A Survey of Video Object Tracking Methods

Mukesh Parmar
PG Student Of Computer Engineering,
Swaminarayan College Of Engineering & Technology, Kalol, India-382721

***Abstract* - Object tracking finds its application in several computer vision applications, such as video compression, surveillance, robotics etc. Object tracking is a process of segmenting a region of interest from a video scene and keeping track of its motion, position and occlusion. The tracking is performed by monitoring objects' spatial and temporal changes during a video sequence, including its presence, position, size, shape Difficulties in tracking objects can arise due to abrupt object motion, changing appearance patterns of both the object and the scene, nonrigid object structures, object-to-object and object-to-scene occlusions, and camera motion. Tracking of an object mainly involves two preceding steps object detection and object representation. Object detection is performed to check existence of objects in video and to precisely locate that object. In this survey, we categorize the tracking methods on the basis of the object and motion representations used, provide detailed descriptions of representative methods in each category, and examine their pros and cons.**

***Keywords* - object detection, object representation, object tracking, point tracking, shape tracking.**

I. INTRODUCTION

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and in expensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms.[1] There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behaviour. Therefore, the use of motion-based recognition, that is, human identification based on gait, automatic object detection, automated surveillance, that is, monitoring a scene to detect suspicious activities or unlikely events video indexing, that is, automatic annotation and retrieval of the videos in multimedia databases ,human-computer interaction, that is, gesture recognition, eye gaze tracking for data input to computers ,traffic monitoring, that is, real-time gathering of traffic statistics to direct traffic flow ,vehicle navigation, that is, video-based path planning and obstacle avoidance capabilities.[2]

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object. Tracking objects can be complex due to loss of information caused by projection of the 3D world on a 2D image, noise in images, complex object motion, nonrigid or articulated nature of objects, partial and full object occlusions, complex object shapes, scene illumination changes, and real-time processing requirements.[2]

There are three key steps in video analysis: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behaviour [3].Actually videos are 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 [3].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 complimentary set contains the background pixels. This 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. Many a times shadow is represented as foreground object which gives improper output. The basic steps for tracking an object are described below:

a) Object Detection
   Object Detection is a process to identify objects of interest in the video sequence and to cluster pixels of these objects. Object detection can be done by various techniques such as temporal differencing , frame differencing , Optical flow and Background subtraction [1]**.**
b) Object Representation
   Object representation involves various methods such as Shape-based representation, Motion-based representation, Color based representation and texture based representation where object can be represented as vehicles, birds, floating clouds, swaying tree and other moving objects.[4]
c) Object Tracking
   Tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. Point tracking, kernel tracking and silhouette tracking are the approaches to track the object.[1]

