A Survey on Object Detection, Classification and Tracking Methods

Kirubaraj Ragland
Post Graduate Scholar, Department of ECE,
Christian College of Engineering and Technology,
Oddanchatram, India.

P. Tharcis
Assistant Professor, Department of ECE,
Christian College of Engineering and Technology,
Oddanchatram, India.

Abstract— **Video tracking is one of the fields of recent development in the field of computer vision. A lot of research has been going on in this area and new algorithms are being proposed to detect and track objects in video. This field of study has experienced a sudden growth especially after robotics has started gaining importance. The objective of this paper is to present the different steps involved in tracking objects in a video sequence, namely object detection, object classification and object tracking. We survey the different methods available for detecting, classifying and tracking objects in a detailed manner. The pros and cons of each of the methods are discussed. Object detection methods are frame differencing, optical flow and background subtraction. Then, objects may be classified based on shape, motion, colour and texture. Tracking methods involve point based tracking, kernel based tracking and silhouette based tracking.**

Keywords— object detection, object classification, object tracking, video processing.

I. Introduction

Tracking objects in a video sequence is an area of constant development which has a wide area of application from surveillance monitoring systems to wildlife monitoring and tracking without the aid of human intervention. This paper explains what the different steps involved in tracking objects, be it humans or wildlife or cars, in a video and what the different methods that are available to perform these steps. Then, different algorithms are studied and compared. We take nine algorithms for our study and analyze the different methods that are used there. The remaining of this paper is arranged as follows: Section II covers an extensive review of literature. Section III gives the basic steps involved in tracking an object. Section IV, V and VI elaborate on the methods for object detection, classification and tracking respectively. Finally we conclude in section VII.

II. Review of Literature

A number of algorithms have been developed over time and the different methods that are used in each of these algorithms is shown as an outline below.

Dong Kwon Park et al.(2000) present a semi-automatic object tracking algorithm in [1]. The two steps involved are intra-frame object extraction and inter-frame object tracking. The human intervention is decreased by using homogeneous region segmentation in intra-frame object extraction while the processing time of the inter-frame object tracking is reduced by the use of 1-D projected motion estimation. An improved flooding method for the conventional water shed algorithm is also proposed.

Yining Deng and B.S. Manjunath (2001) propose a method for unsupervised segmentation in both images and video in [2]. The algorithm, called JSEG, works based on colour-texture regions in image as well as video. The two steps in the proposed algorithm are colour quantization and spatial segmentation. In the first step, colours in the image are quantized to several representative classes which are used to differentiate regions in the image. Then, the pixels are replaced by their corresponding colour class labels. This leaves us with a class map of the image. Using the proposed criterion for "good" segmentation to the class-map by applying it to local windows, we obtain the 'J-image' where the high values correspond to possible boundaries of colour-texture regions and the low values to the interiors. Then, a region growing method is employed to segment the image based on multi-scale J-images. In case of video, an additional region tracking scheme is used along with the above mentioned process to get consistent results even with non-rigid object motion. The limitation of this method is that when a smooth transition of colour (for e.g., from red to orange) occurs, the algorithm oversegments each of the colours. However, if we overcome this problem by checking for smooth transitions, we face a problem when a smooth transition may not indicate one homogeneous region. In case of video on error generated in one frame is carried over to the subsequent frames.

Yaakov Tsaig and Amir Averbuch (2002) present an algorithm for automatic segmentation of moving objects in MPEG-4 videos in [3]. MPEG-4 relies on decomposition of each frame of an image sequence into video object planes (VOPs). Each VOP corresponds to a single moving object in the scene. The basic process is to classify regions as foreground or background based motion information. The segmentation problem is formulated as detection of moving objects over a static background. Motion of camera is compensated by eight parameter perspective motion model. Initially spatial partition is obtained by means of watershed algorithm. Then, Canny's gradient is used to estimate the spatial gradient in the colour space. Then, the optimized rainfall watershed algorithm is applied. Based on the initial partitioning, regions are classified as either foreground or background. The motion of each of the foreground regions is

estimated by region matching in hierarchical frame work. The estimated motion vectors are incorporated into a Markov random field model, which is optimized using highest confidence first (HCF), leading to classification of the regions initially. MRF includes information from previous frame. The final step includes a dynamic memory to ensure temporal coherency of the segmentation process.

