Detecting and Tracking of Moving Objects from Video

ABSTRACT

Background Subtraction methods are widely used for detecting and tracking moving objects in videos. It is useful in many applications such as traffic monitoring, video surveillances. The accurate tracking and detection of moving object is the challenging aspect of such approach. This work proposes a general purpose method which combines the advantage of spatio-temporal differencing with the basic background subtraction method. The results are promising on the proposed method as compared with the basic model of background subtraction and mean shift method.

General Terms

Background subtraction, Spatio-temporal differencing, Mean shift

1. INTRODUCTION

There are many methods or algorithms those are aimed for the simple Object Detection and Tracking. Even though they are proposed for different purposes like to deal with the noise or to be worked on specific images (e.g. Grayscale images) or to reduce complexity, etc; there are still some limitations to the implementation of these various methods or algorithms. Many of the algorithms are concentrated to solve just the object detection problem or just the tracking problem independently. Though they are sufficient to solve their intended problems, all of them cannot solve all the problems. So it is better to solve maximum number of problems by implementing the combination of algorithms to obtain optimal results.

2. RELATED WORK

2.1 OBJECT DETECTION

Moving object detection provides a classification of the pixels in the video sequence into either foreground (moving objects) or background. A common approach used to achieve such classification is background removal, sometimes referred to as background subtraction, where each video frame is compared against a reference or background model, pixels that deviate significantly from the background are considered to be moving objects [1].

Basic Background Subtraction (BBS) algorithm computes the absolute difference between the current frame and a static background frame and compares each pixel to a threshold [2]. Pixels associated with the same object should have the same label; one can accomplish this by performing a connected component analysis. All the connected components are computed and they are considered as active regions if their area exceeds a given threshold. This step is usually performed after a morphological filtering to eliminate isolated pixels and small regions [3]. This is perhaps the simplest object detection algorithm one can imagine.

The Single Gaussian method assumes that each pixel of the background is a realization of a random variable with Gaussian distribution. The mean and covariance of the Gaussian distribution are independently estimated for each pixel [3]. In this method, the information is collected in a vector [Y, U, V], which defines the intensity and color of each pixel [3]. After updating the background, the SGM tries to perform cluster foreground pixels into blobs by binary classification of the pixels into background or foreground. Pixels in the current frame are compared with the background by measuring the log likelihood in color space. If a small likelihood is computed, the pixel is classified as active. Otherwise, it is classified [4] as background.

Another algorithm is denoted as W4 since it is used in the W4 system to compute moving objects. This algorithm is designed for gray-scale images [5]. The background model is built using a training sequence without persons or vehicles. Three values are estimated for each pixel using the training sequence: minimum intensity (Min), maximum intensity (Max), and the maximum intensity difference between consecutive frames.

Mean-Shift (MS) is a clustering approach for joint spatial-color space. Mean-shift approach is used to analyze complex multi-modal feature space and identification of feature clusters [6] [7]. It is a non-parametric technique. Size and shape parameters of its Region of Interest (ROI) are only free parameters on mean-shift process, i.e., the multivariate density kernel estimator [7]. A 2-step sequence of discontinuity preserving filtering and mean-shift clustering is used for mean-shift image.

Optical flow methods make use of the flow vectors of moving objects over time to detect moving regions in an image. They can detect motion in video sequences even from a moving camera, however, most of the optical flow methods are computationally complex and cannot be used real-time without specialized hardware. An OF method can be used to track a region defined by a primitive shape, whose translation is computed by use of the OF [8][7]. OF is based on the idea that the brightness is continuous for most of the points in the image, neighboring points have approximately the same [7] brightness.

An active contour model (ACMs) is another approach, which are in scope of edge-based segmentation and used on object [7] tracking process also. A group of control points connected to each other (or snake) can be considered as snake and can easily be deformed under applied force. This snake is an energy minimizing spine, which is a kind of active contour mapping model. The snake's energy is based on its shape and location within the image [9]. Desired image properties are usually relevant to the local minima of this energy [7].