II. OBJECT TRACKING METHODS

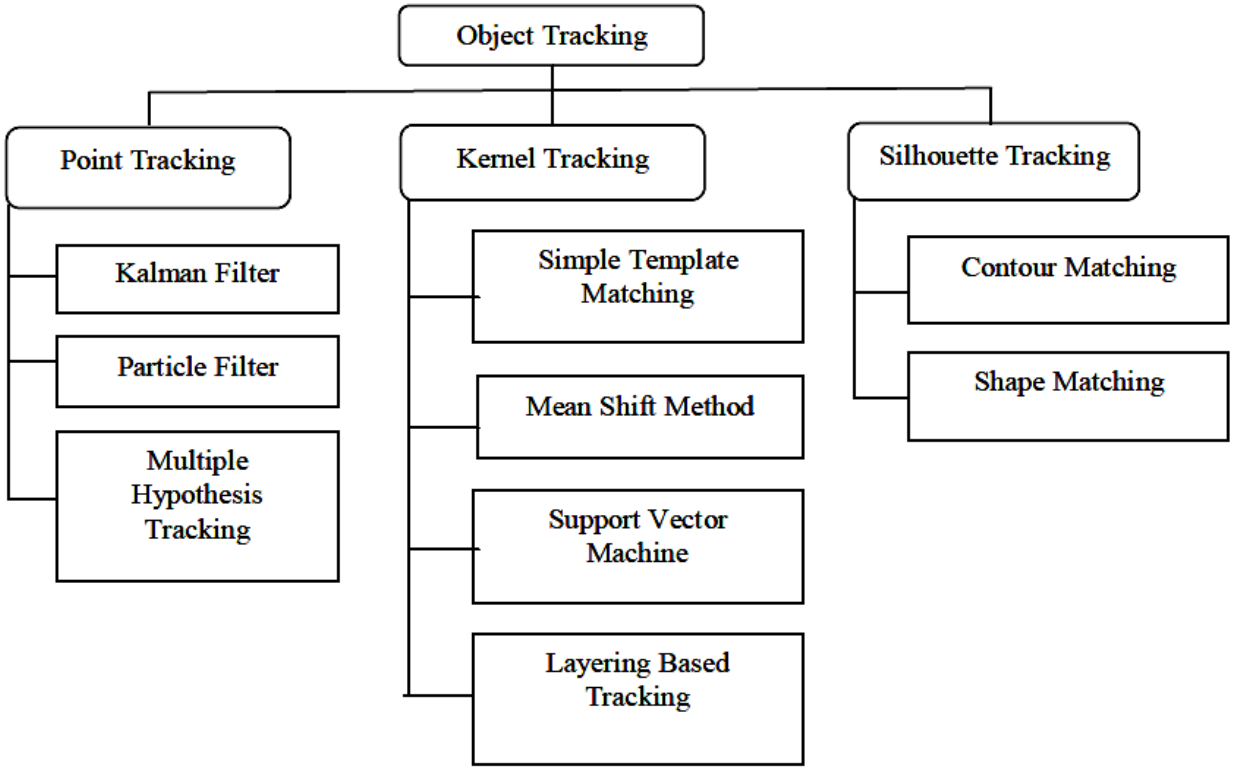

Fig 1.Types of Object Tracking [1]

2.1 Point Tracking

Objects detected in consecutive frames are represented by points,and the 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. An example of object correspondence is shown in figure

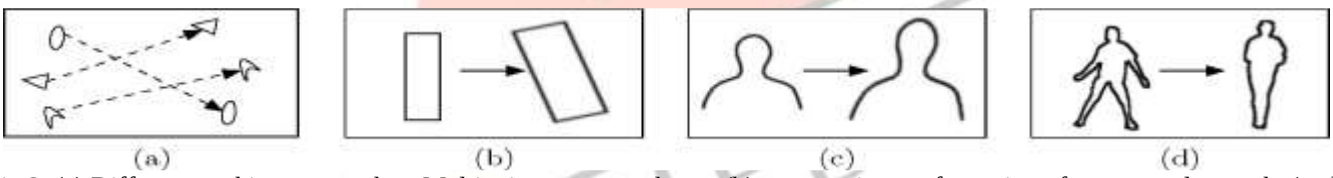

Fig 2. (a) Different tracking approaches. Multipoint correspondence, (b) parametric transformation of a rectangular patch, (c, d) Two examples of contour evolution.[2]

2.1.1 Kalman Filter

A Kalman filter[10] is used to estimate the state of a linear system where the state isassumed to the distributed by a Gaussian. The Kalman filter is a recursive predictive filter that is based on the use of state space techniques and recursive algorithms. It is estimated the state of a dynamic system. This dynamic system can be disturbed by some noise, mostly assumed as white noise. To improve the estimated state the Kalman filter uses measurements that are related to the state but disturbed as well. Kalman filtering is composed of two steps. Thus the Kalman filter consists of two steps:[4]

1. The prediction
2. The correction

In the first step the state is predicted with the dynamic model. The prediction step uses the state model to predict the new state of the variables.

$$X^t = D\,X^{t-1} + W \quad (1)$$

$$\sum^t = D \sum^{t-1} D^t + Q^t \quad (2)$$

Where X^t and Σ^t are the state and covariance predictions at time t. D is the state transition matrix which defines the relation between the state variables at time t and t-1. Q is the covariance of the noise W. Similarly the correction step uses the current observation Z^t to update the object state.

$$K^t = \sum^t M^t [\, M \sum^t M^t + R^t]^{-1} \quad (3)$$

$$X^t = X^t + K^t [Z^t - M\,X^t] \quad (4)$$

where M is the measurement matrix, K is the Kalman gain which is called as the Riccati equation used for propagation of the state models. The updated state *Xt* is distributed by Gaussian. Similarly Kalman filter and extended Kalman filter assumes that the state is distributed by a Gaussian. In the second step it is corrected with the observation model, so that the error covariance of the estimator is minimized. In this sense it is an optimal estimator. Kalman filter has been extensively used in the vision community for tracking .[4] The Kalman filter 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 [3] 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.