R. Venkatesh Babu et al. (2004) present an approach for automatically estimating the number of objects and extracting independently moving video objects from MPEG-4 videos using motion vectors in [4]. Since motion vectors are sparse (i.e. one motion vector per macro-block)in compressed MPEG videos, a method to enrich the motion information from a few frames on either side of the current frame is proposed. Interpolation is performed using median filter so that a motion vector is assigned to each pixel in the frame. Then, segmentation is done. Since sufficient data is not available to estimate motion parameters, Expectation Maximization (EM) algorithm is used. An algorithm for estimating the number of motion models is proposed. Once initially segmented, Video Object Planes (VOPs) are generated by tracking. Finally, the VOs are subject to edge refinement phase, where the pixels at the edges are assigned to the correct VO.

The aim of Vasileios Mezaris et al. (2004) is to segment a video sequence to objects in [5]. There are three stages in this algorithm: Initial segmentation of the first frame using colour, motion and position information, a temporal tracking and finally a region-merging procedure. The segmentation step uses K-means-with-connectivity-constraint algorithm. Tracking is done by means of Bayes classifier. A rule-based processing is done to reassign changed pixels to existing regions and also how new regions introduced in the sequence are handled. Region merging is done on a trajectory-based approach rather than a motion at a frame level. One advantage is, it can track new objects appearing on the scene or fast moving objects effectively.

Weiming Hu et al. (2012) propose an incremental Log-Euclidean Riemannian subspace learning algorithm in [6]. The co-variance matrices of image features are first mapped into a vector space using log-Euclidean Riemannian metric. Both global and local spatial layout information are captured by a log-Euclidean block-division appearance model. Bayesian state inference based on particle filtering are used for single object as well as multi-object tracking with occlusion reasoning. Changes in object appearance are captured by incrementally updating the log-Euclidean block division appearance model.

In [7], Chang Huang et al. (2013) proposes a method in which the input is given as a frame-by-frame target detection results. The first set of target tracklets (tracking fragments) is generated by conservative dual-threshold strategy. So, only reliable detection responses are linked. Doubtful associations are postponed until more evidence is collected. Multiple passes are used by hierarchical association to achieve Maximum A Posteriori (MAP) problem, which hypotheses a trajectory of being a false alarm besides initializing, tracking and terminating them. Hungarian algorithm is used to solve this issue. Next ranking these associations is seen as a bag-ranking problem, which is overcome by a bag-ranking boosting algorithm. Finally, this paper introduces a soft

max/min to facilitate the optimization of the released objective loss function.

Rana Farah et al. (2013) propose a robust tracking method to extract a rodent from a frame under uncontrolled normal laboratory conditions in [8]. It works in two steps: First, three weak features are combined to roughly track the target. Then, the boundaries of the tracker are adjusted to extract the rodent. The newly introduced techniques include Overlapped Histograms of Intensity (OHI) and a new segmentation method which uses an online edge background subtraction and edglet-based constructed pulses. Edglets are discontinuous pieces of edges. A sliding window technique is used to coarsely localise the target.

Shao-Yi Chien et al. (2013) has two major contributions in [9]: First, a threshold decision algorithm for video object segmentation with multi-background model is proposed. Then, a video object tracking framework based on particle filter with likelihood function is composed of diffusion distance measuring colour histogram similarity and motion clue from video object segmentation. This framework can handle drastic changes in illumination, background clutter and it can also track non-rigid moving objects. The threshold decision algorithm determines an appropriate optimal threshold value for segmentation. Colour based histogram is included for better tracking of non-rigid objects. A 1-D colour histogram is used instead of a 3-D colour histogram so that computational complexity can be reduced.

III. BASIC STEPS IN OBJECT TRACKING

The first part of this paper simply expounds on the different steps involved in tracking an object or multiple objects in a video sequence. The tracking process is preceded by two steps which play a vital role in improving the accuracy of the tracking, namely, object detection and object classification. Though the difference between the three is very subtle and may be missed at first glance, it is very crucial to understand it because all three are different and each one is an area of study by itself. The simple flow diagram is as shown in Fig.1. The first step in the process is to detect what objects are present in the video frame. Then, to classify these objects depending on what we want to track. Finally the actual tracking takes place. The three steps are defined below and the different techniques used are also listed under each type.

Fig. 1 Flow Diagram of Basic steps in Object Tracking

A. Object Detection

Object detection is a computer technology that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos [18]. Object detection may be done by using some basic techniques like frame differencing, optical flow and background subtraction.

