Object Detection and Tracking Techniques: A Survey

R Mohan
National Institute of Technology
Tiruchirappalli 620015
India

Midhula Vijayan
National Institute of Technology
Tiruchirappalli 620015
India

Athira AP
National Institute of Technology
Tiruchirappalli 620015
India

Abstract— **The objective of this literature survey is to appraise the avant-garde object detection and tracking methods, categorize them and appreciate the possibilities of the methods by identifying the latest trends in the area. In general, object detection and tracking is a demanding area of research. Object detection methods are used for recognizing the presence of objects in an image or a video sequence whereas Object tracking is used to supervise an object's temporal and spatial transforms throughout a video sequence. Understanding of these techniques is vital for various applications, like video surveillance, human-computer interfaces, and traffic monitoring, robotic systems and health care. There are many algorithms that implement the methods for object detection, classification, representation and essentially tracking. This paper presents different schemes or techniques which are applied for detecting, tracking and identifying objects in a video.**

Keywords— *Object detection, Background subtraction, tracking methods, Mixture of Gaussians*

I. INTRODUCTION

The modern world has enormous mass of digital visual information. To analyze and understand this huge collection of visual data, there subsist several image analysis methods. The technique that detects the object does have significant use and provide a considerable platform for modern applications. The syntactic and semantic meaning of the videos and images can be detected and further analyzed to retrieve the essential information. The probable applications of the image can be recognized using this technique. The most significant content of an image is the objects in it. For that reason, there arises a significant need for better object detection techniques.

Since in most cases objects in motion are the key source of information, many methods have been advised on the detection of these objects. Object detection and tracking decreases hard work, human interpretation and hence improves efficiency. Automatic detection and tracking put in to the smart systems used now days. In tracking, an object is anything which is of concern. An object may be ships in the sea, animals inside a pool, cars on a road, birds in a tree etc. Birds in a tree may be a collection of objects that is required to be tracked for an explicit reason. Representation of an object is very important for obtaining an apt output whilst performing object detection and tracking. There are different ways for representation of objects.

Video in reality is a sequel of images and each image can be further divided into a frame. Furthermore, the successive frames are generally related closely. Frames are played in a higher frequency rate for the eyes to perceive the continuation of content. It is understood from this concept that individual frames may be applied with most of image processing methods.

The three fundamental steps [1] [2] used for tracking an object as per most of the literatures are:-

- Object Detection

In object detection, the object of interest is identified from a video sequence and the pixels of the same are clustered. There are several methods such as background subtraction, frame differencing etc for implementing object detection.

- Object Classification

Any item that is of interest to the user can be termed as an object. It can be of any shape and size. Therefore, there exist different approaches to classify these objects. Classification can be based on shape, colour, texture etc.

- Object Tracking

The path on which the object moves inside an image is approximated as it shifts from scene to scene and is termed as object tracking. Various tracking approaches include point tracking, kernel tracking etc.

In this survey we focus on various object detection and tracking approaches. Extracting the motion part from a video stream is termed as Non-stationery object detection. It separates objects in motion from a background frame. To recognize object of our interest from a video sequel and assembling pixels of that object is referred as Object Detection. A generic step towards object detection is to utilize data from an individual frame or from a set of frames within a time period, which means, temporal information is calculated from a series of frames so as to avoid bogus detections. The pixels are classified as foreground pixels and background pixels of a video stream for further detection of moving object. ie; images from a video can be dispersed into two groups- one with pixels that correspond to background scene and other with pixels of foreground object. The description of foreground and background that is to be recognized is more or less application specific. The product that is detected is mostly showed as a mask or a binary image.

II. DIFFICULTIES IN DETECTION OF OBJECTS

- Illumination: The lightning in the same location may change along the duration of a day. The environmental

circumstances may also influence the light of an image [3] [4].

- Positioning: The location of an object may vary from one position to the other, causing the system not to identify a particular object [3] [4].

- Rotation: The system must be able to detect an image which is in rotated form, in case of methods like shape based object detection [3][4].

- Mirroring: An object detection system must be able to handle situations where mirrored images of the objects are formed.