2.1.2 Particle Filter

Particle filter is a filtering method based on Monte Carlo and recursive Bayesian estimation. The particle filter, also known as condensation filter and they are suboptimal filters. The core idea is that density distribution is present using random sampling particles. There is no restriction to the state vector to deal with nonlinear and non-Gaussian problem, and it is the most general Bayesian approach.[5] The working mechanism of particle filters is given as follows. The state space is partitioned as many parts, in which the particles are filled according to some probability measure. The higher probability, the denser the particles are concentrated. The particle system evolves along the time according to the state equation, with evolving pdf determined by the FPK equation. Since the pdf can be approximated by the point-mass histogram, by random sampling of the state space, we get a number of particles representing the evolving pdf.

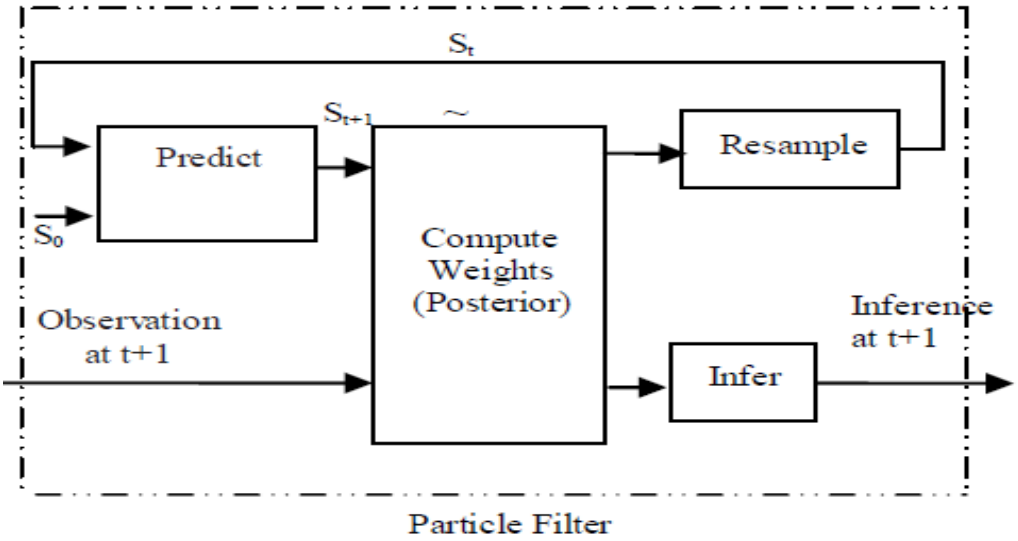

Fig3 Particle Filter

The basic steps block diagram in particle filtering is shown above[6]

The above Fig. 3 represents the particle filtering (PF) scheme. Consider a system whose state is changing in time St = f (St-l , Wt) where St is the system state at time t. The function f is called the state transition model and says that the system is Markovian. That is, St depends on the previous state St-l and the system (process) dynamics Wt, which enable the system to change in time. Also, assume the system is being partially observed using a set of noisy sensors Zt = h(St, Vt). Here h is called the observation model and captures the relationship between the current system state St, the sensor observation Zt, and sensor noise Vt. The randomness associated with Wt and Vt is assumed to be known and captured through pdfs [6].

2.1.3. Multiple Hypothesis Tracking

A key strategy in MHT is to delay data association decisions by keeping multiple hypotheses active until data association ambiguities are resolved. MHT maintains multiple track trees, and each tree represents all of the hypotheses that originate from a single observation . At each frame, the track trees are updated from observations and each track in the tree is scored. The best set of non-conflicting tracks (the best global hypothesis) can then be found by solving a maximum weighted independent set problem. Afterwards, branches that deviate too much from the global hypothesis are pruned from the trees, and the algorithm proceeds to the next frame.[7]

If motion correspondence is recognized using only two frames, there is always a limited chance of an incorrect correspondence. Better tracking outcomes can be acquired if the correspondence choice is overdue until several frames have been observed. The MHT[11] algorithm upholds several correspondences suggestions for each object at each time frame The final track of the object is the most likely set of correspondences over the time period of its observation.[1] MHT is an iterative algorithm. Iteration begins with a set of existing track hypotheses. Each hypothesis is a crew of disconnect 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 dealing with: Tracking multiple object, Ability to tracks for objects entering, exit of Field Of View (FOV). It also handles occlusions, Calculating of Optimal solutions [7].

