected (c) Clean Foreground after noise removal (d) Detected objects (e)Real Time object detection video play CAVIAR [11] database sample (indoor)

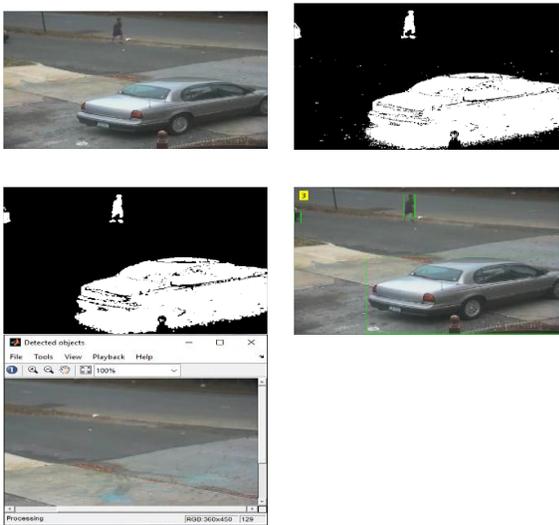


Figure 2 Shows (a)one original frame from video (b)Foreground Detected (c)Clean Foreground (d)Detected objects with number of objects (e) Real Time object detection video play Standard Database sample [2](outdoor)

Table 1 MSE and PSNR for the above shown original frame and foreground detected frame

Dataset	Mse1	Psnr1	Mse2	Psnr2
Caviar	2.1223e+04	4.8967	2.1417e+04	4.8873
Cantata	1.7038e+03	5.8506	1.7081e+03	5.8597

While analysing improved GMM results it is observed that some false object detection is present as the boundary edges in caviar frames also cover the shadow along with object. In order to improve the efficiency of the system, shadow removal is essential to reduce false detection. The MSE and PSNR calculation in *Table 1* shows the reduction in noise

and improvement in PSNR value as a result of morphological operations.

5.Conclusion and future work

In this paper we have presented an efficient visual surveillance system based on MOG [1] and morphological filtering for cleaning the foreground with real time moving object detection. This system operates on both color and grey scale video imagery from a stationary camera. It is possible to detect objects both in indoor and outdoor environments in changing illumination and complex background conditions. In this paper we have successfully applied and implemented our technique on standard database of CAVIAR [11] and one other outdoor video database [2] of road vehicles. In this technique the noise removal is done in order to obtain clean foreground which allows much efficient detection of objects of interest. Our approach is suitable to any of real time video surveillance system for clear and fast detection of moving objects in any video frame sequence. This module can be applied to any of computer vision application for real time moving object detection and foreground extraction. This method has high employability in various object tracking applications in present scenario. This work can be utilized simultaneously for classification and tracking of detected objects trajectories. Also haar wavelet decomposition can be made an essential part of this technique to increase its processing speed. Classification can work for identifying objects of specific classes as human, animals, vehicles and other objects while, on the other hand ,Tracking involves some detected class tracking. In addition to this tracking and classification, some post processing work of shadow removal can also be done here. This work can be introduced to variety of applications such as human computer interaction, scene analysis, driving assistance system, visual security and surveillance, activity recognition, interpretation of video and logical inference.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- [1] Dohare RK, Singh K, Kumar R, Upadhyaya SG, Gupta S. Dynamic model of dividing wall column for separation of ternary system. Malaviya National Institute of Technology Jaipur, India. 2011.
- [2] Naha R, Ray A, Sarkar P, Datta D. Moving Objects Tracking in Real Time Video and Plotting Their Path

- of Movement. International Journal of Advanced Computer Research. 2014; 4(1):242.
- [3] Tank PM, Thakore DG. A fast moving object detection technique in video surveillance system. International Journal of Computer Science and Information Technologies. 2012; 3(2):3787-92.
- [4] Dedeoglu Y. Moving object detection, tracking and classification for smart video surveillance. (Doctoral dissertation, bilkent university). 2004.
- [5] Narayana M. Automatic tracking of moving objects in video for surveillance applications (Doctoral dissertation, University of Kansas). 2007.
- [6] <http://gradworks.umi.com/33/24/3324996.html>. Accessed 16 August 2015.
- [7] Heikkilä J, Silvén O. A real-time system for monitoring of cyclists and pedestrians. Image and Vision Computing. 2004; 22(7):563-70.
- [8] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 24(5):603-19.
- [9] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory. 1975; 21(1):32-40.
- [10] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000; 22(8):888-905.
- [11] <http://groups.inf.ed.ac.uk/vision/CAVIAR>. Accessed 16 August 2015.
- [12] Oliver NM, Rosario B, Pentland AP. A Bayesian computer vision system for modeling human interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000; 22(8):831-43.
- [13] Oliver NM, Rosario B, Pentland AP. A bayesian computer vision system for modeling human interactions. IEEE Transactions on Pattern Analysis And Machine Intelligence. 2000; 22(8):831-43.
- [14] Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction. In European conference on computer vision 2000 (pp. 751-67). Springer Berlin Heidelberg.
- [15] Wang X, Zhao D, Sun G, Liu X, Wu Y. Target detection algorithm based on improved Gaussian mixture model. International Conference on Electrical, Computer Engineering and Electronics 2015 (pp. 846-50). Atlantis Press.



Shikha Mangal was born in rajasthan on 3rd april, 1991. She received her Bachelor degree in Electronic and Communication. Engineering from RTU Kota, in 2012. Presently she is doing her Master's degree from NIT Kurukshetra with specialisation in Instrumentation Engg. and Disseratation work in Image Processing. Her research interests includes image processing, and pattern recognition, intelligent transportation systems and computer vision.
Email:shikhamangal.muskan@gmail.com



Ashavani Kumar received his Ph.D. degree from Aligarh Muslim University Aligarh india in 1994 and presently he is working as HOD of Physics Department NIT Kurukshetra. His area of inetrest includes High energy physics and Nanotechnology.