- Occlusion: Occlusion is a condition where an object is not completely visible in the image. In such situations, there are chances of not detecting useful portions [3] [4].

- Scale: The scale or size of the object in a frame may not be affecting a good object detection system.

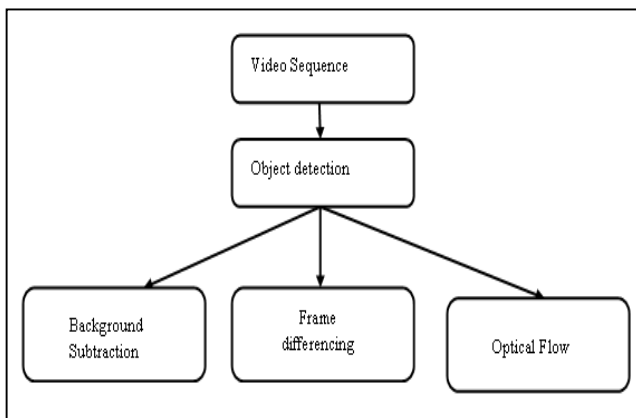- Noisy Images: Image with deprived quality

III. OBJECT DETECTION

Fig.1. Object detection techniques

A. Frame Differencing

Existence of a dynamic object is found by differencing two consecutive frames. The method uses a reference background image for comparison purposes [1][5]. The areas where there are changes are recognized as moving objects. i.e. simple differencing followed by thresholding shows moving objects. This method works well for videos with static background. There are advantages like good accuracy and easiness in implementation. The demerits of this method are difficulty in obtaining an absolute outline as well as difficulty in performing with non-static background.

B. Optical Flow

Optical flow estimation gives a two-dimensional vector field, which means, motion field that denotes velocities and directions of each point of an image sequence. Clustering is done based on the optical flow distribution characteristics [1][5]. It defines direction and time in a time sequence of two subsequent dimensional velocity vectors. The advantage is that it can produce accurate object motion information.

To detect areas in motion inside an image, optical methods [6][7][8] use flow vectors of dynamic objects over time. In this, the perceptible direction and velocity of all pixels in an image frame is to be calculated. The method is effective except for its time intense process. Optic flow can be used to calculate the Background motion model that provides to even out an image in the background. Individual motion can be recognized by this method by the flow of direction of the image gradient which is not envisaged by the background. It can even recognize motion in a video from a non-stationary camera as well as dynamic background but many of these techniques cannot be utilized in real-time applications devoid of dedicated hardware. Moreover, this method is complex as it needs enormous number of calculations.

C. Background Subtraction

This method [1], is one of the most known and used method for movement detection. This technique finds difference of a target image and the background image in order to detect the moving portion, and is usually efficient to deliver information of the object. i.e. the background frame is deducted from the target frame. If the variation is larger than the threshold value (already set), then it is identified as the pixels of dynamic object, or else of the background.

The major step of this technique is the initialization and updation of background frame. The correctness of result is influenced by the effectiveness of these factors. It detects dynamic objects by deducting the present image pixel-by-pixel from a background image (reference image) formed during initialization. If the disparity is beyond the threshold value, then the object is taken as foreground. A fine background subtraction algorithm should be able to manage the objects which at first submerge into the background and after sometime becomes foreground. Moreover, a background subtraction algorithm should be computationally cost-effective and uses less memory, but capable of identifying dynamic objects accurately from a video sequence so as to adjust to the real-time applications.
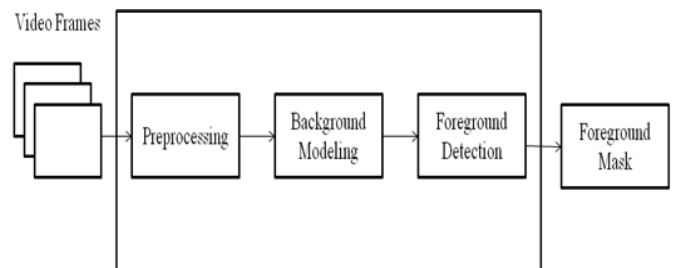

Fig.2. Illustration of Background subtraction

The result of image frames computed by this method can be classified as follows:-

- When there is nil motion of object in a scene.
If there is no motion of objects between two image frames, the output will be a black binary image. This implies that there is no difference between the pixels.

