International Journal of Innovation in Science and Mathematics
Volume 2, Issue 1, ISSN (Online): 2347–9051

A Survey on Moving Object Detection and Tracking

Ashwani Kumar
Department of ECE
BIT MESRA
ashwinisrivastwa@gmail.com

Sudhanshu Kumar Mishra
Department of ECE
BIT MESRA
sudhanshu.nit@gmail.com

Subhendu Kumar Behera
Department of ECE,
DRIEMS, Cuttack, Odisha
subhendu.nit@gmail.com

Jeevan Mishra
OEC, Bhubaneswar,
Odisha, India
jeevanmishra545@gmail.com

***Abstract* – Object tracking in dynamic scenario is one of the most promising research areas in computer vision. It has a wide range of applications which include surveillance, performance analysis, video indexing, smart interfaces, teleconferencing etc. Robust and real time moving object tracking is a problematic issue in the area of computer vision. In object detection methodology, different methods have already been developed by researchers. But the object detection and tracking problem is still a sensitive issue in computer vision, and provide a lot of research opportunity in this area. This paper presents a survey of various techniques related to object detection and tracking. The main goal of this paper is to review the tracking methods, classify them into different categories, and to identify new trends. This paper is an attempt to accentuate on recent and robust methods for object detection and tracking.**

***Keywords* – Object Detection, Background Subtraction, Temporal Frame Differencing, Object Tracking, Video Surveillance, Statistical Method.**

I. INTRODUCTION

One of the most challenging problems in computer vision over the decades is visual tracking, several applications of visual tracking are far reaching these applications include performance analysis, surveillance, video-indexing, smart interfaces, video compression and much-more. A variety of algorithms have been proposed for tracking of an object. These algorithms can be classified into two categories: one is deterministic method and the other is probabilistic method.

In the deterministic method we perform an iterative search to find the resemblance b/w the model image and the existing image. Some of the algorithm that comes under deterministic method are background subtraction (Heikkila&Silven,1999;Stauffer&Grimson,1999;McIvor,2000;Liuet al.,2001), inter-frame difference (Liptonetal., 1998; Collinsetal.,2000), optical flow (Meyeretal.,1998), skin colour extraction (Choetal.,2001; Phungetal.,2003) etc. While in the stochastic method for object tracking is basically divided into two categories: object tracking using Kalmanfilter (Broida & Chellappa, 1986; Arulampala-metal., 2002) and object tracking using particle filter (Isard & Black 1998; Kitagawa, 1996; Gordonetal., 1993; Risticetal., 2004). In this paper we discuss on the some of the techniques used for object detection and tracking. The outline of the paper is as follows: in section 2 and 3 we describe briefly object detection and recognition method. In section 4, we briefly discussed object tracking method.

II. OBJECT DETECTION

Object detection is an important, yet challenging problem in computer vision. It involves image search, image auto-annotation and scene understanding. Object detection mechanism is required in tracking method. Object detection can be done in two ways: first one is either by detecting objects in every frame or by detecting the object when it appears in the video.

The most common approach in the object detection is to use the information in the single frame. The temporal information computed from a sequence of frames is also a common approach for object detection. This approach is basically used to reduce the number of false detections. The temporal information is usually in the form of frame differential. Once the object is detected, it is then the tracker's task to perform object correspondence from one frame to the next frame to track the object.

The object detection problem can be defined as a labeling problem based on the models of known objects. The object detection and segmentation problem are related to each other.Without at least a partial detection of objects, segmentation cannot be done and without segmentation, object detection is not possible. This section reviews the object detection by adaptive and non-adaptive method.

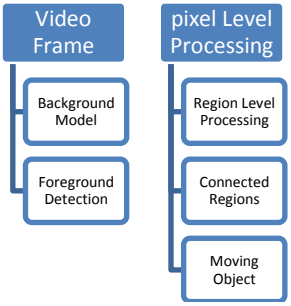

Fig.1. Framework of Moving Object Detection System

For all these detection techniques, the first step is to differentiate the foreground object from stationary background. A commonly used approach to extract foreground object from the image sequence is through background suppression or background subtraction and itsvariant. Many methods have been proposed for real-time applicationfor ground object detection. However, most of them were developed under the assumption that the background consists of stationary object whose color or intensity may change gradually over time. In this paper, for ground pixel mapping at every frame, is done by a combination of various techniques, along with image and post processing have been discussed. Once the foreground pixel is distinguished from the background pixel, the connected regions of the foreground pixels can be grouped to extract the individual object feature such as bounding box, area, perimeter, etc.

