A Survey on Moving Object Detection and Tracking in Video Surveillance System

Kinjal A Joshi, Darshak G. Thakore

Abstract* — This paper presents a survey of various techniques related to video surveillance system improving the security. The goal of this paper is to review of various moving object detection and object tracking methods. This paper focuses on detection of moving objects in video surveillance system then tracking the detected objects in the scene. Moving Object detection is first low level important task for any video surveillance application. Detection of moving object is a challenging task. Tracking is required in higher level applications that require the location and shape of object in every frame. In this survey, I described Background subtraction with alpha, statistical method, Eigen background Subtraction and Temporal frame differencing to detect moving object. I also described tracking method based on point tracking, kernel tracking and silhouette tracking.

Keywords:Object detection, background subtraction, Temporal frame diiferencing, object tracking, video surveillance, statistical methods

I. INTRODUCTION OF VIDEO SURVEILLANCE

Video surveillance is a process of analyzing video sequences. It is an active area in computer vision. It gives huge amount of data storage and display. There are three types of Video surveillance activities. Video surveillance activities can be manual, semi-autonomous or fully-autonomous [10]. Manual video surveillance involves analysis of the video content by a human. Such systems are currently widely used. Semi-autonomous video surveillance involves some form of video processing but with significant human intervention. Typical examples are systems that perform simple motion detection [5]. Only in the presence of significant motion the video is recorded and sent for analysis by a human expert. By a fully-autonomous system [10], only input is the video sequence taken at the scene where surveillance is performed. In such a system there is no human intervention and the system does both the low-level tasks, like motion detection and tracking, and also high-level decision making tasks like abnormal event detection and gesture recognition. Video surveillance system that supports automated objects classification and object tracking. Monitoring of video for long duration by human operator is impractical and infeasible. Automatic motion detection which can provide batter human attention [9].There is varieties of applications in video surveillance like access Control, person identification, and anomaly detection. Intelligent visual surveillance (IVS) refers to an automated visual monitoring process that involves analysis and interpretation of object behaviors, as well as object detection and tracking, to understand the visual events of the scene [11]. Main tasks of IVS include scene interpretation and wide area surveillance control. Scene interpretation detects and track moving objects in an image sequence. It is used to understand their behaviors.

II. MOVING OBJECT DETECTION

Moving object detection is the basic step for further analysis of video. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. It handles segmentation of moving objects from stationary background objects [3]. This focuses on higher level processing .It also decreases computation time. Due to environmental conditions like illumination changes, shadow object segmentation becomes difficult and significant problem. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections [16]. This temporal information is usually in the form of frame differencing, which highlights regions that changes dynamically in consecutive frames. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks. This section reviews three moving object detection methods that are background subtraction with alpha parameter, temporal difference, and statistical methods, Eigen Background Subtraction.

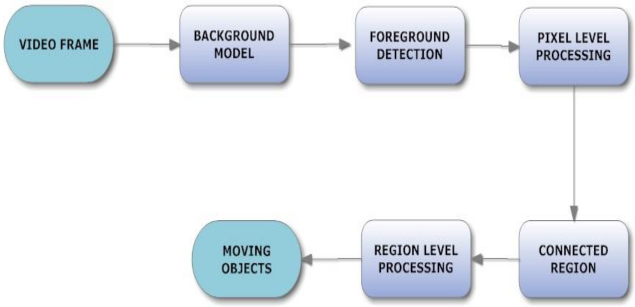

Figure 1: Framework of Moving Object Detection System [3]

The first step is to distinguish foreground objects from stationary background. To achieve this, we can use a combination of various techniques along with low-level image post-processing methods to create a foreground pixel map at every frame. We then group the connected regions in the foreground map to extract individual object features such as bounding box, area, perimeter etc.

Manuscript received on July, 2012

Kinjal A Joshi, PG student, Computer engineering Department, BVM Engineering collage, Anand, India,

Darshak G Thakore, Computer engineering Department, BVM Engineeringcollage,Anand,

2.1 Foreground Detection

The main purpose of foreground detection is to distinguishing foreground objects from the stationary background. Almost, each of the video surveillance systems uses the first step is detecting foreground objects. This creates a focus of attention for higher processing levels such as tracking, classification and behavior understanding and reduces computation time considerably since only pixels belonging to foreground objects need to be dealt with [1].

The first step is the background scene initialization. There are various techniques used to model the background scene. The background scene related parts of the system is isolated and its coupling with other modules is kept minimum to let the whole detection system to work flexibly with any one of the background models [8].