B. Object Classification

Object classification is the process by which the objects that are detected in the frame are classified as what object of interest it is. It is essentially just identifying what object it is. Object identification may be done based on different parameters like shape, motion, colour and texture. So, based on the parameter used, we can perform shape-based classification, motion-based classification, colour-based classification or texture based classification.

C. Object Tracking

Object tracking is a method of following an object through successive image frames to determine how it is moving relative to other objects. This is most commonly done by measuring the position of the centroid of the object in (x, y) in successive frames [19]. Object tracking may be classified as point-based tracking, kernel-based tracking or silhouette based tracking.

IV. METHODS OF OBJECT DETECTION

Object detection is the first step in tracking an object in the video. What actually happens in detection is that "objects" are actually groups of pixels clustered together. The basic idea is to identify these pixel clusters as objects which are not only moving in x and y directions but also in time. Let us examine each of the methods of object detection in detail in the following sections.

A. Frame Differencing

Frame differencing is basically used to detect moving objects on a static background. Because objects are moving with difference to time, the position of the object on one frame is different from the position of the object on the consecutive frame. By finding the difference between the two frames, we can get the exact position of the object on the frame. The computational complexity is very low, but complete outline of the moving object is difficult to obtain, which also leads to lower accuracy.
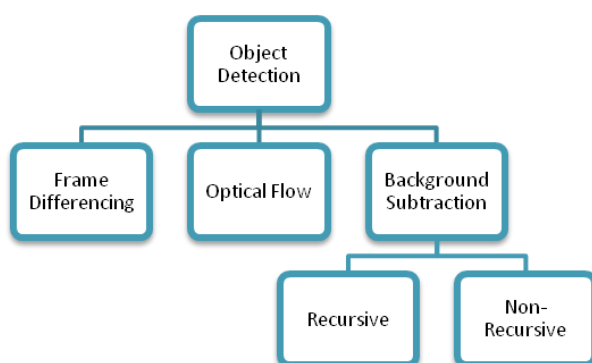

Fig. 2. Methods of Object Detection

B. Optical Flow

Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene [18].Optical flow calculates a velocity for points within the images, and provides an estimation of where points could be in the next image sequence [20]. Optic flow is a vast area of study and [21] provides a summary of the different methods available for estimation of optical flow. Though this method can get better accuracy, it is computationally very costly and its ability to deal with noise is limited.

C. Background Subtraction

A video sequence may be separated into background and foreground. Foreground usually consists of the objects of interest whereas the background data is not important for tracking. If we remove the background data from the video frame, then we are left with just the necessary data in the foreground, which contains the object of interest.

We could get a better accuracy if we already know what the background is. For example, in stationary surveillance cameras as in road traffic monitoring, the background is always constant. The road remains in the same position with respect to the camera. This gives us the advantage of having the background already "modelled" for us. If this is not the case, then background modelling has to be performed before background subtraction. The objective background modelling is to generate a reference model. The video sequence is compared with the reference model and the object is detected by computing the variation between the two. Fig.3. shows how a background is modelled.   There are two types of algorithms for background subtraction. They are recursive algorithm and non-recursive algorithm.

1) Non - recursive algorithm: A non-recursive technique uses a sliding-window approach for background estimation. Select number of previous video frames are stored in a buffer and the background image is estimated based on the temporal variation of each of the pixels within the buffer. Since only select numbers of frames are stored in the buffer, errors caused by frames outside the buffer limit are not taken into consideration. Storage requirements for non-recursive techniques may be very large because of large buffer requirements. This problem is overcome by storing the video at lower frame rates. Some commonly used non-recursive techniques include frame differencing, median filtering, linear predictive filtering and non-parametric modelling [10].
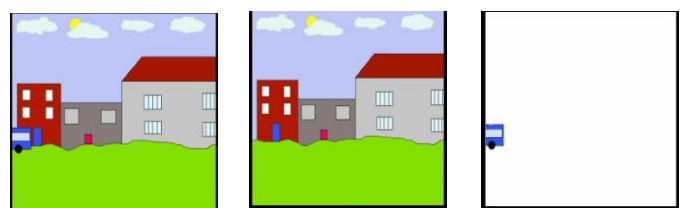

Fig. 3. Background modelling (a): Input video frame, (b): Background model, (c): Object detected after background subtraction.