A. Background/Foreground detection

Background and foreground detection plays a very important role in the video content analysis system. It is a

foundation for various post-processing modules such as object tracking, recognition, and counting. Many approaches are proposed on the topic based on the background module and procedure used to maintain the model. A real-world background and foreground detection should be robust and adaptive to different light conditions in a noisy environment.

There are two types of background and foreground detection methods: non –adaptive and adaptive. Non-adaptive method depends on a certain number of video frames and do not maintain a background model in the algorithm. Adaptive methods usually maintain a background model and the parameters of background model evolve over time. Some well known background and foreground methods are introduced and discussed in this paper.

- Non-adaptive methods
- Background and Foreground detection based on two frames
- Background and foreground detection based on three frames
- Adaptive methods
- Adaptive background and foreground detection
- Statistical background and foreground detection based on Gaussian model.

B. *Background / Foreground Detection Based on Two Frames*

Two-frame-based background and foreground detection is the simplest non-adaptive method. First, a pixel-wise-absolute difference is calculated between the current frame and the previous frame. Second, the absolute difference is compared with a given threshold value. If the absolute difference is greater than the threshold value, the corresponding pixel belongs to the foreground. Otherwise, it belongs to the background. The threshold value is chosen based on image noise level and complexity of the video sequence. Usually, this basic method is used for simple motion detection, not object tracking/recognition because the foreground cannot be effectively extracted from the video sequence. Sometimes, a pixel may be misclassified as foreground because of the noise contribution.

The algorithm of two-frame-based Background/ Foreground detection is described below

- f_i: A pixel in a current frame, where i is the frame index.
- f_{i-1}: A pixel in a previous frame (f_i *and* f_{i-1} are located at the same location).
- d_i: Absolute difference of f_i and f_{i-1}.
- b_i: Background and Foreground mask.
- T: Threshold value.
- $d_i = |f_i - f_{i-1}|$
- If $d_i > T$, f_i belongs to foreground, otherwise it belongs to background.

C. *Background/Foreground Detection Based on Three Frames*

Three frame-based Background/Foreground detection fixes the issue of pseudo object without increasing too much computational cost. First of all a pixel-wise absolute difference is calculated between the current frame and the previous frame. Secondly pixel-wise absolute difference is calculated between the current frame and the next frame. Thirdly both absolute differences are compared to the given threshold value. If both of them are greater than the threshold value the corresponding pixel belongs to the foreground. Otherwise, it belongs to the background. Three frame-based method also reduces false foreground pixels because of noise contribution. This non-adaptive method can enable a short-term video object tracking/recognition in a controlled environment.

The algorithm of three frame based B/F detection is described below.

- f_{i-1}: A pixel in the previous frame.
- f_i: A pixel in a current frame, where i is the frame index.
- f_{i+1}: A pixel in the next frame (f_i , f_{i-1} *and* f_{i+1}) are located at the same location).
- d_i: Absolute difference of f_i and f_{i-1}.
- d_{i+1}: A pixel wise absolute difference between f_i and f_{i+1}
- b_i: Background and Foreground mask.
- T: Threshold value.
- $d_i = |f_i - f_{i-1}|$ and $d_{i+1} = |f_i - f_{i+1}|$.
- If $d_i > T$ *and* $d_{i+1} > T$, then f_i belongs to the foreground; otherwise it belongs to the background.

D. *Adaptive B/F Detection*

Usually the non-adaptive methods are only useful in the highly supervised, short-term tracking applications without significant changes in the video scene. Adaptive background/foreground detection have many advantages over non-adaptive background/foreground detection so, this method of detection is used in many applications. A standard adaptive B/F detection system maintains a background model within the system. For every pixel, the absolute difference between the current frame and background is calculated. If the result is greater than the given threshold value, the corresponding pixel belongs to the foreground. Otherwise, the corresponding pixel belongs to the background. This method is effective for many video surveillance system objects which move continuously and the background is visible over a significant portion of the time.