- When there is motion of object in a scene.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACC - 2016 Conference Proceedings

If there is motion of objects from a frame to another frame, then the binary image output will be the difference between these two frames. The objects that displays movement is shown in white colour and the portion where there is no difference is shown in black.
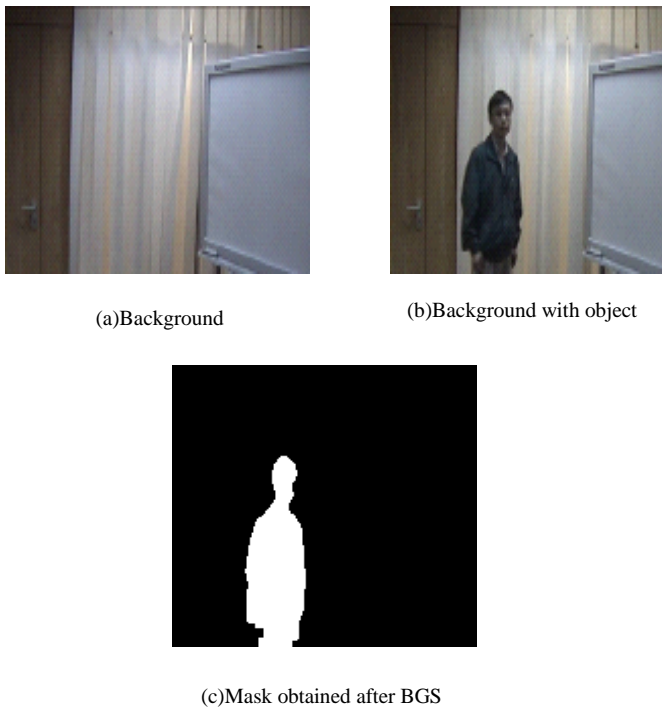

(a)Background   (b)Background with object

(c)Mask obtained after BGS

Fig. 3. Example for background subtraction

1) Basic Classification

(a)   Recursive algorithm

In this type of algorithms [1][9], there is no buffer maintained for background inference. The algorithm recursively updates one background model according to each input frame. Due to this, background model can be affected by frames from distant past. The various methods include Approximate median technique, Adaptive background technique and Mixture of Gaussians.

- *Approximate median Technique*: According to this method, running estimation of median is increased by one if the pixel input is superior to the estimation, and decremented by one if it is inferior. It may not provide smooth result in all conditions is one of the disadvantages of this method. In this case, threshold for each frame can automatically be updated to segment foreground images.

- *Adaptive background Technique*: A background model is constructed using frames during a particular time period. The background keeps getting updated as each frame progresses. The algorithm shows increased efficiency in case of the segmented results.

- *Gaussian mixtures*: In this, it has a background scheme that is parametric. A pixel position can be represented using a mixture of Gaussian functions.

Gaussian mixture method presents enhanced results by modified parameters but is complex and tiresome.

(b)   Non-Recursive algorithm

For background estimation, non-recursive algorithms [1][9] employs sliding-window approach. In this approach, it saves a buffer of preceding video frames, and then approximates a background model according to the temporal difference of the pixels inside the buffer. These algorithms do not rely on the history of the frames saved in the buffer and hence incredibly adaptive. But the necessity for a large storage may arise when a hefty buffer is required to deal with the traffic. Frame difference method and median filtering are among the most common used non-recursive algorithms.

2) Background Subtraction Principles

(a)   When an object appears for the first time in a scene, an object of interest is to be detected using background subtraction [10].

(b) A suitable pixel level criteria has to be defined. Those pixels, that adheres to this criteria are considered as background and disregarded [10].

(c) Both gradual and sudden changes in the background are to be tailored by the background model [10].