2.2. Kernel Tracking

Kernel tracking [1] is usually performed by computing the moving object, which is represented by a embryonic object region, from one frame to the next. The object motion is usually in the form of parametric motion such as translation, conformal, affine, etc.These algorithms diverge in terms of the presence representation used, the number of objects tracked, and the method used for approximation the object motion. In real-time, illustration of object using geometric shape is common. But one of the restrictions is that parts of the objects may be left outside of the defined shape while portions of the background may exist inside. This can be detected in rigid and non-rigid objects .They are large tracking techniques based on representation of object, object features, appearance and shape of the object.

2.2.1. Simple Template Matching

Template matching [1] is a brute force method of examining the Region of Interest in the video. In template matching, a reference image is verified with the frame that is separated from the video. Tracking can be done for single object in the video and overlapping of object is done partially. Template Matching is a technique for processing digital images to find small parts of an image that matches, or equivalent model with an image (template) in each frame. The matching procedure contains the image template for all possible positions in the source image and calculates a numerical index that specifies how well the model fits the picture that position. It can capable of dealing with tracking single image and partial occlusion of object.

Template matching [7]is a brute force method of examining the ROI in the ongoing video a simple way of tracking ithreference image. Here in template matching, a reference image is verified with the frame that is separated from the video. It can track only single object in the video. Translation of motion only can be done in template matching. Capable of dealing with: Tracking single image, Partial occlusion of object, Necessity of a physical initialization.

2.2.2 Mean Shift Method

Mean-shift tracking [1]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.

The task is to first define an Region of Interest (ROI) from moving Object by segmentation and then tracking the object from one frame to next. Region of interest is defined by the rectangular window in an initial frame. Tracked object is separated from background by this algorithm. The accuracy y of target representation and localization will be improved by Chamfer distance transform. Minimizing the distance amongst two color distributions using the Bhattacharya coefficient is also done by Chamfer distance transform. In tracking an object, we can characterize it by a discrete distribution of samples

and kernel is localized . Steps for kernel tracking: Probabilistic distribution of target in firs frame is obtained using color feature. Compare the distribution of first frame with consecutive frame. Bhattacharya coefficient is used to find the degree of similarity between the frames. Loop will continue till the last frame [8].Capable of dealing with: Tracking only single object, Object motion by translation and scaling. Necessity of a physical initialization. Object is partial occlusion [7].

2.2.3. Support Vector Machine (SVM)

SVM [1] is a broad classification method which gives a set of positive and negative training values. For SVM,the positive samples contain tracked image object, and the negative samples consist of all remaining things that are not tracked. It can handle single image, partial occlusion of object but necessity of a physical initialization and necessity of training. SVM is a broad classification method which gives a set of positive and negative trainingvalues. For SVM, the positive samples contain tracked image object, and the negative samples consist of all remaining things that are not tracked. During the analysis of SVM, score of test data to the positive class Capable of dealing with:[7] Tracking single image. Partial occlusion of object. Necessity of a physical initialization. Necessity of training. Object motion by translation.

2.2.*4. Layering based tracking*

This is another method of kernel based tracking[1] where multiple objects are tracked. Each layer consists of shape representation (ellipse), motion such as translation and rotation, and layer appearance, based on intensity. Layering is achieved by first compensating the background motion such that the object's motion can be estimated from the rewarded image by means of 2D parametric motion. Every pixel's probability of calculated based on the object's foregoing motion and shape features [8]. It can capable of tracking multiple images and fully occlusion of object.[7] Capable of dealing with: Tracking multiple images. Fully occlusion of object. Object motion by translation, scaling and rotation.

2.3 Silhouette Tracking

Some object will have complex shape such as hand, fingers, shoulders that cannot be well defined by simple geometric shapes. Silhouette based methods [9] afford an accurate shape description for the objects. The aim of a silhouette-based object tracking is to find the object region in every frame by means of an object model generated bythe previous frames. Capable of dealing with variety of object shapes, Occlusion and object split and merge.[1]

2.3.1 Contour Tracking

Contour tracking methods [9], 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. Contour tracking methods[8], in divergence to shape matching methods, iteratively develop an original contour in the foregoing frame to its new position in the present frame, overlapping of object between the current and next frame. Contour tracking is in form of State Space Models. State Space Models: State of the object is named by the parameters of shape and the motion of the contour. The state is updated for each time according to the maximum of probability. In Contour Tracking, explicitly or implicitly are used for the representation on silhouette tracking. Representation based on explicitly will defines the boundaries of silhouette whereas in case of implicitly, function defined by grid.