The algorithm for Background/Foreground detection is described below.

- f_i: A pixel in a current frame, where i is the frame index.
- μ: A pixel of the background model.
- d_i: Absolute difference of f_i and μ.
- b_i: Background and Foreground mask.
- T: Threshold value.
- α: Learning rate of background.
- d_i: $|f_i - \mu|$

If $d_i > T$, f_i belongs to the foreground ; otherwise it belongs to the background.

E. *Statistical B/F detection based on Gaussian Model*

The most complicated background/foreground detection is based on statistical background model. A background in a given video frame is modelled as a random variable that follows the Gaussian distribution.

$$P_{i,k} \backsim N(\mu_{i,k}, \sigma_{i,k}^2) \quad (1)$$

$P_{i,k}$ is a pixel wise random variable and follows the Gaussian distribution. It is located at the k^{th} position in the i^{th} video frame. $\mu_{i,k}$ and $\sigma_{i,k}$ are corresponding mean and standard deviation parameters of the Gaussian distribution.

Over time, a pixel is modelled as time series called pixel process.

$$\{P_{i,k}: P_{i,k} \backsim N\left(\mu_{i,k}, \sigma_{i,k}^2\right) 0 < i < M, k = 0,1,2 \dots \dots\}$$

Were M is the size of the image.

A pixel-wise Gaussian distribution is totally determined by its mean and standard deviation. The statistical property (mean and standard deviation) of a pixel process evolved over time based on live video data. Basically, a background can be seen as a collection of pixel-wise mean per frame at a given time. Every background pixel has its own threshold value derived from the corresponding standard deviation. This method is very efficient when different regions have different lighting conditions or different noise levels. A uniform threshold may result in object disappearing when they enter a low noise level.

The algorithm details are described below.

- f_i = A pixel in a current frame (the i^{th} video frame)
- μ_i: Mean of a pixel-wise background Gaussian distribution (f_i and μ_i are located at the same location)
- σ_i: Standard deviation of a pixel-wise background Gaussian distribution.
- d_i: Absolute difference between f_i and μ_i.
- T_i: A pixel wise threshold.
- α : Learning rate of background.
- η: Threshold gain.
- $d_i = |f_i - \mu_i|$.
- $T : \eta\sigma_i$
- If $d_i > T$, then f_i belongs to the foreground otherwise, it belongs to the background.

If f_i belongs to the background, update the pixel wise mean and standard deviation of the corresponding pixel distribution at a given learning rate.

III. Pixel Level Post-Processing

The output of the foreground detection contains noise. Generally it is affected by various noise factors. In order to overcome this dilemma of noise requires further pixel level processing. There are various factors that cause the noise in foreground detection such as:

Camera noise: camera noise presents due to camera's image acquisition components. This noise is produced because of the intensity of a pixel that correspond to an edge between two different color objects in the scene may be set to one of the object's color in one frame and to other's color in the next frame.

Background colored object noise: Reflectance noise is caused by the light source. When a light source moves from one position to another some parts in the background scene reflects light.

We can use a low pass filter for morphological operations, erosion and dilation for the foreground pixel map to remove noise that is caused by the items listed above. Our aim of applying these operations is removing noisy foreground pixels that do not correspond to actual foreground regions to remove the noisy background pixels near and inside object regions that are actually foreground pixels. Low pass filters are used for blurring and for noise reduction. Blurring is used in preprocessing tasks such as the removal of small details from an image prior to larger object extraction used for pixel level post processing. A Gaussian filter smoothes an image by calculating theweighted averages in a filter co-efficient. Gaussian filter modifies the input signal by convolution with a Gaussian function.

A. Detecting Connected regions

After detecting foreground regions and applying post-processing operations to remove noisy regions .The filtered foreground pixels are grouped into connected regions that correspond to objects the bounding boxes of these regions are calculated.

B. Region Level Post processing

Even some pixel-level noise gets removed, still some artificial small region remains just because of the bad segmentation. To remove this type of regions, the region that has smaller sizes than a pre-defined threshold, are deleted from the foreground pixel map.

Once segmenting the regions we can extract the features of the corresponding object from the current image. These features withcenter of mass or just the centroid are bounded area of the connected component. These features are used for object tracking and classification for the further processing in event detection.