3)   Background Subtraction methods

(a) Threshold based BS

The various approaches to Threshold based background subtraction are in respect of background maintenance, post processing and foreground region detection. In [11] authors have used this scheme where a background model B_m is created and then the difference between background model and intensity value of current frame is calculated. If the value obtained is larger than the predefined threshold, it is identified as an object or else considered background. Small sized regions are eliminated and morphological closing is performed after creation of a foreground pixel map.
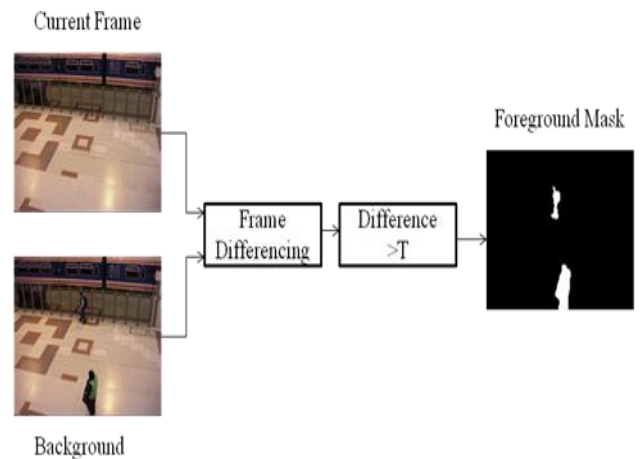

Fig. 4. Threshold based background subtraction

(b) Mean and Median based BS

The mean value of first n frames is calculated to create the background model. B (α, β, t) is to be calculated from n frames for the tth frame. If the variation or difference between intensity value of the present frame, and the background model, is larger than the predefined threshold, then the pixel (α, β) belongs to the object. Instead of mean value, the algorithm uses median value in median background subtraction model. Both the methods are straightforward to implement and use. The background models change over time and not constant and this feature is the biggest advantage when considering these models. Object speed and frame rate are the key factors that decides the accuracy of the frame. High memory requirements are a demerit for both these methods.

(c) Gaussian Averaging

For detecting non-stationary objects from dynamic background scenarios, Gaussian averaging [12] and multi-level Gaussian averaging [13] can be applied. In these approaches, firstly a background model is constructed by scrutinizing a series of formerly available frames. This is performed by taking into account the variance and average of intensity value for a particular position of the frame from the series of formerly observed frames of that video. Only a solitary set of parameters are necessary, if single Gaussian probability density function (PDF) is considered. Likewise, a multi-valued Gaussian PDF requires multiple set of parameters.

IV. OBJECT CLASSIFICATION

A. Template Matching

The method of finding the fractions of an image that matches with the template is termed as Template matching, which is a straightforward process. Template Matching is perhaps the finest method for several precise situations. It is correct enough but occasionally there is a possibility of lack of novelty in the object detected. The object can be recognized for a particular video by means of a template that has been chosen from that video. Nevertheless, there is no assured accurateness since all that it knows is the finest match for every frame. The algorithm barely works if in case, the object is constantly present in the video, or else it may show fake detection.

B. Colour Based

The object detection using colours involved in the objects is also significantly used and provide a simple to implement method. It provides potent information for object recognition. Color histograms prove to be simple and efficient and provide an edge for the same. This information has been segmented into two approaches which is the part based approach and the efficient sub-window approach. Feature combination, photometric invariance and compactness are the three major features that need to be taken into account while integrating or appending the color attributes with the object detection.

C. Shape Based

Lately, shape has proved to be of great importance in object recognition. They have been looked at considerably to recognize objects in suitable real world images. These features also provide an upper hand over local features like SIFT as most of the objects are illustrated and described by their shapes and textures such as different animals and other varying objects. They are most likely used to add an advantage to the local features.

V. OBJECT REPRESENTATION

In a tracking environment, an object is something that is of interest to the user for advanced analysis. Shapes and appearances can be used to represent an object. In the case of object detection and tracking, representation of an object is an important factor. There are several ways for representing objects and here we will be considering object shape depictions generally used for tracking and deal with shape and appearance representations.

