A Survey on Moving Object Detection using Background Subtraction Methods in Video

Aseema Mohanty
Scholar, DIMAT, CSVTU
Raipur, India

Sanjivani Shantaiya
Assistant Professor, DIMAT, CSVTU
Raipur, India

ABSTRACT
Nowadays moving object detection has become a very prime area for research due to its use in various computer vision applications. Beside from the vital benefit of being able to differentiate video streams into moving and background content, detecting moving objects provides a purpose of attention for recognition, classification and activity scrutiny making these later steps more effective. This research paper presents the thorough survey of background subtraction methods for object detection with a brief information about other methods for object detection. The background subtraction methods discussed here includes Frame Difference, Mixture of Gaussians (MoG), Approximated Median Filter and Eigen Background.

General Terms

Moving object detection techniques, video processing.

Keywords
Moving objects, Object detection, Background subtraction, Frame difference, Mixture of Gaussians, Approximated median filter, Eigen Background.

1. INTRODUCTION
Moving object detection is defined as extracting the motion part from a video stream. It handles separation of moving objects from the immobile background scene. To recognize objects of worth in the video sequence and to group together pixels of these objects is termed as Object Detection. Since moving objects are normally the principal source of information, most methods intend on the detection of such objects. A general approach for object detection is to use information in an individual frame. However, some object detection methods require usage of the temporal information computed from a succession of frames to diminish the number of false detections [1]. The classification of the pixels in the video stream into either foreground or background is said as detection of the moving objects in the video [2].

In the recent years, various approaches are applied to detect the moving objects and still there is hope of improvement in the methods for this area. The optical flow method as described by J.Barron et.al [3] is to compute the image optical flow field, and do bunch processing accordingly to the optical flow distribution characteristics of image or the individual video frame. The fundamental optical flow approach is that characteristic of the frame are tracked over a period to determine the relative swiftness of objects in the scene [4]. The method tries to compute the motion among two image frames which are taken at times t and t + δt at each position. Even though this method is accurate but has high complexity and its practicability is very poor thus is complicated for real-time adaptation [4].

The purpose of next method which is image segmentation algorithm is to separate the image into perceptually the same regions. Every segmentation algorithm addresses two problems, the conditions for a good partition and the way for obtaining effective partitioning [5]. The segmentation method includes mean shift and graph cut techniques. The mean shift clustering as described by Comaniciu and Meer [6] is to find clusters in the joint spatial and color space, [l, u, v, x, y], where [l, u, v] represents the color and [x, y] represents the spatial location. For an image, the method is initialized with a huge number of hypothesized cluster centers randomly selected from the data. Then, each cluster center is moved to the mean of the data lying within the multidimensional ellipsoid centered on the cluster center. The vector defined by the original and the new cluster centers is called the mean-shift vector. The mean-shift vector is calculated iteratively until the cluster centers do not shift their positions. In image segmentation using graph cut method as described by Wu and Leahy [7] the vertices of a graph or image are partitioned into N disjoint subgraphs by cropping the weighted edges of the graph and the total weight of the cropped edges between the subgraph is called a cut and for each new segment the procedure is continuously executed till a threshold is achieved. Although Segmentation method is robust to complex background and sudden illumination changes, it is expensive in terms of memory requirements and processing speed for large image.

In the temporal differencing method, moving regions are detected by making usage of the pixel-by-pixel difference of successive frames (two or three) in a video sequence. Lipton et al. [8] offered a two-frame differencing arrangement where the pixels that fulfill the following equation 1 are taken as foreground where Ʈ is the threshold value-

$$It(a, b) - It{-}1(a, b)| > Ʈ \qquad (1)\ [8]$$

As the two frame differencing raised problems in some cases Collins et al. [9] developed a hybrid technique that associates three-frame differencing with an adaptive background subtraction model for their VSAM project. The hybrid algorithm effectively segments moving regions in video devoid of any previous flaws of temporal differencing. Temporal differencing method is computationally less complex and adaptive to dynamic variations in the video frames but is extra sensitive to the threshold value when determining the changes within difference of video frames and it needs distinct supportive algorithm to detect stationary objects [10].