Next step in the detection method is detecting the foreground pixels by using the background model and the current image from video. This pixel-level detection process is dependent on the background model in use and it is used to update the background model to adapt to dynamic scene changes [5]. Also, due to camera noise or environmental effects the detected foreground pixel map contains noise. Pixel-level post-processing operations are performed to remove noise in the foreground pixels. Once we get the filtered foreground pixels, in the next step, connected regions are found by using a connected component labeling algorithm and objects' bounding rectangles are calculated. The labeled regions may contain near but disjoint regions due to defects in foreground segmentation process. Hence, some relatively small regions caused by environmental noise are eliminated in the region-level post-processing step [20]. In the final step of the detection process, a number of object features like area, bounding box, perimeter of the regions corresponding to objects are extracted from current image by using the foreground pixel map.

2.2 Pixel Level Post-Processing

The output of foreground detection contains noise. Generally, it affects by various noise factors. To overcome this dilemma of noise, it 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 is the noise caused by the camera's image acquisition components. This noise is produce because of the intensity of a pixel that corresponds to an edge between two different colored 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 [16].

Background Colored Object Noise: The color of the object may have the same color as the reference background. difficult to detect foreground pixels with the help of reference background [16].

Reflectance Noise: Reflectance noise is caused by light source. When a light source moves from one position to another, some parts in the background scene reflect light [16].

We can use low pass filter and morphological operations, erosion and dilation, to the foreground pixel map to remove noise that is caused by the items listed above [3]. Our aim in applying these operations is removing noisy foreground pixels that do not correspond to actual foreground regions, and 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 pre-processing tasks, such as removal of small details from an image prior to large object extraction, and bridging of small gapes in lines or curves. Gaussian low pass filter is use for pixel level post processing [8].A Gaussian filters smoothes an image by calculating weighted averages in a filter co-efficient [10]. Gaussian filter modifies the input signal by convolution with a Gaussian function.

2.3 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. After finding individual regions that correspond to objects, the bounding boxes of these regions are calculated.

2.4 Region Level Post-Processing

As pixel-level noise removed, still some artificial small regions remain just because of the bad segmentation. To remove this type of regions, regions that have smaller sizes than a pre-defined threshold are deleted from the foreground pixel map. Once segmenting regions we can extract features of the corresponding objects from the current image. These features are size, center-of-mass or just centroid and Bounded Area of the connected component. These features are used for object tracking and classification for the further processing in event detection.

A. Background Subtraction with Alpha

Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. Usually, a connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is referred to as the background subtraction [6].

Heikkila and Silven [6] presented this technique. At the start of the system reference background is initialized with first few frames of video frame and that are updated to adapt dynamic changes in the scene. At each new frame foreground pixels are detected by subtracting intensity values from background and filtering absolute value of differences with dynamic threshold per pixel [8] .The threshold and reference background are updated using foreground pixel information. It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialized period [6]. The pixels where the difference is above a threshold are classified as foreground. After creating foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt to dynamic scene changes.

Pixel is marked as foreground if the inequality is satisfied [3],

$$| I_t(x, y) - B_t(x, y) | > T \qquad (1)$$

Where T is a pre-defined threshold. The background image Bt is updated by the use of a first order recursive filter as shown in Equation

$$B_{t+1} = \alpha I_t + (1 - \alpha)B_{t\,a} \qquad (2)$$

Where α is an adaptation coefficient. The basic idea is to provide the new incoming information into the current background image. After that, the faster new changes in the scene are updated to the background frame. However, α cannot be too large because it may cause artificial "tails" to be formed behind the moving objects. The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions.

B. Statistical Methods

To overcome the shortcoming of the basic background methods, statistical Methods are used. Statistical methods are used to extract change regions from background. These statistical methods are mainly inspired by the background subtraction methods. It uses characteristics of individual pixels of group of pixels to construct advance background model. That statistics of background are updated dynamically during processing. At each frame this method keeps and updates dynamic statistics of pixels that belongs to background image process [3]. Foreground pixels are identified by comparing each pixel's statistics with that of the background model. This approach is becoming more popular due to its reliability in scenes that contain noise, illumination changes and shadow [8]. One of the example of statistical methods, Stauffer and Grimson [5] described an adaptive background mixture modeled by a mixture of Gaussians which are updated on-line by incoming image data. In order to detect whether a pixel belongs to a foreground or background process, the Gaussian distributions of the mixture model for that pixel are evaluated.