Next approach, Eigen background Subtraction, presents an Eigen space model for segmentation of moving object. In this method, dimensionality of the space constructed from sample images is reduced by the help of Principal Component Analysis (PCA). It is proposed that only the static parts of the scene are represented by the reduced space after PCA [10].

Moving object detection in a video is the process of identifying different object regions which are moving with respect to the background [11]. After detecting the moving objects from the image frames, it is required to track them. Tracking of a moving object from a video sequence helps in finding the velocity, acceleration, and position of it at different instants of time. In visual surveillance, sometimes it may be required to obtain the speed/velocity of a moving vehicle so as to keep an eye on the movement of a particular vehicle [11]. Other statistical methods are used to extract change regions from background. Background subtraction methods mainly inspire these statistical methods. It uses characteristics of individual pixels of grouped pixels to construct advance background model. During processing the statistics of background frame were updated dynamically. The dynamic statistics of pixels belonging to background image were keeping and updated at each frame by using this method. Background model is used to compare the statistics of foreground frame to its pixels statistics [4]. This approach is reliable in scenes that shadow effects, noise and illumination changes so becoming more popular among the image detection community.

2.2 OBJECT TRACKING

Computer vision researchers interested in tracking because it's a significant and difficult problem to address. The objective of tracking is to establish correspondence of objects and its parts between consecutive frames of video. Surveillance applications uses tracking as important task since it provides to enable higher level data extraction such as activity analysis and behavior recognition and cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation. Inaccurate segmentation of objects makes it difficult to track them. Common problems of erroneous segmentation are partial or full occlusion of objects with each other, and long shadows. Thus, dealing with shadows and coping with occlusions is important for robust tracking.

You can categorize tracking in video according to the methods used for its solution or according to the needs of the applications it is used in. Partial tracking is adequate for indoor surveillance applications and higher level behavior understanding applications and whole body tracking is generally used for outdoor video surveillance whereas objects' part tracking is necessary for some indoor surveillance.

There are two common approaches in tracking objects as a whole [12]: one is based on correspondence matching and other one carries out explicit tracking by making use of position prediction or motion estimation. On the other hand, the methods that track parts of objects (generally humans) employ model-based schemes to locate and track body parts [13].

Various methods for tracking the object in the video are as follows:

In Point tracking, points represent the objects detected in consecutive frames, and the association of such points is based on the previous object state which can be object position and motion. To detect the objects in every frame, this approach requires an external mechanism. Tracking can be formulated as the correspondence of detected objects which are represented by points across frames. Point trackers are suitable for tracking very small objects which follow the single point representation. To track larger objects, multiple points are needed. An important problem in the context of tracking objects using multiple points is the automatic clustering of points that belong to the same [14] object. Point correspondence is a complicated problem-specially in the presence of occlusions.

In Kernel Tracking, Kernel is nothing but the object shape and appearance. For example, the kernel can be a rectangular template or an elliptical shape. By computing the motion of the kernel in consecutive frames, objects are tracked. Kernel tracking is typically performed by computing the motion of the object, which is represented by a primitive object region, from one frame to the next [14].

In Silhouette Tracking, to track the object, the object region in each frame is estimated. Silhouette tracking methods use the information encoded inside the object region [15]. This information can be in the form of appearance density and shape models which are usually in the form of edge maps. Either shape matching or contour evolution is used for Silhouettes tracking. Both of these methods can essentially be a part of object segmentation applied in the temporal domain. Silhouettes can be represented in different ways. The most important advantage of silhouette tracking is their flexibility to handle a large variety of object shapes. Occlusion handling is another important aspect of silhouette tracking methods.

3. BACKGROUND SUBSTRACTION

In video surveillance systems, stationary cameras are typically used to monitor activities at outdoor or indoor sites. Since the cameras are stationary, the detection of moving objects can be achieved by comparing each new frame with a representation of the scene background. This process is called background subtraction and the scene representation is called the background model. Typically, background subtraction forms the first stage in an automated visual surveillance system. Results from background subtraction are used for further processing, such as tracking targets and understanding events [17].

