International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 02, Issue: 02 (MAR-APR, 2016)

Survey on Video Object Detection & Tracking

Astha Dixit[1], Manoj Verma[2], Kailash Patidar[3]
Computer Science & engineering Department
SSSIST, Sehore, India
[1]dastha67@gmail.com, [2]manoj.verma1511@gmail.com, [3]kailashPatidar123@gmail.com

Abstract: **-** Now a day's detection of video object is extensively used for sensing the motion positioning and occlusion of an image or video. The object detection and object classification are foregoing steps for tracking an object in succession of images. Object detection is performed to confirm the existence of objects in video and to specifically position that object. Subsequently detected object can be classified in different categories such as humans, automobiles, birds, floating clouds and other moving objects. Object tracking is performed using monitoring objects spatial and temporal transformation during a video sequence, together with its presence, location, size, and shape etc. This is used in numerous applications such as video surveillance, traffic monitoring robot vision, video in painting and animation. In this paper, we present the literature study of the formerly work done in the field of video object detection and techniques with their merits and demerits.

Keywords: **-** Object tracking, Traffic monitoring, Spatial, Temporal.

I. INTRODUCTION

Image processing is a term which indicates the processing on image or video frame which is taken as an input and the result set of processing is may be a set of related parameters of an image. The purpose of image processing is visualization which is to observe the objects that are not visible. Analysis of human motion is one of the most recent and popular research topics in digital image processing. In which the movement of human is the important part of human detection and motion analysis, the aim is to detect the motions of human from the background image in a video sequences. Object Tracking is a process of locating the object to associate the target in successive video frame over time and it finds wide scale applications in the field of security and surveillance, video communication, augmented reality, traffic control, medical imaging etc. Object Tracking is a complex process to be implemented in hardware mainly because of the amount of data associated with the video. Videos are actually sequences of images, each of which called a frame, displayed in fast enough frequency so that human eyes can percept the continuity of its content. It is obvious that all image processing techniques can be applied to individual frames. Besides, the contents of two consecutive frames are usually closely related [1]. The identification of regions of interest is typically the first step in many computer vision applications including event detection, video surveillance, and robotics. A general object detection algorithm may be desirable, but it is extremely difficult to properly handle unknown objects or objects with significant variations in color, shape and texture. Therefore, many practical computer vision systems assume a fixed camera environment, which makes the object detection process much more straightforward [2]. An image, usually from a video sequence, is divided into two complimentary sets of pixels. The first set contains the pixels which correspond to foreground objects while the second and complimentary set contains the background pixels. This output or result is often represented as a binary image or as a mask. It is difficult to specify an absolute standard with respect to what should be identified as foreground and what should be marked as background because this definition is somewhat application specific. Generally, foreground objects are moving objects like people, boats and cars and everything else is background [3]. Many a times shadow is classified as foreground object which gives improper output. The steps required to detect the features of the video objects is shown in figure 1. In this paper, presents the literature survey of the earlier work done for detection and tracking of the video object also discusses various techniques of video object detection. The organization of remaining section of paper is done as follows: In Section II discusses the literature of the work done for object tracking. Section III describes the various techniques of object tracking & detection and last section gives overall conclusion of the paper.

II. RELATED WORK

In this section discusses the previous work done in the field of detection and tracking of video object by several researchers using different image processing techniques.

Xu et al. [4] proposed a contour based object tracking algorithm to track object contours in video sequences. In their algorithm, they segmented the active contour using the graph-cut image segmentation method. The resulting contour of the previous frame is taken as initialization in each frame. New object contour is found out with the help of intensity information of current frame and difference of current frame and the previous frame.

Dokladal et al. [5] proposed approach is active contour based object tracking. For the driver's-face tracking problem they used the combination of feature-weighted gradient and contours of the object. In the segmentation step they computed the gradient of an image. They

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 02, Issue: 02 (MAR-APR, 2016)

proposed a gradient-based attraction field for object tracking.