Lastly, the background subtraction method as described by A.McIvor [11] encompasses comparison of an observed image with an estimate of the image if it included no objects of interest. The areas of the image plane where there is a substantial difference amid the observed and estimated images show the location of the objects of interest. The designation background subtraction derives from the simple procedure of subtracting the observed image from the estimated image and thresholding the obtained to generate or find the objects of interest in the scene. Background subtraction is an extensively used technique for detecting moving objects in videos from static cameras. The reasoning behind the method is that of detecting the moving objects from the variance amid the current frame and a reference frame, frequently called the “background image” or “background model”. As a fundamental, the background image must be a description of the scene without moving objects and must be kept frequently updated so as to become accustomed to the unpredictable luminance circumstances and geometry situations [12]. This can be established either by processing the frame, pixel-by-pixel, or block-by-block. Block centered practice embodies rich contextual information and choices are made for every block of the frame, however in pixel-based approaches, unrestricted conclusion is taken on every individual pixel of the image. Here, the image is eventually split into foreground layer and background layer [13]. Though this method is fast and precise yet it is susceptible to background and illumination variations of the image thus manipulating the intensity values to a large degree, eventually affecting the whole performance [4].

The remaining portion of this paper is systematized in several sections. Section 2 describes about the moving object detection, Section 3 explains the various methodologies that comes under background subtraction techniques, while the Section 4 consists of the conclusion part.

2. MOVING OBJECT DETECTION

The moving object detection is the elementary step for most of the applications based on video processing. Videos are in reality sequel of images, each of which called is a frame, played in fast enough frequency so that human eyes can percept the continuousness of its content. It is evident that all image processing methods can be applied to separate frames. Moreover, the contents of two successive frames are generally closely correlated. An image, generally from a video stream, is distributed into two complementary groups of pixels. The first group include the pixels which correspond to foreground objects while the other and complimentary group include the background pixels. This result which is the detected object is often showed as a binary image or as a mask. It is troublesome to mention an absolute standard regarding what should be recognized as foreground and what should be distinguished as background because this description is rather application specific [9].

Real-time moving object detection is important for a variety of embedded applications like security surveillance, traffic monitoring, robotics, video processing, biomedicine, visual tracking, video compression, human-computer interfaces, medical imaging, content-based indexing, and retrieval. Moving object detection typically acts as an initial step for more processing like classification of the detected moving object, tracking of the detected object etc. So as to perform a lot of subtle operations like classification, tracking lots of methods are developed which are economical and effective for moving object detection. The most common techniques are optical flow method, segmentation method, temporal difference method and background subtraction method.

The complicacy of scenes may be more due to the environmental circumstances, such as the interference consequent of the loathsome weather, abrupt illumination changes, shadow etc. thus object segmentation gets laborious and vital problem which makes object detection a relatively troublesome work. Many a times shadow is categorized as foreground object which gives inaccurate output [7, 9].The difficulty level of this problem greatly depends on the factorhow you describe the object to be detected.

3. BACKGROUND SUBTRACTION TECHNIQUES

Background subtraction is a broadly used methodology for detecting moving objects in videos streams from static cameras. It is the general way of motion detection. It is a process that finds the difference of the current image and the background image to detect the motion region, and it is commonly efficient to deliver data included object information. The keynote of this process lies in the initialization and update of the background image. The efficiency of both will influence the precision of test results [15]. It tries to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image which is composed by averaging images over time in an initialization period. The pixels where the variation is beyond a threshold value are categorized as foreground.

A decent background subtraction algorithm must control the moving objects that first immerse into the background and then become foreground at advanced time. Furthermore, to adapt the real-time requirements of many applications, a background subtraction algorithm must be computationally economical and have little memory requirements, although still being capable to precisely identify moving objects in the video sequence [16]. Also the algorithm should be susceptible to suddenenvironmental changes like illuminance change etc. and should be able to decide that a shadow is not a foreground object.

The four chief phases in a background subtraction algorithm are preprocessing, background modeling, foreground detection and data validation [16] as shown in figure 1.

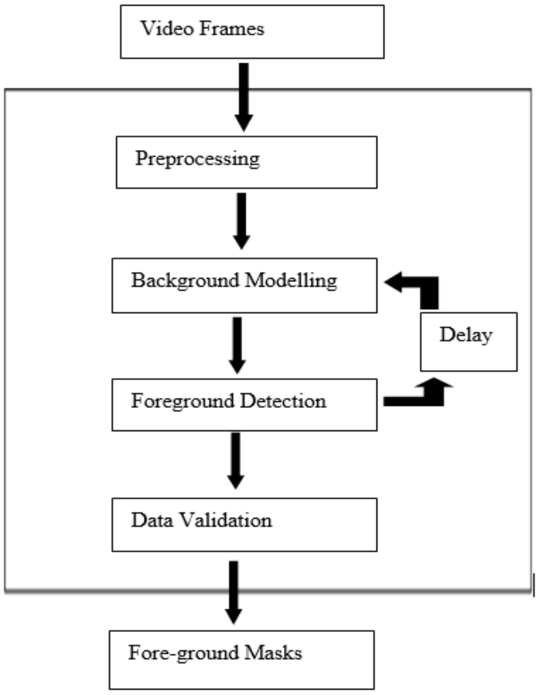


Fig 1: General flow diagram of a background subtraction algorithm [16]