IV. Object Tracking

Object tracking in video sequences is a challenging task and has various applications. It is an important task within the field of computer vision. Tracking involves matching detected foreground objects between consecutive frames using different feature of object like motion, velocity, color and texture. There are three major steps involved in video tracking, detection of interesting moving object, tracking of such object from frame to frame and analysis of object tracks to recognize their behaviors. In tracking approach, the objects are represented using the shape appearance models. The model selected to represent object shape limits the type of motion. For example, if an object is represented as a point, then only a translational model can be used. In caseof a geometric shape representation like an ellipse is used for the object, parametric motion model like affine or projective transformations are appropriate.

The motion of rigid objects in the scene is approximated by this representation. Silhouette or contour is the most descriptive representation of a non-rigid object. Different object tracking methods are described as follows.

A. Point tracking

Point tracking is robust, reliable and accurate tracking method developed by Veenman [9]. This method generally is used to track the vehicles. This approach requires a good

level of fitness of detecting object. This method is based on deterministic or probabilistic methods of tracking [10].

Object tracking is based on a point which is represented in detectingobjects in consecutive frames and association of the points is based on the previous object state which can include object position and motion. This approach requires an external mechanism to detect the object in every frame.

B. Kernel Tracking

In this approach kernel require shape and appearance of the object [9]. In this approach any features object is used to track objects as kernel like rectangular template or an elliptic shape with an associated histogram. After computing the motion of the kernel between consecutive frame object can be tracked.

In [4], Mean-Shift tracking is based on the kernel tracking method .In this method E-kernel is used. It representsa histogram feature based by spatial masking with an isotropic kernel.

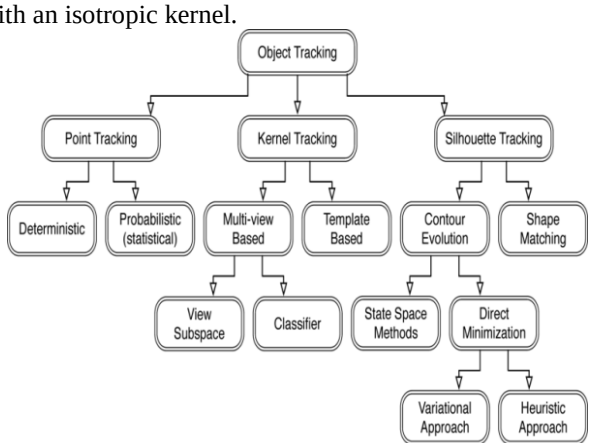

Fig.2. Taxonomy of tracking methods [9]

C. Silhouette Tracking

In this approach silhouette is extracted from detected object. By shape matching or contour evolution silhouettes are tracked either by calculating object region in consecutive frame tracking is done. The silhouette tracking method makes use of the information stored inside the object region [6]. The appearance of the region can have appearance density and shape models which are given in the object models.

Tracking of an object is based on the features, requires selecting the right features, which plays a critical role in tracking. In general the features used for tracking must be unique so that the object can easily be distinguished in the feature space. Following this various features is used for object tracking:

a. Color: The apparent color of an object is influenced primarily by two physical factors .First is the spectral power distribution of the illuminate and the second is the surface reflectance of the object [12]. In image processing the RGB (red, green, blue) color space is usually used to represent color.

b. Edges: Object boundaries usually generate strong chances in image intensities [9]. Edge detection is used to identify these changes. An important property of Edges is that they are less sensitive to illumination changes compared to color features.

c. Centroid: The center of mass (centroid) is a vector of 1-by-n dimensions in length that specifies the center point of a region. For each point it is worth mentioning that the first element of the centroid is the horizontal coordinate(x-coordinate)of the center of mass, and the second element is the vertical coordinate (y-coordinate) [11].

d. Texture: Texture is used for classification as well as tracking purpose. This feature is used to identify a region or object in which we are interested in.It is a measurement of intensity variation of a surface which quantifies properties such as smoothness and regularity [20]. Compared to color, texture requires a processing set up to generate the descriptors.

Among all features,color and texture feature are widely used to track the object. A color bandis sensitive to illumination variation.