Chen et al. [6] modeled an active contour based object tracking by Neural Fuzzy network. Contour based model is used to extract object's feature vector. For training and recognizing moving objects their approach uses the self-constructing neural fuzzy inference network. In this paper, they have taken the histograms of the silhouette of human body in horizontal and vertical projection and then transform it by Discrete Fourier Transform (DFT).

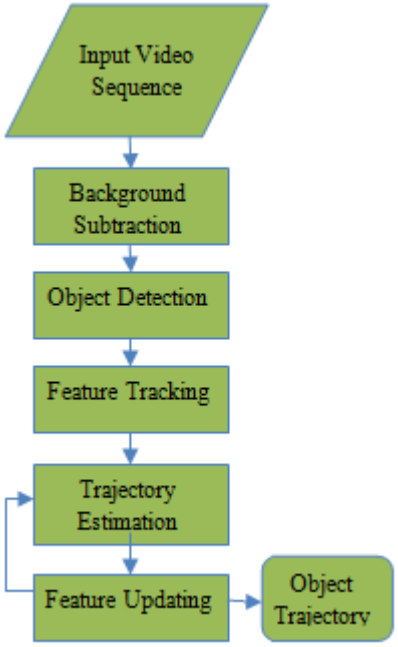

Fig. 1 Basic Video object tracking steps

Ling et al. [7] given an object tracking approach based on contours. The object rough location is found though multi-feature fusion strategy. For accurate and robust object contour tracking, they have extracted the contours with the help of region-based object contour extraction. In their model the object rough location is obtained by color histogram and Harris corner features fusion method. In the particle filer method they have used the Harris corner feature fusion method. Their model of region-based temporal differencing is applied in object contour detection step, and the resultant is the rough location tracking result.

Zhao et al. [8] presented an algorithm in which they first calculate the average of the values of the gray of the continuous multi-frame image in the dynamic image, and then get background image obtained by the statistical average of the continuous image sequence, that is, the continuous interception of the N-frame images are summed, and find the average. In this case, weight of object information has been increasing, and also restrains the static back-ground. Eventually the motion detection image contains both the target contour and more target information of the tar-get contour point from the background image, so as to achieve separating the moving target from the image. The simulation results show the effectiveness of the proposed algorithm.

Arunachalam et al. [9] presented the advance techniques for object detection and tracking in video. Most visual surveillance systems start with motion detection. Motion detection methods attempt to locate connected regions of pixels that represent the moving objects within the scene; different approaches include frame-to-frame difference, background subtraction and motion analysis. The motion detection can be achieved by Principle Component Analysis (PCA) and then separate an objects from background using background subtraction. The detected object can be segmented. Segmentation consists of two schemes: one for spatial segmentation and the other for temporal segmentation. Tracking approach can be done in each frame of detected Object. Pixel label problem can be alleviated by the MAP (Maximum a Posteriori) technique.

Cucchiara et al. [10] proposed an approach for detecting Vehicles in urban traffic scenes by means of rule-based reasoning on visual data. The strength of the approach is its formal separation between the low- level image processing modules (used for extracting visual data under various illumination conditions) and the high- level module, which provides a general purpose knowledge-based framework for tracking vehicles in the scene. The image-processing modules extract visual data from the scene by spatial-temporal analysis during daytime and by morphological analysis of headlights at night. The high-level module is designed as a forward chaining production rule system, working on symbolic data, i.e., vehicles and their attributes (area, pattern, direction, and others) and exploiting a set of heuristic rules tuned to urban traffic conditions. The synergy between the artificial intelligence techniques of the high- level and the low- level image analysis techniques provides the system with flexibility and robustness.