The very first step which is preprocessing comprises of a group of some simple image processing jobs that alter the raw input video into a format that can be processed by succeeding steps [16]. It is done so as to remove noise present in the video so as to smoothen the result which is to be obtained. Background modelingwhich is the backbone of the background subtraction algorithm defines the sort of model chosen to be shown as the background, and the model representation can merely be a frame at time (t-1) formula such as the median model. Model Adaption is the operation employed to set right the background changes that may appear in a scene by continuously updating the current background frame according to the changes in it [17]. This background model offers a statistical portrayal of the whole background scene. The next step foreground detection then recognizes pixels by matching with a certain threshold value for the video frame that cannot be sufficiently described by the background model, and outputs them as a foreground mask.In foreground detection the main job ofthresholding is done. Thresholding is a process that removes an unwanted series of pixels in the scene with respect to certain threshold standards. Lastly, data validation inspects the result and gives the final foreground object by rejecting the pixels that do not correspond to real moving objects.

Now the most popularly used background subtraction methods namely frame differencing method, Gaussian mixture model, approximated median filter method and eigen background method are discussed in the following subsections.

3.1 The Frame Differencing method

In this method the difference is calculated between two frames out of which one is the current frame while the other one is the background frame to detect the presence of any moving object in the video. The foreground object is obtained by the following equation 2:

$$|\text{frame } I_c - \text{frame } I_b| > T \qquad (2)\ [18]$$

Where, frame I_c is the Current frame, frame I_b is the Background frame and T is the Threshold value.

When the pixel difference is more than the threshold value T, it is taken as foreground object.

For this method, the algorithm is as follows [18]:

I. Define the background frame and current frame from video stream.

II. Calculate the gray scale converted image of those frames.

III. Fix the frame dimension for further calculation of pixels.

IV. Calculate the difference amid pixels of the two frames and match with a defined threshold value.

V. If the difference is above threshold value take it as foreground object otherwise as a background.

VI. Update the threshold value according to the changes in successive frames.

The benefit of using this method is that it is fast, easier to apply and performs well for static background but it needs a background not having moving objects [24] otherwise they can be taken as moving object by this method.

(a) Background frame

(b) Current frame

(c) Difference image

Fig 2: The frame difference method (a) the background frame, (b) the current frame and (c) the difference image [18]

3.2 Gaussian Mixture Model method

The Gaussian mixture model (GMM) algorithm is based on a supposition that background is more regularly visible than the foreground, and background variance is little. As single Gaussian is not a decent model for outdoor scenes this method for background subtraction was proposed by Stauffer and Grimson [19] in which every pixel in the background is modelled as a mixture of Gaussian. Each and every pixel value is matched with the current set of models to discover the match. If no match is found, the least model that is acquired is rejected and it is substituted by new Gaussian with initialization by the existing pixel value means the pixel values that don't suit into the background are taken to be background. For a mixture of H Gaussian (H=3) the probability of observing the pixel intensity at time t is expressed using (3) [13],

$$\text{P (z)} = \sum_{j=1}^{H} \omega_{j,z}\, \eta(X_z, \mu_{j,z}, \Sigma_{j,z}) \qquad (3)$$

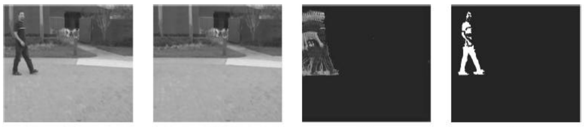
Where,

- X_z represent the pixel intensity at time t of a pixel and the history $\{X_1 \ldots\ldots X_z\}$ of the pixel is modeled using a mixture of K distributions of Gaussian
- H represents the number of Gaussian clusters for modeling pixel history
- $\omega_{j,z}$ is the weight factor associated with cluster i at time t
- η is the Gaussian pdf
- $\mu_{j,z}$ and $\Sigma_{j,z}$ are the mean and covariance matrix of j_{th} Gaussian cluster

The parameter P is assessed dynamically by summation of weights for arranged order of Gaussians till threshold T is achieved (T=0.25).