A. *Points*. In this type of representation, an object is represented as a point. ie; by considering the centroid or by using a collection of points. Ideally, points are appropriate for tracking objects which engage only modest region in the image.
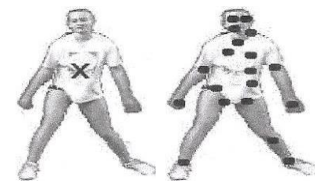

Fig.5. Point Representation

B. *Primitive geometric shapes.* Ellipse, rectangle etc can be used to represent the shape of an object. Affine, translation, or projective transformation is used for representing motion of the object. These primal geometric shapes are apt for representing objects with rigid shape. However, can also be used for tracking objects with non-rigid shape.
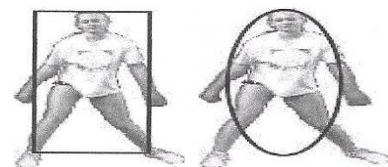

Fig. 6. Primitive shape Representation

C. *Object silhouette and contour.* Contours are used to define the boundary of an object. Silhouette of an object is the area inside this contour. Non-rigid shaped objects can be tracked by Silhouette and contour representations.
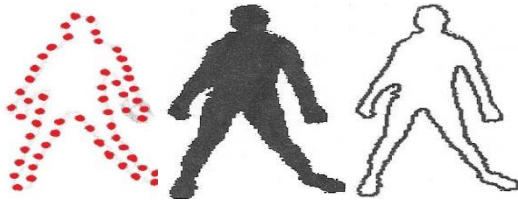
Fig.7. Object Silhoutte & Contour Representation

D.   *Articulated shape models.* Body parts which are connected together by joints compose articulated objects. That is, in case of human body, it is articulated by joints connecting chest, hands, legs, head, and feet. Kinematic motion models govern the relationship between the parts of the body, ie; joint, angle, etc. Body parts can be modeled using ellipses or cylinders, so as to represent an articulated object.
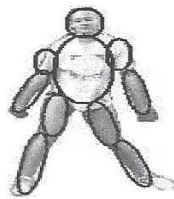

Fig.8. Articulated representation

E.   *Skeletal models.* Application of medial axis transform to the profile of an object is used for extracting an object skeleton. This is usually used for representing shapes to recognize objects. Both articulated and firm (rigid) objects can be represented using this.
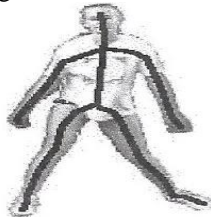

Fig. 9. Skeletal model representation

VI.   OBJECT TRACKING

One of the demanding applications these days is tracking an object from a video sequel. Tracking objects have wide range of applications like surveillance, Industrial inspection, Robotics, lane detection etc and is much more challenging to get better performance for recognition and tracking. It is a vital problem in analysis of human motion. Like every methods, existing object tracking methods have some drawbacks. The procedure by which an object is located by its position in every frame in a video for tracking anything of interest in it by time is termed as object tracking [14]. Appearance or shape models are used to represent an object in a tracking method [15]. The type of movement is limited by the model used to represent the object.

That is, only translational model can be applied if an object is shown as point or parametric models like geometric representations. Movement of the rigid objects in a scene can be approximated using these representations. For non rigid objects, parametric and nonparametric models can be used to specify the motion. There exist various models for object tracking like, region-based, contour-based and feature point-based models.

A.   *Contour-based object tracking model*

Finding outline of object from an image can be done by using active contour model. The objects can be tracked by considering the boundary contours in a contour-based tracking algorithm. Subsequently the contours are dynamically updated in succeeding frames. Nevertheless, the algorithm is very susceptible to tracking initialization and hence hard to initiate automatically.

A contour based tracking algorithm was proposed by Xu and Ahuja [16] to track contours throughout a video. In the paper, they have implemented the algorithm by applying the graph-cut image segmentation technique so as to segment the active contour. Resultant contour of preceding frame is the initialization of each frame. Contour of the new object is obtained using intensity information of present frame and difference of present frame and the preceding frame.

B.   *Region based object tracking model*

Colour distribution of the tracked object has been considered in region based object model for tracking an object [17]. The object is represented by its colour and therefore, is computationally competent. On the other hand, its competency gets degraded when a number of objects move together in an image. Due to occlusion in the presence of multiple object movement, it is difficult to attain accuracy in tracking. In addition, tracking is mostly reliant on the background model utilized for the extraction of object outline, in the deficiency of any object shape information.