Nanda et al. [19] presented a novel algorithm for moving object detection and tracking. The proposed algorithm includes two schemes: one for spatial-temporal spatial segmentation and the other for temporal segmentation. A combination of these schemes is used to identify moving objects and to track them. A compound Markov random field (MRF) model is used as the prior image attribute model, which takes care of the spatial distribution of color, temporal color coherence and edge map in the temporal frames to obtain a spatial-temporal spatial segmentation. In this scheme, segmentation is considered as a pixel labeling problem and is solved using the maximum posteriori probability (MAP) estimation technique. The MRFMAP framework is computation intensive due to random initialization. To reduce this

International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 02, Issue: 02 (MAR-APR, 2016)

burden, we propose change information based heuristic initialization technique. The scheme requires an initially segmented frame. For initial frame segmentation, compound MRF model is used to model attributes and MAP estimate is obtained by a hybrid algorithm [combination of both simulated annealing (SA) and iterative conditional mode (ICM)] that converges fast. For temporal segmentation, instead of using a gray level difference based change detection mask (CDM), we propose a CDM based on label difference of two frames. The proposed scheme resulted in less effect of silhouette. Further, combination of both spatial and temporal segmentation processes used to detect the moving objects. Results of the proposed spatial segmentation approach are compared with those of JSEG method and edgeless and edge based approaches of segmentation. It is noticed that the proposed approach provides a better spatial segmentation compared to the other three methods.

III. VIDEO OBJECT DETECTION TECHNIQUES

In this section describes the various video object detection and tracking methodologies:

3.1 Background Subtraction

First step for background subtraction is background modeling. It is the core of background subtraction algorithm. Background Modeling must sensitive enough to recognize moving objects [11]. Background Modeling is to yield reference model. This reference model is used in background subtraction in which each video sequence is compared against the reference model to determine possible Variation. The variations between current video frames to that of the reference frame in terms of pixels signify existence of moving objects [11]. Currently, mean filter and median filter are widely used to realize background modeling. The background subtraction method is to use the difference method of the current image and background image to detect moving objects, with simple algorithm, but very sensitive to the changes in the external environment and has poor anti- interference ability. However, it can provide the most complete object information in the case background is known. As describe in [12], background subtraction has mainly two approaches:

3.1.1 Support Vector Machines: For a linear system, the available data can be clustered into two classes or groups by finding the maximum marginal hyper plane that separates one class from the other with the help of Support Vector Machines [14]. The distance of hyper plane and the closest data points helps in defining the margin of the maximized hyper plane. The data points that lie on the hyper plane margin boundary are called the support vectors. For object detection purpose the objects can be included in two classes, object class (positive samples) and the non-object class (negative samples). For applying SVM classifier to a nonlinear system, a kernel trick has to be applied to the input feature vector which is extracted from the input.

3.1.2 Adaptive Boosting: Boosting [13] is done by combining many base classifiers to find accurate results. In the first step of training phase of the Ad boost algorithm is an initial distribution of weights over the training set is constructed. The first step of Adaptive boosting is that the boosting mechanism selects the base classifier with least error. The error of the classifier is proportional to the misclassified data weights. Next, the misclassified data weights are increased which are selected by the base classifier. In the next iteration the algorithm selects another classifier that performs better on the misclassified data.

3.2 Frame Difference

The frame difference is the most effective method for detecting change of two adjacent frames in the video image [14]. Suppose the video frame at time *t* is given by f(x, y, t), then the next frame at $t + 1$ is f (x, y, t+1). The binary image operation results of frame difference can be defined as:

$$D(x, y, t+1) = f(x)$$

$$= \begin{cases} 1 & |f(x,y,t) - f(x,y,t+1)| > Th \\ 0 & Otherwise \end{cases}$$

Where T_h represents the threshold for decision, if the frame difference image value is greater than the thresh-old, then put the point as a foreground pixel. Similarly, when less than the threshold, regarding the point as a background pixel.

3.3 Optical flow

The translation of each pixel in a region can be found out by a dense field of displacement vectors defined as optical flow. Brightness constraint is taken as a measure while computing optical flow, assuming that brightness of corresponding pixels is constant inconsecutive frames. Optical flow feature is mostly used in motion-based object segmentation and tracking applications. Furthermore it is also used in video segmentation algorithms.

3.4 Spatio-temporal features

In recent times local spatio-temporal features are mostly used. These features provide a visual representation for recognition of actions and visual object detection [15]. Salient and motion patterns characteristics in video are captured by local spatiotemporal features. These features provide relative representation of events independently. While presenting events the spatio-temporal shifts and scales of events, background clutter and multiple motions in the scene are considered. To show the low level presentation of an object such as pedestrian space-time contours are used. To covert a one-dimensional contour

into three-dimensional space a 3D distance transform is used.

