Scientific Journal of Impact Factor (SJIF): 4.72
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 10, October -2017

A Review of Object Detection and Tracking Methods

Dr. Rachna Verma[1]

[1]Department of Computer Science and Engineering, Faculty of Engineering, JNV University, Jodhpur, Rajasthan, India

Abstract: *Moving object tracking in videos is an actively researched area for the last two decades due to its practical applications in many areas, such as event analysis, human computer interaction, crowd analysis, etc. Extensive research reported some success applications in highly constrained domains, but there are still many challenges that need to be resolved, such as abrupt object motion, changes in appearance of the object, non-rigid objects, occlusion, illumination, etc. This paper presents a comprehensive review of various object tracking approaches reported in literature and proposes a new categorization to group various tracking approaches to streamline future researches. It also discusses, in detail, methods used in each category for tracking single and multiple objects. Finally, the paper concludes by setting directions for further research.*

Keywords: *Object detection, object tracking, particle filter, Kalman filter, background subtraction, optical flow*

I. INTRODUCTION

Now a days, many researchers are actively involved in the development of computer vision systems that try to simulate the basic abilities of biological systems, such as the abilities to understand scenes, detect objects(static or moving), understand surrounding, recognize events, analyze crowd, count people, detect people and vehicles detection, etc. Object detection refers to finding an object of some interest in a scene, for example detecting people, vehicles, etc. in a scene. Object tracking refers to estimate the trajectory of a moving object in a scene, for example, tracking the trajectory of a moving car to find lane violation. For tracking objects, videos are generated either from static cameras, such as surveillance cameras, or moving cameras, such as cameras mounted on a mobile robot. In a static camera, the background is always static and objects move, while in a moving camera, objects move in a dynamic background.

In order to design a robust visual tracker, there are some fundamental problems, such as abrupt object motion, changes in appearance of the object, occlusion, illumination, non-rigid objects and real time processing requirements, which need to be resolved. The appearance of the object often varies during tracking, for example appearance changes due to articulation, rotation or geometrical transformations of the object. Therefore, to achieve the robustness in tracking, adaptability to appearance changes is necessary.

II. OBJECT TRACKING FRAMEWORK

A typical object tracking framework, as shown in Figure 1, generally consists of three modules: Object Detection, Object Modeling and Object Tracking. They interact with each other during a tracking process. These are discussed in detail in the following sections.

III. OBJECT DETECTION

Object detection, a pre-requisite for initializing a tracking process, refers to locate the object of interest in every frame of a video sequence. There are generally two approaches of object detection strategies commonly used to initialize a tracking process: (1) manually locating the object in the first frame and let the system detects features, such as corners, to track the object in the next frame and (2) automatic detection of the object using predefined features, such as color.
There are many techniques to detect moving objects: Background subtraction, Kalman filter, particle filter.

Background Subtraction [1-3] is widely used in video sequences having static background. The method segments the image into foreground and background. The foreground contains moving objects such as moving people, cars while the background contains static objects, like road, building, trees, stationary cars, etc.

In this technique, a reference background image is first captured when the objects of interest are not present in the scene. The moving object is detected by subtracting the current image frame from the reference background image. The resulting difference image has values below a predefined threshold in the background area of the current image except the area occupied by the object. The pixels where the difference is above the threshold are classified as foreground. As, in practice, the background of any scene gradually changes with time, the reference background image should be updated from time to time to avoid false detection of objects.

Temporal differencing [4] is a method most suitable for scenarios where the camera is moving. It detects objects by taking differences of consecutive frames (two or three), pixel by pixel. In a moving camera situation, the motion of the camera and the motion of the object are mixed up. Therefore, some researchers [2] proposed to estimated and adjust camera motion first and then apply the background subtraction method. The temporal differencing method fails to detect the overlapping areas the moving objects and wrongly detects the trailing region of the object, known as ghost region, for a fast moving object.