Basic Background Subtraction (BBS) algorithm computes the absolute difference between the current image and a static background image and compares each pixel to a threshold [10]. Pixels associated with the same object should have the same label. This can be accomplished by performing a connected component analysis. All the connected components are computed and they are considered as active regions if their area exceeds a given threshold [3]. This step is usually performed after a morphological filtering to eliminate isolated pixels and small regions. This is perhaps the simplest object detection algorithm one can imagine.

3.1 Algorithm for Basic Background Subtraction and tracking

Basic algorithm steps for detection and tracking of moving objects using simple background subtraction are as given below.

1. Take sample video as input -Read the sample video file.
2. Divide the video into images.
3. Convert RGB TO Gray scale image.
4. Read the input image.
5. For (present position = initial position: final position)
6. Difference between pixel values at present position and the background image (which is the first frame) is calculated.

7. Calculate the absolute value.
8. Store the difference in new image at same pixel position that is at present position.
9. Display the black and white image with the bounding box indicating the moving object.

4. BACKGROUND SUBSTRACTION ADDED WITH SPATIO-TEMPORAL ANALYSIS

This approach for the segmentation of a moving foreground combines temporal image analysis with a background subtraction approach. In short, temporal features are used for the background updating so that the background subtraction for the dynamically changing environment can be achieved. Then afterwards makes use of spatial features to tackle with intensity changes.

4.1 Algorithm for Basic Background Subtraction with Spatio-Temporal Analysis

Basic algorithm steps to detect and track moving objects with the proposed approach are as given below.

1. Take sample video as input -Read the sample video file.
2. Divide the video into images.
3. Convert RGB TO Gray scale image.
4. Read the input image.
5. For (present position = initial position: final position)
6. Difference between pixel values at present position and pixel values at previous (difference between two consecutive frames) position is calculated.
7. Calculate the absolute value.
8. Store the difference in new image at same pixel position that is at present position.
9. Display the black and white image with the bounding box indicating the moving object.

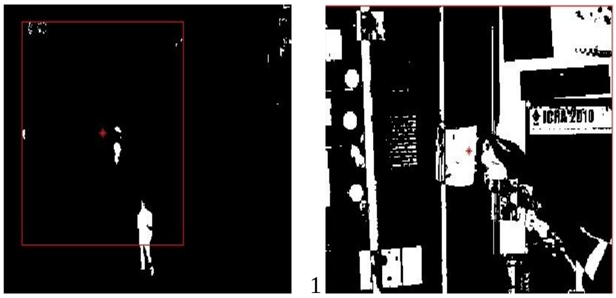
5. RESULTS AND DISCUSSIONS

1.1 Movie 1 **1.2 Movie 2**

1.3 Movie 3 **1.4 Movie 4**

Figure 1: Sample input video frames

Figure 1 shows the sample video frames taken for testing the algorithm with the two of the background subtraction and mean shift algorithms.



2.1 Movie 1 2.2 Movie 2

2.3 Movie 3 2.4 Movie 4

Figure 2: Detecting and tracking of Moving object Using BS method. The method fails to locate the newly arrived moving object in the video.

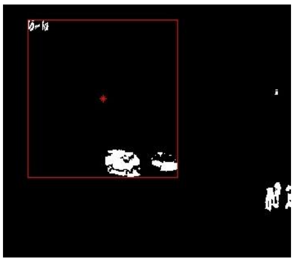
Figure 2 shows the moving object detected and tracked using the simple background Subtraction method. As can be seen from the figure the moving object is detected properly but it considers the position of the moving object in the first frame. The object in first frame seems to be obstructing in the tracking of moving object. The method draws the bounding box by considering both positions of the object. Thus it fails to properly locate the object.