3.5 Kalman Filter

This filter technique is used for point tracking and they are based on Optimal Recursive Data Processing Algorithm. This filter technique performs the restrictive probability density propagation. Kalman filter [16] is a set of mathematical equations that provides an efficient Computational (recursive) means to estimate the state of a process in several aspects: it supports estimations of past, present, and even future states, and it can do the same even when the precise nature of the modeled system is unknown. It estimates a process by using a form of feedback control. The filter estimates the process state at some time and then obtains feedback in the form of noisy measurements. The equations for Kalman filters fall in two groups: time update equations and measurement update equations. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the priori estimate for the next time step. The measurement update equations are responsible for the feedback. Kalman filters always give optimal solutions.

3.6 Contour Tracking

Contour tracking methods [17], iteratively progress a primary contour in the previous frame to its new position in the current frame. This contour progress requires that certain amount of the object in the current frame overlay with the object region in the previous frame. Contour Tracking can be performed using two different approaches. The first approach uses state space models to model the contour shape and motion. The second approach directly evolves the contour by minimizing the contour energy using direct minimization techniques such as gradient descent. The most significant advantage of silhouettes tracking is their flexibility to handle a large variety of object shapes.

3.7 Multi-object Data Association & State Estimation

Kalman filter, extended kalman filter and particle give very good results when the objects are not close to each other. For tracking multiple objects in the video sequences by using Kalman or particle filters, the most likely measurement for a particular moving object needs to be associated with the object's state. This is called the correspondence problem [20]. So for multiple objects tracking the most important step we have solved is the correspondence problem before kalman or particle filters are applied. Nearest neighbor approach is the very simplest method to solve the correspondence problem. Data Association algorithms are used to associate the objects state like position, velocity, size with the available filters. Some of the methods to solve the data association are Linear Assignment problem (LAP), Stable Marriage problem (SMP) and Munkers algorithm etc. However the correspondence problem is hard to deal with when the moving objects are close to each other, and then the correspondence shows incorrect results. These filters fail to converge when incorrectly associated measurement occurs. There exist several statistical data association techniques to tackle this problem. Two mostly used techniques for data association in this complex scenario are Joint Probability Data Association Filtering (JPDAF) and Multiple Hypothesis Tracking (MHT).

3.8 Multiple Hypothesis Tracking (MHT)

Multiple hypotheses tracking (MHT) [18] is generally accepted as the preferred method for solving the data association problem in modern multiple target tracking (MTT) systems. It is an iterative algorithm, several frames have been observed for better tracking outcomes. Iteration begins with a set of existing track hypotheses. Each hypothesis is a crew of disconnected tracks. For each hypothesis, a prediction of object's position in the succeeding frame is made. The predictions are then compared by calculating a distance measure. MHT is capable of tracking multiple object, handles occlusions and calculation of optimal solutions.

3.9 Mean Shift Method

Mean-shift tracking [21] tries to find the area of a video frame that is locally most similar to a previously initialized model. The image region to be tracked is represented by a histogram. A gradient ascent procedure is used to move the tracker to the location that maximizes a similarity score between the model and the current image region. In object tracking algorithms target representation is mainly rectangular or elliptical region. It contain target model and target candidate. To characterize the target color histogram is chosen. Target model is generally represented by its probability density function (pdf). Target model is regularized by spatial masking with an asymmetric kernel.

IV. CONCLUSION

Object detection and tracking is becomes much essential these days. It is used in various surveillance system applications such as Traffic Monitoring, understanding of human activity, observation of people and vehicles within a busy environment, Security in Shopping Malls or Offices etc. Various techniques and algorithm has been developed to detect and track the motion of the video objects but those has some advantages and disadvantages. In this we present literature review about different approaches with their merits and demerits. In future work overcome the drawbacks of the above techniques design an algorithm by using the useful features of two or more approach which helps in detection and tracking the video object more accurately for the particular application.