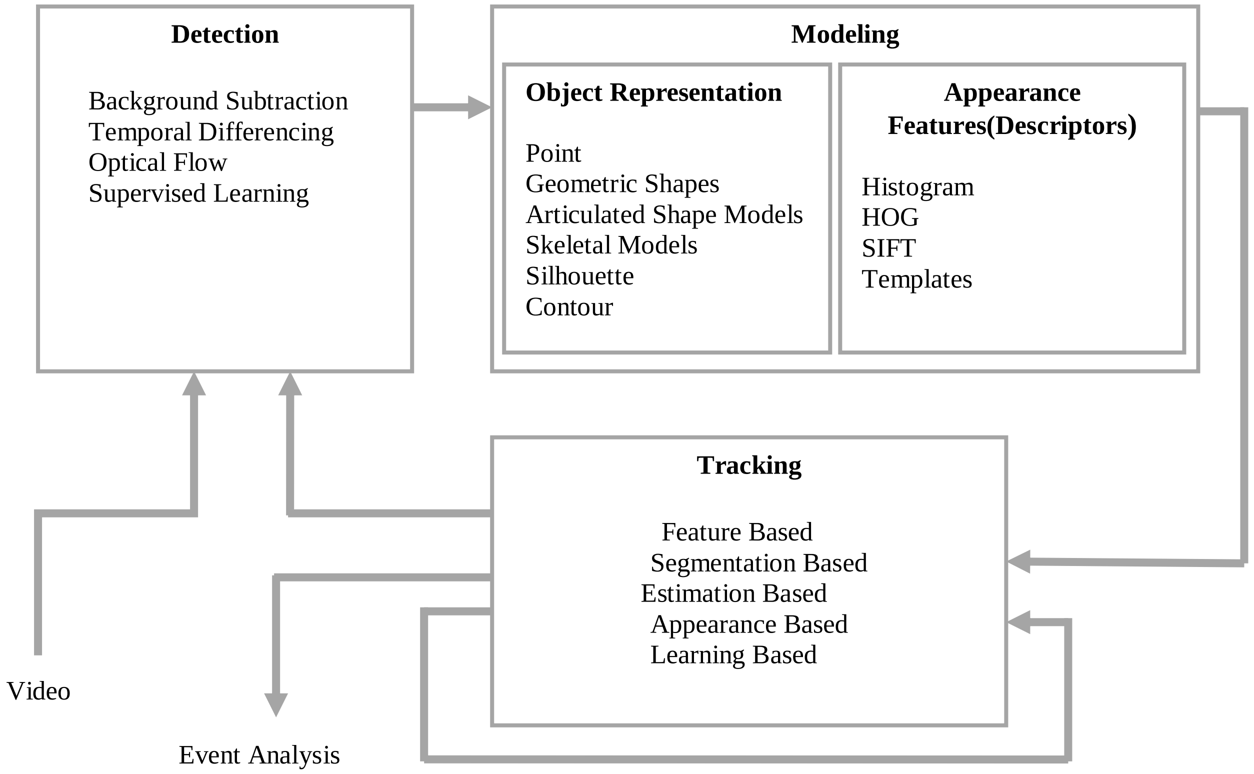

Figure 1: A typical object detection and tracking framework

Optical flow [5, 6] is another technique to find moving objects in video frames. It gives a two-dimensional vector field, also called motion field, that represents velocities and directions of each point in consecutive image sequences [5]. Discontinuities in the optical flow help in segmenting images into regions that correspond to different objects. The method, being computationally expensive, has an advantage that it can detect motion in video sequences having dynamic background.

Object detection can be done by training a classifier that learns different object views and appearances by means of supervised learning methods. After a classifier is trained, the decision is made on the test region whether it is a target object or not. Various learning methods are described later in the paper.

IV. OBJECT MODELING

Object modeling represents the object of interest in a scene. To represent an object, features are extracted that uniquely defines an object. These features or the descriptors of an object are then used to track the object. A feature is an image pattern that differentiates an object from its neighbourhood. The features of an object are converted into descriptors, also referred as appearance features, using some operations around the features.