2.3.2 Shape Matching

These approaches examine for the object model in the existing frame. Shape matching performance is similar to the template based tracking in kernel approach. Another approach to Shape matching [9] is to find matching silhouettes detected in two successive frames. Silhouette matching, can be considered similar to point matching. Detection based on Silhouette is carried out by background subtraction. Models object are in the form of density functions, silhouette boundary, object edges. Capable of dealing with single object and Occlusion handling will be performed in with Hough transform techniques. These approaches examine for the object model in the existing frame. Shape matching performance is similar to the template based tracking in kernel approach. Another approach to Shape matching is to find matching silhouettes detected in two successive frames. Silhouette matching, can be considered similar to point matching which is described. Detection based on Silhouette is carried out by background subtraction. Models object are in the form of density functions, silhouette boundary, object edges.[8] Capable of dealing with:Edge based template, Silhouette tracking feature of shape matching are able to track only single object. Occlusion handling will be performed in with Hough transform techniques.

Table:1 Qualitative Comparison for Tracking methodologies.(#:no of objects tracking, S:single,M:multiple,P:partial,F:full, Symbols √ and × denote whether the tracker can or cannot handle occlusions, and requires or does not require training).

Sr.no.	Method	Category	#	Entry	Exit	Occlusion	Optimal	Training Rule
1	Kalman filter	Point tracking	S	√	√	×	√	-
2	MHT	Point tracking	M	√	√	√	√	-
3	Particle Filter	Point tracking	M	×	×	√	√	-
4	Template matching	Kernel Tracking	S	-	-	P	-	×
5	Mean shift	Kernel Tracking	S	×	×	P	-	×
6	SVM	Kernel Tracking	S	-	-	P	-	√
7	Layering based tracking	Kernel Tracking	M	-	-	F	-	×
8	Shape matching	Silhouette tracking	S	-	-	×	-	×
9	Contour matching	Silhouette tracking	M	-	-	F	√	√

III. **ACKNOWLEDGEMENTS**

I am very grateful and would like to thank my guide and teacher Prof.N R Patel, Prof. Darshana Mistry for their advice and continued support without them it would not have been possible for me to complete this report. I would like to thank all my friends, colleague and classmates for all the thoughtful and mind stimulating discussions.

IV. REFERENCES

[1] Himani S. Parekh, Darshak G. Thakore, Udesang K. Jaliya 2014. A Survey on Object Detection and Tracking Methods, International Journal of Innovative Research in Computer and Communication Engineering 2, p.2970

[2] Alper Yilmaz,Omar Javed,Mubarak Shah 2006. Object Tracking: A Survey, *ACM Comput . Surv. 38*, 4, Article 13 p.45.

[3] Gandham Sindhuja, DR.Renuka Devi S M 2015. A Survey on Detection and Tracking of Objects in Video Sequence, International Journal of Engineering Research and General Science 3,p.418 ISSN 2091-2730

[4] Barga Deori, Dalton Meitei Thounaojam 2014. A survey on moving object tracking in video, International Journal on Information Theory (IJIT), 3 p.31

[5] Md. Zahidul Islam, Chi-Min Oh and Chil-Woo Lee 2009. Video Based Moving Object Tracking by Particle Filter, International Journal of Signal Processing, Image Processing and Pattern . 2, p.119

[6] G.Mallikarjuna Rao 2013. Visual Object Target Tracking Using Particle Filter: A Survey, I.J. Image, Graphics and Signal Processing, 6,p. 57

[7] J.Joshan Athanesious, P.Suresh 2012. Systematic Survey on Object Tracking Methods in Video, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)Volume 1, Issue 8, October 2012

[8] Rahul Mishra, Mahesh K. Chouhan, Dr. Dhiiraj Nitnawwre 2012 .Multiple Object Tracking by Kernel Based Centroid Method for Improve Localization, International Journal of Advanced Research in Computer Science and Software Engineering, p 137.
[9] Mr. Joshan Athanesious J; Mr. Suresh P, March 2013.Implementation and Comparison of Kernel and Silhouette Based Object Tracking, International Journal of Advanced Research in Computer Engineering & Technology, p. 1298
[10] Rupesh Kumar Rout: A Survey on Object Detectionand Tracking Algorithms, Department of Computer Science and EngineeringNational Institute of Technology Rourkela,June 2013
[11] Chanho Kim , Fuxin Li, Arridhana Ciptadi , James M. Rehg : Multiple Hypothesis Tracking Revisited.