C. Temporal Differencing

Temporal differencing method uses the pixel-wise difference between two or three consecutive frames in video imagery to extract moving regions. It is a highly adaptive approach to dynamic scene changes however, it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly [3]. When a foreground object stops moving, temporal differencing method fails in detecting a change between consecutive frames and loses the object. Let $I_n(x)$ represent the gray-level intensity value at pixel position x and at time instance n of video image sequence I, which is in the range [0, 255]. T is the threshold initially set to a pre-determined value. Lipton et al.[3] developed two-frame temporal differencing scheme suggests that a pixel is moving if it satisfies the following [3]:

$$| I_n(x) - I_{n-1}(x) | > T \qquad (3)$$

This method is computationally less complex and adaptive to dynamic changes in the video frames. In temporal difference technique, extraction of moving pixel is simple and fast. Temporal difference may left holes in foreground objects, and is more sensitive to the threshold value when determining the changes within difference of consecutive video frames [5]. Temporal difference require special supportive algorithm to detect stopped objects.

D. Eigen background Subtraction

Eigen background subtraction [2] proposed by Oliver, *et al.* It presents that an Eigen space model for moving object segmentation. 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 the reduced space after PCA should represent only the static parts of the scene, remaining moving objects, if an image is projected on this space. The main steps of the algorithm can be summarized as follows [8]:

- A sample of N images of the scene is obtained; mean background image, μb, is calculated and mean normalized images are arranged as the columns of a matrix, A.
- The covariance matrix, C=AAT, is computed.
- Using the covariance matrix C, the diagonal matrix of its Eigen values, L, and the eigenvector matrix, Φ, is computed.
- The M eigenvectors, having the largest Eigen values (Eigen backgrounds), is retained and these vectors form the background model for the scene.
- If a new frame, first projected onto the space spanned by M eigenvectors and the reconstructed frame I' is obtained by using the projection coefficients and the eigenvectors.
- The difference I - I' is computed. Since the subspace formed by the eigenvectors well represents only the static parts of the scene, outcome of the difference will be the desired change mask including the moving objects.

III. OBJECT TRACKING

Object tracking is the important issue in human motion analysis. It is higher level computer vision problem. Tracking involves matching detected foreground objects between consecutive frames using different feature of object like motion, velocity, color, texture. Object tracking is the process to track the object over the time by locating its position in every frame of the video in surveillance system. It may also complete region in the image that is occupied by the object at every time instant [9]. In tracking approach, the objects are represented using the shape or appearance models [7]. 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 the case where a geometric shape representation like an ellipse is used for the object, parametric motion models like affine or projective transformations are appropriate [15]. These representations can approximate the motion of rigid objects in the scene. For a non rigid object, silhouette or contour is the most descriptive representation and both parametric and nonparametric models can be used to specify their motion. Different object tracking methods are described as follows.

Point Tracking: Point tracking is robust, reliable and accurate tracking method developed by Veenman et al[9]. This method is generally is used for to track the vehicles. This approach requires good level of fitness of detected object. This method require deterministic or probabilistic methods [10].object is tracked is based on point which is represented in detected object 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 objects in every frame..

Kernel Tracking: In this approach kernel require shape and appearance of the object [9]. In this approach any feature of object is used to track object as kernel like rectangular template or an elliptic shape with an associated histogram. After computing the motion of the kernel between consecutive frames object can be tracked. In [4]. Mean-shift tracking is based on the kernel tracking method used. In this method E-kernel is used. It represents histogram feature based by spatial masking with an isotropic kernel.

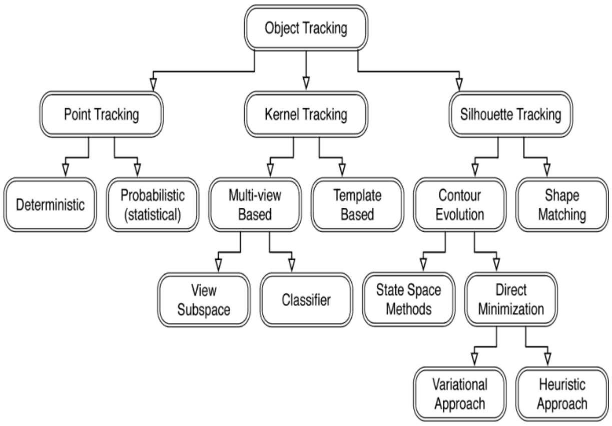

Figure 2: Taxonomy of tracking methods [9]

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. Silhouette tracking methods make use of the information stored inside the object region [6]. This information of the region can be appearance density and shape models. Given the object models,

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

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

Edges: Object boundaries usually generate strong changes in image intensities [18]. 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.