4.1 Object representation

Object representation refers to how an object is modeled and represented for tracking. In general, the choice of an object representation depends on an application domain. The commonly used object representations for tracking are: Centroid, multiple points, rectangular patch, elliptical patch, part-based multiple patches, object skeleton, complete object contour, control points on the object contour and object silhouette.

Point representation – In the point representation [7-9], an object is represented by either a point, i.e. the centroid pixel of an object, or by a set of points on an object. Points are obtained using some statistics on the object of interest. The main advantage of using the point representation is its efficient processing and manipulation.

Primitive geometric shapes - The object is represented by primitive shapes like rectangle, square, ellipse, circle, etc. Rectangle is the most widely used shape for tracking people [10] and vehicles [11].

Articulated shape models - Articulated objects, like human body, are composed of body parts that are held together with joints. The body parts consist of primitive geometric shapes, such as rectangles, cylinders and ellipses. The relationship between the parts is governed by kinematic motion models, for example, joint angle, etc. Part-based model for pedestrian detection, based on the detection of individual body parts or limbs, is well described by Andriluka et al.[12].

Skeletal models - Skeleton representation is used to model both articulated and rigid objects. The skeleton is defined as a set of articulations within an object that describes the dependencies and defines constraints between the representations of the parts [17]. In [12] the authors utilized the skeletal model for detecting significant number of people.

Object silhouette - Silhouette [13], also called Blobs, is the region inside the contouring boundary of the object. A blob is binary mask that represents an object of interest, the object regions are represented by ones and the non object regions are by zeros.

Contour - Contour [14] defines the boundary of an object. It only describes the edges enclosing the object, a representation suitable for non-rigid objects. For contour based methods, the silhouette is represented either explicitly or implicitly.

4.2 Appearance features (descriptors)

There are a number of ways to represent the appearance features, also called descriptors, of an object, such as probability densities of object appearance and template.

The color histogram [9,15] is the most popular approach for probability density estimates of the object appearance. The major limitation of using histograms is that two objects that have very similar color histogram may have totally different appearance.

Templates[16], suitable only for tracking objects whose poses do not vary considerably during the course of tracking, are formed using simple geometric shapes or silhouettes. It is an ordered list of appearance observations inside its region and carries both spatial and appearance information. Templates, however, only encode the object appearance generated from a single view.

V. OBJECT TRACKING

Object tracking refers to estimating the trajectory of a moving object in a video. For tracking, it is necessary to find the positions of the object in two consecutive frames to generate the trajectory and this process is repeated for the entire video sequence. Sometimes, the object is occluded by some other object in a few numbers of frames and then appears again. Therefore, the tracker must be robust enough to handle partial or full occlusion.

There are number of issues, such as abrupt object motion, noise in the image sequences, changes in appearance of the object and the scene, illumination, etc., that need to be handled while designing a robust visual tracker. In general, a robust tracker should perform the following:

1. It must detect all the objects that entered or moved in the scene. For example a parked car which remains static in many frames, suddenly starts moving must be tracked.
2. It must differentiate between multiple objects. For example, in a football match, two or more person crossing over each other. To track all such objects unique labels must be assigned and maintained throughout the tracking.
3. It must handle the general problems as discussed above, like occlusion, appearance change, non rigid objects, etc.

A number of literature surveys [17-21] have been published on the state-of-the-art methods, their limitations and applications for object tracking. Yilmaz et al. [16] reviewed the object tracking methods, presenting a classification, detailed analysis and comparisons of various tracking methods before 2006. Yang et al. [22] presented the review on tracking methods, their classification and future trends upto 2011. Porkli and Yilmaz[23] wrote a book chapter highlighting various state-of-the-art tracking methods in detail. Li et. al.[21] intensively reviewed various methods till 2013, focusing on the detailed analysis of 2D appearance modeling for visual object tracking. Review of various leading methods upto 2016 classified by different thinking and technologies is presented in [18]. Review of various detection and tracking techniques highlighting vehicles as objects is presented in [19-20].