REFERENCES

[1]. Abhishek Kumar Chauhan, PrashantKrishan, "Moving Object Tracking Using Gaussian Mixture Model And Optical Flow", International Journal of Advanced Research in Computer Science and Software Engineering, April 2013.

[2]. Cheng-Laing Lai; Kai-Wei Lin, "Automatic path modeling by image processing techniques," Machine Learning and Cybernetics (ICMLC), 2010 International Conference on , vol.5, no., pp.2589,2594, 11-14 July 2010

[3]. Sen-Ching S. Cheung and Chandrika Kamath, "Robust techniques for background subtraction in urban traffic video".

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

[5]. P Dokladal, R Enficiaud, E. Dejnozkova , 'Contour-based object tracking with gradient-based contour attraction field', EEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP '04) ,pp. 17-20, 2004.

[6]. T. Chen, 'Object Tracking Based on Active Contour Model by Neural Fuzzy Network', IITA International Conference on Control Automation and Systems Engineering, pp. 570-574, 2009.

[7]. X. Lu, Li Song, Songyu Yu, Nam Ling,' Object Contour Tracking Using Multi-feature Fusion based Particle Filter', IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 237 –242, 2012.

[8]. Kuihe Yang, ZhimingCai, Lingling Zhao "Algorithm Research on Moving Object Detection of Surveillance Video Sequence", Optics and Photonics Journal, 2013, 3, 308-312.

[9]. V. Arunachalam, I. Sorimuthu, V. Raja gopal and B. Sankaragomathi "Automatic Fast Video Object Detection And Tracking On Video Surveillance System", ICTACT Journal On Image And Video Processing, August 2012, Volume: 03, Issue: 01, ISSN: 0976-9102(ONLINE).

[10]. R. Cucchiara, M. Piccardi, and P. Mello, "Image analysis and rule-based reasoning for a traffic monitoring system," IEEE Trans. Intel. Transp. Syst., vol. 1, no. 2, pp. 119–130, Jun.2000.

[11]. J. Joshan Athanesious, P. Suresh, "Systematic Survey on Object Tracking Methods in Video", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) October 2012, 242-247.

[12]. Sen-Ching S. Cheung and Chandrika Kamath, "Robust techniques for background subtraction in urban traffic video".

[13]. Viola, P., Jones, M., And Snow, D. 2003. Detecting pedestrians using patterns of motion and appearance. In IEEE International Conference on Computer Vision (ICCV). Pp.734–741.

[14]. Papageorgiou, C., Oren, M., And Poggio, T. "A general framework for object detection". In IEEE International Conference on Computer Vision (ICCV). Pp.555–562, 1998.

[15]. Hanxuan Yang , Ling Shao, FengZheng , Liang Wangd, Zhan Song, "Recent advances and trends in visual tracking: A review", Elsevier Neuro computing 74 (2011) pp. 3823–3831, 2011

[16]. Greg Welch, Gary Bishop," An introduction to the Kalman Filter", In University of North Carolina at Chapel Hill, Department of Computer Science. Tech. Rep. 95-041, July-2006.

[17]. Joshan Athanesious J; Mr. Suresh P, "Implementation and Comparison of Kernel and Silhouette Based Object Tracking", International Journal of Advanced Research in Computer Engineering & Technology, March 2013, pp. 1298-1303.

[18]. J. J. Du, "Research on Detection and Tracking of Moving Object in Intelligent Video Surveillance System," master's degree thesis of Southwest Jiaotong University, 2009.

[19]. Badri Narayan Subudhi, Pradipta Kumar Nanda, Member, IEEE, and AshishGhosh "A Change Information Based Fast Algorithm for Video Object Detection and Tracking", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 7, July 2011.

[20]. Samuel S. Blackman, "Multiple Hypothesis Tracking for Multiple Target Tracking", IEEE A&E Systems magazine Vol. 19, No. 1,pp. 5-18,January 2004.

[21]. JoshanAthanesious J; Mr. Suresh P, "Implementation and Comparison of Kernel and Silhouette Based Object Tracking", International Journal of Advanced Research in Computer Engineering & Technology, March 2013, pp. 1298-1303.