Supposing the Gaussian mixture model comprises of the combination of Gaussian probability density function, the

Gaussian probability density function of each has its own mean, standard deviation, and weight, the weights can be inferred by the equivalent Gaussian model of the frequency, the more frequently they appear in the Gaussian model the greater the weight. The greater frequency of occurrence, then find the maximum weight on the Gaussian probability density function, lastly, the Gaussian probability density function of the means pixel value is background image [14].



(a) (b) (c) (d)

Fig 3: Mixture of Gaussian modelling for background subtraction. (a) Image from a series. (b) The mean of the highest-weighted Gaussians at each pixels location. (c) The means of the Gaussian with the second-highest weight. (d) Background subtraction result [1]

This method requires less memory to work and gives very accurate result as well as can deal with slow lighting variations although it cannot handle multimodal background and involves rigorous computations [2, 24, 25].

3.3 Approximated Median Filter method

McFarlane and Schofield [20] had proposed a simple recursive filter to evaluate the median of an image pixels in which the running estimate of the median is augmented by one if the input pixel is greater than the estimate and so on decremented by one if the input pixel is lesser than the estimate. This estimate ultimately converges to a value for which half of the input pixels are bigger than and half pixels are lesser than this value that is this value is the median.

In this process, the median filtering buffers the preceding N frames of the video stream. After this the background frame is computed from the median of the buffered frame and the background is subtracted from the current frame to give the foreground pixel.

This is denoted mathematically by equation 4 and 5 [18],

$$\text{Fr} > \text{Bg} \rightarrow \sum_{n-1,1}^{l,m} \text{Bg}(l,m) + 1 \qquad (4)$$

$$\text{Fr} < \text{Bg} \rightarrow \sum_{n-1,1}^{l,m} \text{Bg}(l,m) - 1 \qquad (5)$$

The drawback of this technique is that it does not offer smoother results in all circumstances as it is a recursive technique it does not keeps a buffer for background estimation in its place it regularly updates a single background frame thus any input frame from a very distant past could affect the current background model. Although it means it requires less memory requirements as it doesn't needs to maintain a buffer [24].

(a) (b)

(c)

Fig 4: The Approximated median method. (a) The background frame, (b) The current frame and (c) The detected object [21]

3.4 Eigen Background method

It incorporates region-based (spatial) scene information from the video unlike most of the other methods which are based on color information of the pixels. As proposed by Oliver et al. [22] it uses an Eigen space to model the background for moving object segmentation. This technique's ability is to learn the background model from unconstraint video, even though when they have moving foreground objects segmentation. PCA is used to lessen the dimensionality of the space. After PCA is executed, reduced space should represent only the unmoved portions of the image, even if moving objects is existent in the space. An Eigen vector formed in subspace represents the static unmoved parts of the scene [2].The key steps of this algorithm is described as following [23]:

• A sample consisting of P images from the scene is acquired; mean background image, μb, is calculated and mean normalized images are organized as the columns of a matrix, A.

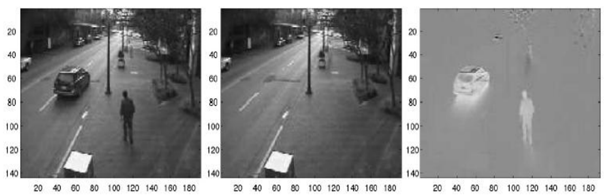
• The covariance matrix, C=AAT, is calculated.

• By means of this covariance matrix C, the diagonal matrix of its Eigen values, L, and the eigenvector matrix, Φ, is calculated.

• The M eigenvectors, having the largest Eigen values (Eigen backgrounds), is kept and these vectors form the background model for the scene.

• If a fresh frame, first projected on the space spanned by M eigenvectors and the recreated frame I' is acquired by using the projection coefficients and the eigenvectors.

• The difference I - I' is calculated. Since the subspace made by the eigenvectors well denotes only the still parts of the scene, consequence of the difference will be the wanted alteration mask comprising of the moving objects.



(a) (b) (c)

Fig 5: Eigenspace based background subtraction method (a) an input frame with objects, (b) reconstructed image after projecting input image onto the eigenspace, (c) difference image.

This model works well for unstable background even though it has some restrictions. It is not able to model dynamic scenescompletely. So, it is not favored for system having outdoor surveillance necessities [2, 12].

4. CONCLUSION

This research paper gives a brief information about various common moving object detection techniques like optical flow, segmentation, temporal differencing and background subtraction method. Further one of the most popular method of object detection which is background subtraction method is described along with its summarized working. Some of the most used background subtraction method are discussed here which involves frame differencing, Gaussian mixture model, approximated median filter and eigen background method. A brief overview of each algorithm along with its mathematical form and pictorial result is provided here. The review shows that the Gaussian mixture model method is best in terms of accuracy for object detection while frame difference method is best to get the result fast. With the recent advances in the field of video processing, moving object detection has become an important area for research and thus this survey can come handy for studying the methods to apply it in many applications.