2) Recursive algorithm: No buffer is used in the case of a recursive technique. A single background model is updated based on each input frame. This means that even frames from the distant past could cause an error in the current model. This also reduces the storage space, as no memory would be necessary to buffer the data. An error caused can linger for a long time. Some recursive techniques include approximated median filtering, Kalman filtering and Mixture of Gaussians (MoG).

V. METHODS OF OBJECT CLASSIFICATION

After the objects have been detected in a video sequence, the next step would be to identify these objects and classify them according to our requirement. Classification is done based on the parameter we select. Depending on what parameter we select for classification, the methods are defined as follows: shape-based classification, motion-based classification, colour-based classification and texture-based classification. Each of the methods is explained below.

A. Shape based classification

Shape based classification is done based on shape analysis. Shape analysis is the automatic analysis of geometric shapes by a computer to detect similarly shaped objects by comparing against entries on a database. Mostly boundary based representation is used. However, volume based representation or point based representation of shapes is also possible. The simplified representation is called shape descriptor. A complete shape descriptor consists of all the information required to reconstruct the shape. Shape descriptors may be invariant with respect to congruency, isometry (intrinsic shape descriptors).Graph based descriptors are another class [18]. Shapes may also be classified based on part structure. But capturing part structure is not a trivial task considering the non-linearity of shapes. Part structure capturing can basically be classified into three categories. The first one builds part models from sample images. This requires some prior knowledge of the part. The next two categories capture part structures from only one image. In the second category, individual parts are compared with each other and a similarity measure is obtained. In the third category, the part structure is captured considering the interior of shape boundaries [16].

B. Motion based classification

Motion based classification works on periodicity of the motion. A system can be made to learn how the object moves and then classify it better. Motion based classification has to be addressed for both rigid and non-rigid objects. Though it is easier to track objects which are rigid and show periodicity in motion, a limited amount of periodicity has been known to exist in non-rigid objects as well. Optical flow is also used for motion-based classification [10].

Fig.4. Methods of Object Classification

C. Colour-based classification

Colour-based approach is based on studying the colour features in an image. The two main colour features that are used to classify based on colour are spectral power distribution of the illuminant and the object's surface

reflectance property. Colour information is usually represented in the most commonly used RGB colour space. The problem with RGB is that it is not a uniform colour spaceTherefore, we have to consider the use of other colour spaces like L*a*b and L*u*v which are perceptually uniform. HSV (Hue, Saturation, and Value) is a relatively uniform colour space.

In all the above mentioned spaces, features to define an object are not efficient. Therefore in the recent years, colour descriptors have been classified into histogram based colour descriptors and SIFT (Scale Invariant Feature Transform) based colour descriptors [12].

D. Texture-based classification

Texture is an innate property of virtually all surfaces, the grain of wood, the weave of fabric, the pattern of crop in fields, etc. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Since the textural properties of images appear to carry useful information, for discriminating purpose features have always been calculated for textures [17].

Although it is quite easy for a human observer to recognize and describe in empirical terms, texture has been extremely adverse to precise definition and analysis by computer. Texture is represented by means of texture descriptors. They observe region homogeneity and histograms of region borders. Different texture descriptors include homogeneous texture descriptor (HTD), texture browsing descriptor (TBD) and edge histogram descriptor (EHD) [18].

VI. METHODS OF OBJECT TRACKING

Having detected the objects and classified them, the next step would be the actual tracking process. According to [21], tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. There are three methods of object tracking which are discussed in detail below. They include point tracking, kernel tracking and silhouette tracking [18].

A. Point Tracking

In the point tracking approach objects are represented as points and are generally tracked across frames by evolving their state (object position and motion). Point tracking may be Kalman filtering, particle filtering or Multiple Hypothesis Tracking (MHT).

1) Kalman Filtering: Kalman filtering is an algorithm that uses a series of measurements observed over time, containing noise (random variations) and other inaccuracies, and produces estimates of unknown variables that tend to be more precise than those based on a single measurement alone. More formally, the Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system state [14].