C.   *Feature point based tracking*

Feature points [18] are used to define objects in a Feature point based model. Three fundamental steps are there in feature point based tracking algorithm. Recognition and tracking of object by extraction of elements is the first step. Clustering the elements into advanced level features is the next step. The final step is matching of extracted features with images in the succeeding frames. Feature extraction and correspondence are significant steps in feature point based tracking. Feature correspondence is the demanding issue in feature based object tracking because feature point in an image may perhaps have numerous analogous points in another image. This may be a cause of ambiguity in feature correspondence.

VII.   CONCLUSION

This survey paper elaborates object detection and tracking techniques that can enable readers to assimilate the techniques promptly. The process of tracking an object can be divided into three groups, object detection, object classification and object tracking. The paper makes a note of various methods for detecting moving objects as well as tracking them. Various techniques like background subtraction, optic flow and frame differencing have been portrayed along with frequently used object tracking models. Further revision might be performed out to study proficient method that reduces computational cost and time.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICACC - 2016 Conference Proceedings

REFERENCES

[1] Himani Parekh, Darshan Thakore, Udesang Jaliya,"A Survey On Object Detection And Tracking Methods", International Journal of Innovative Research in Computer and Communication Engineering, February 2014.

[2] Sheetal Khadse, Priti Vetal, Rucha Saraf, Snehal Gite , "Survey Paper on detection and tracking of moving objects", International Journal of Research In Science & Engineering, Volume: 1, Issue: 6

[3] Payal Panchal, Gaurav Prajapati, Savan Patel, Hinal Shah and Jitendra Nasriwala, "A Review on Object Detection and Tracking Methods", International Journal of Research In Emerging Science and Technology", Volume-2, Issue-1, January-2015

[4] Khushboo Khurana, Reetu Awasthi, "Techniques for Object Recognition in Images and Multi-Object Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 4, April 2013.

[5] Hemangi R. Patil, Prof. K. S. Bhagat," Detection and Tracking of Moving Object: A Survey", Int. Journal of Engineering Research and Applications, Vol. 5, Issue 11, (Part - 5) November 2015.

[6] N. Paragios, R. Deriche; "Geodesic active contours and level sets for the detection and tracking of moving objects", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (3), pp. 266–280, 2000.

[7] L.Wixson, "Detecting Salient Motion by Accumulating Directionally-Consistent Flow", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (8), 2000.

[8] Robert Pless, Tomas Brodsky and Yiannis Aloimonos, "Detecting Independent Motion: The Statistics of Temporal Continuity", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 (8), 2000.

[9] K.Srinivasan 1, K.Porkumaran 2, G.Sainarayanan, "Improved Background Subtraction Techniques for Security in Video Applications".

[10] At ICCV 1999, MS Research presented a study, Wallflower: Principles and Practice of Background Maintenance, by Kentaro Toyama, John Krumm, Barry Brumitt, Brian Meyers.

[11] J. Heikkila and O. Silven, "A real-time system for monitoring of cyclists and pedestrians", Proc. of 2nd IEEE Workshop on Visual Surveillance, pp. 74–81, 1999.

[12] Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: realtime tracking of the human body. IEEE Trans. Pattern Anal.Mach. Intell. **19**(7), 780–785 (1997).

[13] Stauffer, C., Grimson, W.: Learning patterns of activity using real time tracking. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 747–767 (2000)

[14] Yilmaz, A., Javed, O., and Shah, M. 2006. Object tracking: A survey. ACM Computer Survey 38, 4, Article 13,December 2006

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

[16] N Xu, N Ahuja, 'Object contour tracking using graph cuts based active contours', International Conference on Image Processing, pp. 277-280 vol.3, 2002.

[17] L. Li, S. Ranganath, H. Weimin, and K. Sengupta, "Framework for Real-Time Behavior Interpretation from Traffic Video", IEEE Tran. On Intelligen Transportation Systems, , Vol. 6, No. 1, pp. 43-53, 2005.

[18] Z Zivkovi, "Improving the selection of feature points for tracking", In Pattern Analysis and Applications, vol.7, no. 2, Copyright Springer-Verlag London Limited, 2004.