5. REFERENCES

[1] Alper Yilmaz, Omar Javed and Mubarak Shah, “Object Tracking: A Survey” ACM Computing Surveys, Vol. 38, No. 4, Article 13, December 2006.

[2] A.Ramya and Dr. P.Raviraj, “A Survey and Comparative Analysis of Moving Object Detection and Tracking”, International Journal of Engineering Research & Technology, Vol. 2 Issue 10, October – 2013.

[3] J.Barron, D.Fleet and S.Beauchemin, "Performance of optical flow Techniques", International Journal of Computer Vision, 1994.

[4] Mingyang Yang, “Moving Objects Detection Algorithm in Video Sequence”, 978-1-4799-3903-9/14/$31.00 ©2014 IEEE.

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

[6] Comaniciu, D. and Meer, P., “Mean shift: A robust approach toward feature space analysis”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, NO.5, IEEE, 2002.

[7] Wu, Z. and Leahy, R., “An optimal graph theoretic approach to data clustering: Theory and its applications to image segmentation”, IEEE Transactions on Pattern Analysis and Machine Intelligence. 11, 1993.

[8] A. J. Lipton, H. Fujiyoshi, and R.S. Patil, “Moving target classification and tracking from real-time video”, In Applications of Computer Vision, 1998. WACV'98. Proceedings, Fourth IEEE Workshop on, pages 8–14. IEEE, 1998.

[9] 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.

[10] Kinjal A Joshi and Darshak G. Thakore, “Survey on Moving Object Detection and Tracking in Video Surveillance System”, International Journal of Soft Computing and Engineering, Volume-2, Issue-3, July 2012.

[11] A.McIvor, “Background subtraction techniques”, in Proceedings of Image and Vision Computing , Auckland, New Zealand, 2000.

[12] M. Piccardi, “Background subtraction techniques: a review”, in Proceedings IEEE International Conference Systems, Man, Cybernetics, 2004, pp. 3099–3104.

[13] Soumya Varma and Sreeraj M, “Object Detection and Classification in Surveillance System”, IEEE Recent Advances in Intelligent Computational Systems, 2013.

[14] Abhishek Kumar Chauhan and Prashant Krishan, “Moving Object Tracking using Gaussian Mixture Model and Optical Flow”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

[15] RupaliS.Rakibe and BharatiD.Patil, “Background Subtraction Algorithm Based Human Motion Detection”, International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013.

[16] Deepak Kumar Rout and SharmisthaPuhan, “Video Object Detection using Inter-frame Correlation Based Background Subtraction” IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2013.

[17] R.Manikandan and R.Ramakrishnan, “Human Object Detection and Tracking using Background Subtraction for Sports Applications” International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 10, October 2013.

[18] K.Srinivasan, K.Porkumaran and G.Sainarayanan, “Improved Background Subtraction Techniques for Security in Video Applications”, IEEE 3rd International Conference on Anti-counterfeiting, Security and Identification in Communication, August 2009.

[19] C Stauffer and W Grimson, "Adaptive background mixture models for real - time tracking”, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999.

[20] N. McFarlane and C. Schofield, “Segmentation and tracking of piglets in images", Machine Vision and Applications, 1995.

[21] Asim R. Aldhaheri and Eran A. Edirisinghe, “Detection and Classification of a Moving Object in a Video Stream”, Proc. of the Intl. Conf. on Advances in Computing and Information Technology-- ACIT 2014.

[22] N.M. Oliver, B. Rosario, and A.P. Pentland, “A Bayesian computer vision system for modeling human interactions”, IEEE Trans. on Pattern Anal. and Machine Intell., vol. 22, no. 8, 2000.

[23] S. Y. Elhabian and K. M. El-Sayed, “Moving object detection in spatial domain using background removal techniques- state of the art”, Recent patents on computer science, Vol 1, Apr, 2008.

[24] Himani S. Parekh, Darshak G. Thakore and Udesang K. Jaliya “A Survey on Object Detection and Tracking Methods” International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2014.

[25] Man Zhu and Shuifa Sun, Shuheng Han and Hongying Shen “Comparison of Moving Object Detection Algorithms” Proceedings of 2010 Conference on Dependable Computing (CDC’2010) Yichang, China November 20-22, 2010.