There are two steps in the algorithm. The prediction step produces estimates of the current state variables along with their uncertainties. Then, the outcome of the next measurement
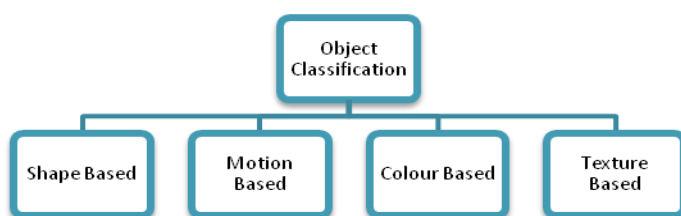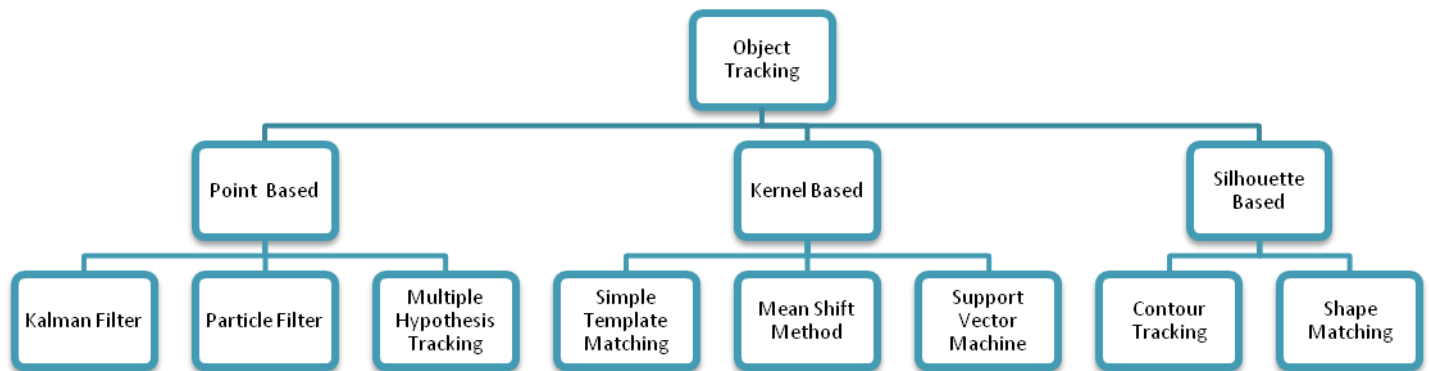
Fig. 5. Methods of Object Tracking

is observed and the estimates are updated using weighted average, with more weight being given to estimates with higher certainty. Since it is a recursive algorithm, only the present value, previous value and uncertainty matrix are enough to calculate in real time.

2) Particle Filtering: Particle filters are a set of on-line posterior density estimation algorithms that estimate the posterior density of the state-space by directly implementing the Bayesian recursion equations. Posterior density is represented by a set of particles. No assumptions about the dynamics of the state-space or the density function are made, but they provide a well-established method for generating samples from the required distribution. The samples are represented by a set of particles and each particle has a weight assigned to it. The weight of each particle represents the probability that that particle is being sampled from the probability density function. Resampling is done so as to avoid weight disparity which leads to weight collapse. When the state varibles are not distributed normally (Gaussian), Kalman filter provides a poor approximation. Particle filters are used to overcome this problem. This algorithm uses contours, colour features or texture mapping [18].

3) Multiple Hypothesis Tracking(MHT): The MHT algorithm begins with the set of hypothesis of the previous iteration also called parent hypothesis set and the set of measurements from the beginning until that iteration. Each hypothesis represents a different set of assignments of the set of measurements to the different tracks. Taking into account the new set of measurements and one of the previous hypotheses, a new hypothesis is generated, making a specific assignment of the current measurements. The set of plausible assignments that can be done for a parent hypothesis is named ambiguity matrix (sometimes also called hypothesis matrix). Each element of the matrix, a_{ij}, can take a value of 1 or 0, representing the possibility that measurement i is associated to a previous track, a new track, is considered noise, etc., or not. Associated to the ambiguity matrix, a cost

4) matrix must also be defined. Each element of the matrix, c_{ij}, represents the probability that measurement i has been originated due to j. MTH is capable of tracking multiple targets, handling occlusions and calculating optimal solutions [15].

B. ernel Based Tracking

In kernel based approach, an object is tracked based on computing the motion of the rectangular or elliptical kernel in consecutive frames. Motion may include translation, rotation and affine transformations.

The problem with this method is that a part of the object may be left outside the kernel while a part of the background, which is not necessary, may be added into the kernel. Tracking may be based on geometric shape of the object, object features and appearance. Different kernel based tracking approaches include simple template matching, mean shift method, support vector machine (SVM) and layering based technique. These are explained below.