Approaches to simplify visual tracking have been proposed in the literature. This section presents a review focusing on the methodologies used for tracking objects and proposes a classification based on the approaches used for object tracking, as given in Figure 2. This classification is mainly inspired from Yang et. al.[22].

5.1 Feature based methods

Image features capture different characteristics of the object. To track objects, features, such as texture, gradient, colour, etc., are extracted first. These extracted features must be unique so that objects can be easily distinguishable in the feature space. Once the features are extracted, then the next step is to find the most similar object in the next frame, using those features, by exploiting some similarity criterion. Some methods use only single type of features while others have integrated multiple features to increase robustness in tracking. The following section describes the recent advances in feature based methods.

5.1.1 Color based methods

Color is one of the features that is used extensively in literature to track objects. Recent advances using color as a feature often use color histograms to model the object. Besides having low computational cost, color histogram distribution is robust against non-rigidity, scale and rotation.

In [15], Mean-shift algorithm, a nonparametric density gradient estimator is used that finds the image window that is most similar to the object's color histogram in the current frame. The mean-shift tracker maximizes the appearance similarity iteratively by comparing the histogram of the object and the histogram of the window around the candidate object location. This has simple implementation but it cannot handle scale variations and clutter. For computing similarity between two histograms, Bhattacharya distance is used.

Using the Bhattacharya color histogram distance, under HSV color space, but within the probabilistic framework is used in [24]. A Monte Carlo estimation method, called particle filter, is used to better handle color clutter in the background. Li and Zheng[25] used RGB color histogram and particle filter by adding two auxiliary variables in the particle state space. These variables control the updating speed of the color observation mode. Another method based on histogram based particle filtering is proposed by Fotouhi et al. [26]. To handle the object appearance changes, focusing on pose and illumination changes, Fotouhi et al [26] proposed a novel adaptation target histogram according to FIFO queue concept, to save the histogram of object at each frame. The model histogram is computed recursively, based on weighted averaging of histograms for each bin. The method is robust against partial occlusion, rotation, scaling and object deformation.

Zhao [27] presented the use of matrix filters, by keeping the important information, for building the target's histogram model for robust object tracking. The author also presented an approach for combining histogram with different feature spaces into an enhanced histogram. A hybrid histogram matching, using combination of Bhattacharyya and Chi-squared similarity measures is presented in [28]. They concluded that in general, Bhattacharyya metric performs better than Chi squared, but the combined approach gives better results in case of multimodal histograms.

The histogram based models, however, is sensitive to noise and occlusion, since it loses spatial information. To alleviate this, algorithms based on multi-region/patch-based representations [29] have been proposed. The patch based tracker divides the target into several regions and extracts features from these regions.