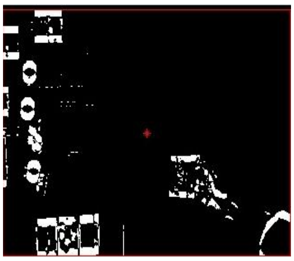
3.1 Movie 1 3.2 Movie 2

3.3 Movie 3 3.4 Movie 4

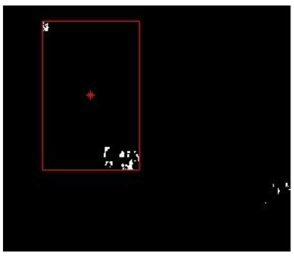
Figure 3: Detecting and tracking of Moving object Using Mean Shift method

Figure 3 shows the detection and tracking of moving objects in video using Mean Shift Method. The main drawback of this method is that you need to select the target first. If the target is not selected properly then target loss occurs and object is not properly detected.





4.1 Movie 1 4.2 Movie 2



4.3 Movie 3 4.4 Movie 4

Figure 4: shows the moving object detected and tracked using the proposed background Subtraction and spatiotemporal method.

As can be seen from the figure the moving object is detected as well as tracked properly as it considers the difference between two consecutive frames. The method draws the bounding box by considering proper position of the object.

The method is also able to detect the newly arrived object in the video during execution. So the proposed algorithm performs well in case of occlusions also.

Table 1 Time required for Object detection and tracking

Video clip	**Time in sec. for background subtraction method**	**Time in sec. for mean shift method**	***Time in sec. for proposed method***
Movie 1	3336	12760	3303
Movie 2	54384	63193	52379
Movie 3	127400	104318	124408
Movie 4	20542	37900	22236

To evaluate the performance of the proposed algorithm we have used some videos. Table 1 shows the time required for the object detection and tracking for three methods.

As can be seen from the table the proposed method is faster as compared with basic background subtraction and mean shift method.

The occlusion detection performance of the algorithms is depicted in table 2.

Table 2 Occlusion detection performance

Video clip	**background subtraction method**	**mean shift method**	***proposed method***
Movie 5	Yes (But it detects both occluded object and moving object)	No (We fail to select object initially)	Yes (It detects moving object after occlusion)
Movie 6	No (It fails to detect moving object after occlusion)	No (We missed the object selection)	Yes (It detects moving object after occlusion)
Movie 7	Yes (It detects moving object after occlusion)	Yes (We selected the object successfully)	Yes (It detects moving object after occlusion)
Movie 8	No (It fails to detect moving object after occlusion)	No (We missed the object selection)	Yes (It detects moving object after occlusion)

We have used some videos to evaluate the performance of the proposed algorithm even in occluded videos. Table 2 shows the success in detection occluded object after occlusion for three methods.

The overall performance of the algorithms is depicted in table 3.

Table 3 Overall Performance of the algorithms

Video clip	**background subtraction method**	**mean shift method**	***proposed method***
Movie 1	No(Considering the whole frame as the moving target)	Yes (Not Considering New arrival)	Yes (Considers the exact bounding box with new arrival)
Movie 2	Yes (But Second car looks like coming through first car)	Yes (Not considered second car arrival)	Yes (Considered both cars very well)
Movie 3	Yes(It detected cars and persons well)	No(No moving object was in sight to be selected)	Yes(Considered movement of object very well)
Movie 4	Yes (Edges were clear)	No(Found and Lost the object in early detection)	Yes (Detected the movement even in background sunlight)

6. CONCLUSION AND FUTURE SCOPE

In this paper we have proposed a new method for detecting and tracking of moving objects in video bu combining the basic background subtraction method is with the spatiotemporal analysis and the results are promising. We have evaluated the performance of the algorithm using sample videos. The performance is compared with the well known mean shift method and the basic background subtraction method. As can be seen, the proposed method is faster and efficient. Future scope will be to resolve the problems of locating multiple moving objects separately in the real time scenarios.

7. REFERENCES

[1] Shireen Y. Elhabian, Khaled M. El-Sayed and Sumaya H. Ahmed, “Moving Object Detection in Spatial Domain using Background Removal Techniques - State-of-Art” , 2008, Recent Patents on Computer Science, Vol. 1, No.