1) Simple Template Matching: Template matching is a technique used in digital image processing for finding small parts of an image or video that match a template image. Since there is no specific way of right and wrong, it is termed to be a brute force method of examining regions of interest in a video. The matching procedure calculates a numerical index for how well the image in the frame matches with the image in the template.

Motion based classification works on periodicity of the motion. A system can be made to learn how the object moves and then classify it better. Motion based classification has to be addressed for both rigid and non-rigid objects. Though it is easier to track objects which are rigid and show periodicity in motion, a limited amount of periodicity has been known to exist in non-rigid objects as well. Optical flow is also used for motion-based classification.

2) Mean Shift Method: Mean shift algorithm is a non-parametric feature space analysis technique for locating the maxima of a density function [18]. It is an iterative algorithm and is based on predicting the future values based on the past values. One of the simplest forms would be to calculate the confidence map of the new image based on colour histogram

of the previous images. The peak of the confidence map will occur at the next predicted position of the object. Thus, the confidence map is a probability density function and the peak of the function may be calculated using mean shift method [15].

The region of interest (ROI) is selected by a bounding box from the first frame of the video. The probability density function is calculated based on colour information. Then, the present probability density function is compared with the probability density functions of the consecutive frames. Degree of similarity between the frames is represented by using Battacharya coeffecient. The same procedure will be followed till the last frame of the video is reached [11].

This method has a lot of drawbacks. Only one object can be tracked. The ROI has to be initialized manually. It cannot track an object if the object is moving with high speed within the frame.

3) Support Vector Machine (SVM): Support Vector Machine uses algorithms which are supervised learning models and their associated learning algorithms which analyze data and recognize patterns based on previously learnt data. Support vector machines are used mostly for classification and regression [18].

The basic idea behind a SVM algorithm is that it classifies the points into two sets of hyperplanes. The objects that are tracked are classified as one set and the objects that don't have to be tracked as another set. It can track only a single object and cannot handle partial occlusions. The algorithm needs to be initialized and also takes more computational overhead as it involves learning.

4) Layer based tracking: Layer based tracking is a multiple object tracking method. Here, each layer corresponds to a particular shape and based on that, the particular object is tracked in that layer. Since many such layers can be present, this facilitates multiple tracking of objects simultaneously.

Layering is achieved by features like motion or colour. Each layer has three features which define what to track in that particular layer. Each of the layers have a shape representation, a motion representation and a layer appearance, based on intensity.

This method is capable of tracking multiple objects at the same time. It can also able to handle tracking of fully occluded objects on the scene [11].

C. Silhouette Tracking

Silhouette based tracking is done when we have to track complex shapes that cannot be represented by simple geometric shapes. Complex shapes include fingers, hands, shoulders etc. The aim of silhouette based tracking is to generate a object model from the previous frame which is compared with the object region in the next frame and thus achieve tracking.

1) Contour tracking: Contour tracking is a method in which the contour of the object is taken from the previous frame and it iteratively proceeds to calculate the contour of the next frame. The requirement of this method is that a certain amount of contour from the last frame overlap with

the contour of the next frame. Only if this requirement is satisfied, proper tracking results can be achieved.

There are two ways to perform contour tracking. In the first approach, the contour shape and motion are modeled using state space models. The second method is more direct and direct minimization techniques such as gradient descent for minimizing the contour energy, thereby evolving the contour. The most significant advantage is that this can be used to track objects of irregular shapes.

2) Shape matching: Shape matching is similar to the shape-based classification technique. But, instead of classifying the object, we use shapes to track the particular shape. It is also similar to template matching, because the shapes are stored in a database and then the shape on the frame is compared with the shape in the database and thus tracking is done. Silhouettes from two successive frames can also be matched to obtain the required results. A single object can be tracked and occlusions are handled by means of using Hough transform [11].

VII. CONCLUSION

We have surveyed the various steps involved in tracking an object from a video sequence and also the different methods to perform object detection, object classification and object tracking were discussed. The advantages and disadvantages of each of these methods were discussed. In the future, we plan to propose an algorithm that overcomes the disadvantages of the existing object detection methods and track objects in a video with capability to handle multiple objects and occlusions.