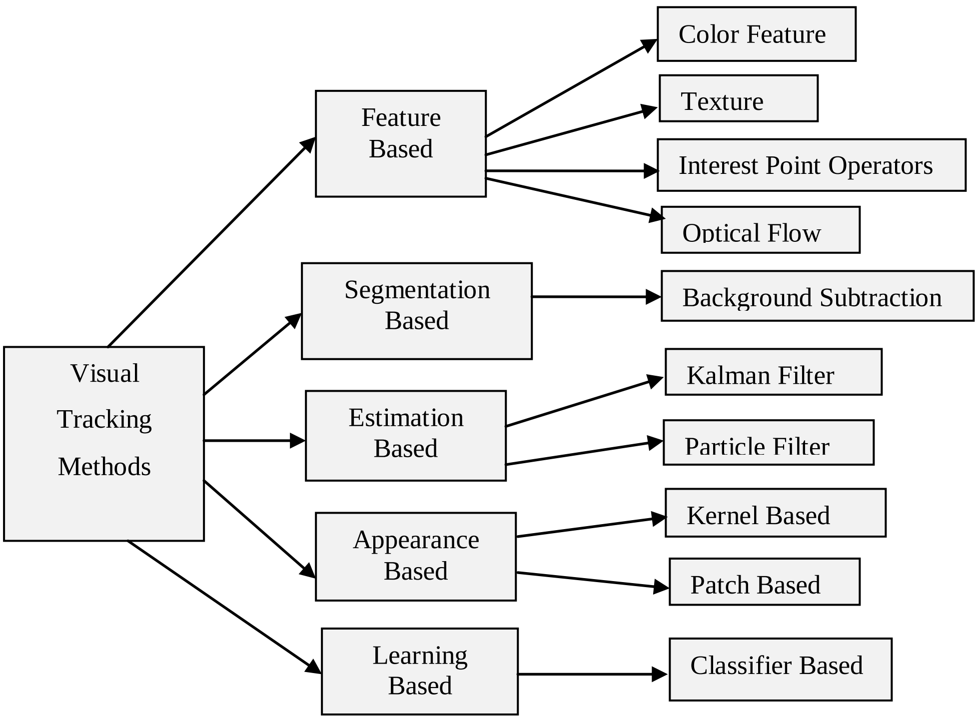

Figure 2: A classification of visual tracking methods

5.1.2 Interest point operators

A local feature is an image pattern that differentiates the object from its nearest neighbourhood. The features can be points, edge segments or small image patches. When the location derived from the feature extraction process is used in applications then these features are referred as interest points.

Lowe [30] used SIFT (Scale-Invariant Feature Transform) features that computes a histogram of local oriented gradients around the interest point. The features are highly distinctive and are invariant to image scaling and rotation, and partially invariant to change in illumination and 3D camera viewpoint. However, its computation involves descriptors which are computational intensive. Over the years, many improvements to SIFT exist. SIFT is designed mainly for gray images. Adgel-Hakim and Farag[31] extended SIFT from grey images to colored images and introduced colored local invariant feature descriptor(CSIFT) which is more robust than SIFT with respect to color and photometric variations.

Many researchers combined SIFT features with other tracking methods to increase the robustness. Wang and Hong[32] combined SIFT with Camshift in which the algorithm adaptively choose SIFT features tracking or Camshift tracking. Yan[33], added SIFT matching features into the particle filter tracking framework based on color histogram feature. In this, the particle weights are calculated considering both the SIFT matching features and the color histogram feature.

Another feature [34] that is used in tracking is SURF (Speeded Up Robust Features). It provides faster calculation as compared to SIFT, thus making it appropriate for real time processing. The detector is based on the Hessian matrix and relies on integral images to reduce the computation time. To improve the robustness, SURF features are combined with color based object tracking methods. Zhou[35] integrated SURF and Mean-shift tracking to overcome the limitations of traditional mean-shift. Shuo et. al.,[36] selected SURF features from test region. In the matching stage, they calculated the Euclidean distance between the SURF descriptor vectors in the test image and ones in the reference image, thus making the tracking robust against appearance variations, scale change and cluttered scenes.

Dalal and Triggs[37] presented Histograms of Oriented Gradient (HOG) descriptors in which a local 1-D histogram of gradient directions of small connecting regions is accumulated and normalized before using them. They used linear SVM as a classifier for human detection. In [38], HOG features are used for tracking using on-line boosting. Yucai et. al[39] used HOG features for a service robot to detect and follow people in indoor environments.

5.1.3 Texture based methods

Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity [23]. Gabor wavelet is one the most studied texture feature. The most important property of Gabor filters is their invariance to illumination, rotation, scale and translation which makes it suitable for object tracking. A method to detect the location and body shapes of moving animals using Gabor filter is presented in [40].

Local Binary Patterns (LBP) is another texture based descriptor that is used for many applications, such as face recognition and motion analysis. The LBP operator, which is rotation invariant, is defined as a grayscale invariant texture measure, derived from a general definition of texture in a local neighborhood [22]. Also, the operator is computationally simple, as the operators can be realized with a few comparisons in a small neighborhood and a lookup table. LBP features are combined with other features and tracking methods to increase the robustness of tracking. Zhao et. al.[41] used LBP to describe moving objects and used Kalman filter for target tracking. In [42] Camshift algorithm is implemented using LBP combined with Hue information. The LBP is sensitive to noise, therefore another variant of LBP called Local Ternary Patterns (LTP)[43], which is more discriminate and less sensitive to noise, is proposed.