D. Correspondence Based Matching Algorithm

In correspondence based matching algorithm, we take objects of the previous frame and match the pairs which are close. In this method we compute the distance between the centroid that is smaller than the pre –defined threshold T [7]. For example suppose two object Oc and Op where the subscripts c stands for the current frame and p for the previous frame with centre of mass (Xc ,Yc) and (Xp ,Yp) respectively, then the Euclidian distance between centers are expressed as shown in equation

$$\sqrt{(Xc-Xp)^2+(Yc-Yp)^2}<\mathrm{T} \qquad (2)$$

There are various numbers of object (blobs) in the current and previous frame $I_n$and I_{n-1}. Let L_{n-1} and L_n be the number of objects (blobs) in these frames, respectively. There are three possible cases:

Case I :$L_n > L_{n-1}$

Case II: $L_n < L_{n-1}$

Case III: $L_n = L_{n-1}$

Case I: In this case the number of objects in the current frame is more than the number of objects in the previous frame. In this case we find a correspondence of objects in the current frame that have a correspondence with the previous frame rest of the object in the current frame not tracked [5]. Here, a number of not tracked object is ($L_n - L_{n-1}$).

Case II: In this case the number of objects in the current frame is less than the number of objet in the previous frame. In this case we find correspondence of all the objects in the current frame that have that have corresponded with previous frame.

Case III: In this case the number of objects in the current frame is same as the number of objects in the previous frame .In this case we find correspondence of all objects in the current frame with all objects in the previous frame .In this case all objects are tracked.

V. CONCLUSION

To analyze images and extract high level information image enhancement, motion detection, object tracking and behavior understanding researchers have studied. In this paper, we have studied and presented different methods of

moving object detection used in video surveillance. We have discussed adaptive and non-adaptive methods for foreground and background detection. The drawback of non adaptive method is that it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly. This article gives valuable insight into this important topic and encourages regorous further research in the area of moving object detection as well as in the field of computer vision. This paper can be considered as educational work based on the emerging field of machine vision. Although it is an attempt to present most comprehensive set of references. There is some possibility of omission. We apologize for any errors or omissions.

REFERENCES

[1] M. Kass , A Witkin and D TerzPoulos , snakes: active contour models, in j. comput. Vision vol.1,page 321-332,year 1988.

[2] V. C aselles, R .kimmel ,and G Sapiro ,Geodsic active contours,In IEEE international conference on computer vision ,page 694-699, year 1995.

[3] N. Paragios, and R. Deriche.. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intell. 22, 3, page 266–280, year 2000

[4] Comaniciu, D. And Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Analy. Mach. Intell. 24, 5, page 603–619.

[5] S. Zhu, and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Patt. Analy. Mach. Intell. 18, 9, page 884–900, year1996.

[6] Elgammal, A. Duraiswami, R.,Hairwood, D., Anddavis, L. 2002. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. Proceedings of IEEE vol 90, 7, page 1151–1163.

[7] Isard, M. And Maccormick, J. 2001. Bramble: A bayesian multiple-blob tracker. In IEEE International Conference on Computer Vision (ICCV). Page 34–41.

[8] 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, page 32-54, year 2008.

[9] Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Comput.Surv. 38, 4, Article 13,year 2006 .

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

[11] A. M. McIvor. Background subtraction techniques. Proc. of Image and Vision computing, year 2000

[12] C. Stauffer and E. Grimson, "Learning patterns of activity using realtime tracking," IEEE Trans. On Pattern Analysis and MachineIntelligence, vol. 22, no. 8, pp. 747-757, August 2000.

[13] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 809-830, August 2000.

[14] PrajnaParimita Dash, DiptiPatra, Sudhansu Kumar Mishra, "Kernel based Object Tracking using Color Histogram Technique " International Journal of Electronics and Electrical Engineering (IJEEE), ISSN : 2277-7040 Vol. 2, Issue. 4, PP.28-35, April 2012.

[15] PrajnaParimita Dash, DiptiPatra "Efficient Object Tracking Method Using LBP Based Texture Feature and OhtaColour Moment" International Journal of Electronics and Communication Engineering (IJECE), 1(2), 15 – 22,2012

[16] PrajnaParimita Dash, DiptiPatra, "Evolutionary Neural Network for Noise Cancellation from Image Data" Int. J. Computational Vision and Robotics, Inderscience publisher.Vol.2,no.3. pp.206-217,Oct 2011.