REFERENCES

[1] Dong Kwon Park, Ho Seok Yoon and Chee Sun Won, "Fast Object Tracking In Digital Video," *IEEE Trans. Consumer Electronics*, vol. 46, no. 3, pp. 785-790, Aug. 2000.

[2] Yining Deng, Member, IEEE, and B.S. Manjunath, Member, IEEE, "Unsupervised Segmentation of Color-Texture Regions in Images and Video," *IEEE Trans. Pattern Analysis And Machine Intelligence,* vol. 23, no. 8, pp. 800-810, Aug. 2001.

[3] Yaakov Tsaig and Amir Averbuch, "Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach*," IEEE Trans. Circuits And Systems For Video Technology,* vol. 12, no. 7, pp. 597-612, Jul. 2002.

[4] R. Venkatesh Babu, K. R. Ramakrishnan, Member, IEEE, and S. H. Srinivasan, "Video Object Segmentation: A Compressed Domain Approach," *IEEE Trans. Circuits And Systems For Video Technology,* vol. 14, no. 4, pp. 462-474, Apr. 2004.

[5] Vasileios Mezaris, Student Member, IEEE, Ioannis Kompatsiaris, Member, IEEE, and Michael G. Strintzis, Fellow, IEEE, "Video Object Segmentation Using Bayes-Based Temporal Tracking and Trajectory-Based Region Merging," *IEEE Trans. Circuits And Systems For Video Technology,* vol. 14, no. 6, pp. 782-795 Jun. 2004.

[6] Weiming Hu, Xi Li, Wenhan Luo, Xiaoqin Zhang, Stephen Maybank, and Zhongfei Zhang, "Single and Multiple Object Tracking Using Log-Euclidean Riemannian Subspace and Block-Division Appearance Model," *IEEE Trans. Pattern analysis and machine intelligence,* vol. 34, no. 12, pp. 2420-2440, Dec. 2012.

[7] Chang Huang, Yuan Li, and Ramakant NevatiaFellow, IEEE, "Multiple Target Tracking by Learning-Based Hierarchical Association of Detection Responses," *IEEE Trans. Pattern analysis and machine intelligence,* vol. 35, no. 4, pp. 898-910, Apr. 2013.

[8] Rana Farah, Student Member, IEEE, J. M. Pierre Langlois, Member, IEEE, and Guillaume-Alexandre Bilodeau, Member, IEEE, "Catching a Rat by Its Edglets," *IEEE Trans. Image Process.,* vol. 22, no. 2, pp. 668–678, Feb. 2013.

[9] Shao-Yi Chien, *Member, IEEE,* Wei-Kai Chan, Yu-Hsiang Tseng, and Hong-Yuh Chen, "Video Object Segmentation and Tracking Framework With Improved Threshold Decision and Diffusion Distance," *IEEE Trans. Circuits and systems for video technology,* vol. 23, no. 6, pp. 921-934, Jun. 2013.

[10] K.Srinivasan, K.Porkumaran, G.Sainarayanan, "Improved Background Subtraction Techniques For Security In Video Applications"

[11] Himani S. Parekh, Darshak G. Thakore , Udesang K. Jaliya, "A Survey on Object Detection and Tracking Methods," *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2, Issue 2, pp. 2970-2978, Feb. 2014.

[12] Barga Deori and Dalton Meitei Thounaojam "A Survey On Moving Object Tracking In Video*," International Journal on Information Theory (IJIT),* vol.3, no.3, pp. 31-46 Jul. 2014.

[13] R.J.Bhiwani, M.A.Khan, S.M.Agrawal, " Texture Based Pattern Classification," *International Journal of Computer Applications,* vol. 1, no. 1, pp. 54-56, 2010.

[14] Ma¨el Primet and Lionel Moisan, "Point tracking: an a-contrario approach."

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

[16] Haibin Ling David W. Jacobs, "Shape Classification Using the Inner-Distance."

[17] Ritika and Gianetan Singh Sekhon, "Moving Object Analysis Techniques In Videos - A Review," *IOSR Journal of Computer Engineering,* vol. 1, Issue 2, pp. 07-12, May-June 2012.

[18] http://en.wikipedia.org/

[19] http://www.sstgroup.co.uk/go/products/video-analytics/video-analytics-glossary

[20] http://users.ecs.soton.ac.uk/msn/book/new_demo/opticalFlow/

[21] http://www.scholarpedia.org/article/Optic_flow