[2] R.C.Gonzalez and R.E.Woods., “Digital Image Processing”, Prentice Hall, 2002.

[3] Jacinto Nascimento, Jorge Marques,” Performance evaluation of object detection algorithms for video surveillance”, Project CAVIAR, 2001

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

[5] I. Haritaoglu, D. Harwood, and L. S. Davis, “W4: real-time surveillance of people and their activities,” *IEEE Trans. Pattern Anal.Machine Intell.*, vol. 22, no. 8, pp. 809–830, August 2000.

[6] Shan, C., Tan, T., Wei, Y., “Real-time hand tracking using a mean shift embedded particle filter”, *Pattern Recognition*, Vol. 40, No. 7, pp. 1958-1970, 2007.

[7] Bahadir Karasulu, “Review And Evaluation Of Well-nown Methods For Moving Object Detection And Tracking In Videos”, July 2010, Journal of Aeronautics and Space Technologies

[8] Jodoin, P.M., Mignotte, M., “Optical-flow based on an edge-avoidance procedure”, *Computer Vision and Image Understanding*, Vol. 113, No. 4, pp. 511-531, 2009.

[9] Dagher, I., Tom, K. E., “WaterBalloons: A hybrid watershed Balloon Snake segmentation”, *Image and Vision Computing*, Vol. 26, pp. 905–912, doi:10.1016/j.imavis.2007.10.010, 2008.

[10] S. Y. Elhabian, 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, pp 32-54, Apr, 2008.

[11] Badri Narayan Subudhi, Pradipta Kumar Nanda, and Ashish Ghosh, “A Change Information Based Fast Algorithm for Video Object Detection and Tracking”, , IEEE Transactions on Circuits and Systems for Video Technology, July 2011.

[12] A. Amer. “Voting-based simultaneous tracking of multiple video objects”. *Proc. SPIE Int. Symposium on Electronic Imaging*, pages 500–511, Santa Clara, USA, January 2003.

[13] Abhishek Kumar Chauhan, Deep Kumar, “Study of Moving Object Detections and Trackings for Video Surveillance”, 2013.

[14] Alper Yilmaz, Omar Javed, Mubarak Shah, “Object tracking: A survey”, *ACM Comput. Surv.* 38, 4, Article 13,December 2006.

[15] In Su Kim, Hong Seok Choi, Kwang Moo Yi, Jin Young Choi, and Seong G. Kong, “Intelligent Visual Surveillance - A Survey” 2010, International Journal of Control, Automation, and Systems, 8(5):926-939. [

[16] Ahmed Elgammal, Ramani Duraiswami, David Darwood, and Larry S. Davis , “Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance”, July 2002,Proceedings of the IEEE, vol.90, no.7

[17] Jong Sun Kim, Dong Hae Yeom, and Young Hoon Joo, “Fast and Robust Algorithm of Tracking Multiple Moving Objects for Intelligent Video Surveillance Systems” August 2011, IEEE Transactions on Consumer Electronics, Vol. 57, No. 3

[18] Vibha L, Chetana Hegde, P Deepa Shenoy, Venugopal K R, L M Patnaik, “Dynamic Object Detection, Tracking and Counting in Video Streams for Multimedia Mining”, July 2008,IAENG International Journal of Computer Science, 35:3, IJCS 35_3_16

[19] Dorin Comaniciu and Peter Meer, "Mean Shift – A Robust Approach Toward Feature Space Analysis", May 2002, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5

[20] Teddy Ko (2011). A Survey on Behavior Analysis in Video Surveillance Applications, Video Surveillance, Prof.Weiyao Lin (Ed.), ISBN: 978-953-307-436-8, InTech.

[21] P. Spagnolo, T.D' Orazio, M. Leo, A. Distante, "Moving object segmentation by background subtraction and temporal analysis", Image and Vision Computing 24 (2006).