Centroid: The Center of mass (centroid) is 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 (or x-coordinate) of the center of mass, and the second element is the vertical coordinate (or y-coordinate) [16].

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

Among all features color and texture features are widely used to track the object. Color bands are sensitive to illumination variation.

E. Correspondence Based Matching Algorithm

In correspondence based object matching algorithm, we take the objects of the previous frame and the objects of the current frame, and match the pairs which are close. In this method we compute the distance between the centroid that is smaller than a pre-defined threshold T [7]. For example, suppose two objects (Oc and Op, c for current frame and p for previous frame) with center of mass (xc, yc) and (xp, yp) respectively, then the Euclidian distance between centers expressed as shown in equation

$$\sqrt{(x_c - x_p)^2 + (y_c - y_p)^2} < \mathrm{T} \qquad (4)$$

There are varies number of objects (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 numbers of objects in the current frame are more than the number of object in the previous frame. In this case we find correspondence of the objects in the current frame that have correspondence with the previous frame rest of the objects in the current frame not tracked [5]. Here, numbers of not tracked objects are (L_n – L_{n-1}).

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

Case III: In this case the numbers of objects in the current frame are less than the number of object in the previous frame. In this case we find correspondence of the all objects in the current frame that have correspondence with the previous frame.

IV CONCLUSION

To analyze images and extract high level information, image enhancement, motion detection, object tracking and behavior understanding researches have been studied. In this paper, we have studied and presented different methods of moving object detection, used in video surveillance. We have described background subtraction with alpha, temporal differencing, statistical methods. Detection techniques into various categories, here, we also discuss the related issues, to the moving object detection technique. The drawback of temporal differencing is that it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly. When a foreground object stops moving, temporal differencing method fails in detecting a change between consecutive frames and loses the track of the object. We presented detail of background subtraction method in deep because of its computational effectiveness and accuracy. This article gives valuable insight into this important research topic and encourages the new research in the area of moving object detection as well as in the field of computer vision. Here research on object tracking can be classified as point tracking, kernel tracking and contour tracking according to the representation method of a target object. In point tracking approach, statistical filtering method has been used to estimating the state of target object. Kalman filter and particle filter are the most popular filtering method. In kernel tracking approach, various estimating methods are used to find corresponding region to target object. Now a day, the most preferred and popular kernel tracking techniques are based on Mean-shift tracking and particle filter. Contour tracking can be divided into state space method and energy function minimization method according to the way of evolving of contours.

REFERENCES

[1] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. Int. J. Comput. Vision 1, 321–332, 1988.

[2] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In IEEE International Conference on Computer Vision . 694–699, 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, 266–280, 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, 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, 884–900, 1996.

[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 90, 7, 1151–1163.

[7] Isard, M. And Maccormick, J. 2001. Bramble: A bayesian multiple-blob tracker. In IEEE International Conference on Computer Vision (ICCV). 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, pp 32-54, Apr, 2008.

[9] Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Comput. Surv. 38, 4, Article 13,December 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):926-939

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

[12] C. Stauffer and E. Grimson, “Learning patterns of activity using real time tracking,” IEEE Trans. On Pattern Analysis and Machine Intelligence, 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] Maniciu, D. And Meer, p. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Analy. Mach. Intell. 24, 5, 603–619.

[15] ISARD, M. AND MACCORMICK, J. 2001. Bramble: A bayesian multiple-blob tracker. In IEEE International Conference on Computer Vision (ICCV). 34–41.

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

[17] Dockstader, S. And Tekalp, A. M. 2001a. Multiple camera tracking of interacting and occluded human motion. Proceedings of the IEEE 89, 1441–1455.

[18] Christopher R. Wren, Ali J. Azarbayejani, Trevor Darrell, and Alex P.Pentland, ”Pfinder: Real-Time Tracking of the Human Body” in IEEETransactions on Pattern Analysis and Machine Intelligence, July 1997,19(7), pp. 780-785.

[19]] Chris Stauffer and Eric Grimson, ”Learning Patterns of Activity UsingReal-Time Tracking” in IEEE Transactions on Pattern Recognition and Machine Intelligence (TPAMI),

[20] WuU Z, and Leahy R. “An optimal graph theoretic approach to data clustering: Theory and its applications to image segmentation”. IEEE Trans. Patt. Analy. Mach. Intell. 1993.