5.1.4 Optical flow based methods

An apparent change of a moving object's location in the form of motion vectors between frames is represented by Optical flow. Discontinuities in the optical flow help in segmenting images into regions that correspond to different objects. At present there are many methods for its calculation, which presumes the conservation of brightness intensity. The two widely used methods are: Lucas-Kanade and Horn-Schunck [5]. Hariyono et al[44]., presented a pedestrian detection method using optical flow to extract a moving object and used HOG features to recognize the object using linear SVM. kim et. al.[45] used background subtraction and optical flow simultaneously, thus eliminated the stationary camera restriction for background subtraction. The actual detection of moving object was done by optical flow.

5.2 Estimation Based Methods

Estimation or filtering based methods formulate the tracking problem as a state estimation problem, in which an object is repersented by a state vector. The state vector describes the dynamic behaviour of the system and includes motion characteristics of the concerned object, such as position, velocity and accleration.

The general framework for dynamic state estimation problems is given by Bayesian methods. In this approch the probability density function of the state vector is computed using all availaible information at that time. The posterior probability density function includes all the information about the object state, thus it provides a complete solution of the state estimation problem. To compute the posterior probability density functions of the states, two models are used. They are state model and observation model. The state model describes the evolution of the system, wheras, the observation model describes the relationship between observation and state. Kalman Filter and Particle Filter are two typical Bayesian methods that are widely used to estimate the tarjectory of objects. A detailed description of these two filters is presented in [46]. The filters consist of two steps: prediction and Correction. The prediction step estimates the new state of variables at the next time step, using the state model, whereas, the correction step uses the current observations to update the object's state, using the observation model. The prediction and correction steps are carried out in each image frame of the video.

5.2.1 Kalman filter

If the state posterior density is a Guassain, Kalman Filter[47] is be used. The Kalman filter is applied only for tracking objects with linear motion models. For non-linerar motion models, the extended Kalman filter is used [48], through the first-order linearization of the nonlinear systems. However, this can introduce large errors in the true posterior mean and covariance of the transformed Gaussian random variable, which may sometimes lead to divergence of the filter procedure. To overcome this problem, the unscented Kalman Filter[49] can be used. It uses a set of discretely sampled points to parameterize the mean and covariance of the posterior density[15].

5.2.2 Particle filter

The Kalman filter assumes that the state variables follow a Gaussian distribution, thus it gives poor estimate of state variables that do not follow a Gaussian distribution. Most real tracking problems are non-linear and non-Gaussian. For such problems, the particle Filter [7-8, 25] is widely used by researchers. The particle filter is a recursive Monte Carlo statistical computing method, which is used to solve a Bayesian estimation problem under the condition that measurement models are corrupted with noise which may be non-Gaussian and multimodal. A detailed description on particle filters is given in [46]. The main idea of the particle filter is to represent the posterior density function by a set of samples with weights corresponding to their sampling probability. These samples are called particles. The weight of a particle defines its importance, i.e. its observation frequency [23]. The main problem associated with the particle filter is sample degeneracy and impoverishment. The sample impoverishment refers to a situation in which particles that have high weights are statistically selected many times. It is severe in case of small process noises. To reduce the effect of degeneracy, resampling is used. Resampling means drawing samples either randomly or using some statistical methods, with replacement from a set of data points.

It is observed that the particle filter does not perform well when the dynamic system has a very small noise or if the observation noise has very small variance. To overcome this problem, the Regularized Particle Filter (RPF)[52] and Kernel Particle Filter[51] are proposed, with some success.

5.3 Segmentation based methods

Segmenting foreground objects from a video frame is a fundamental and the most critical step in visual tracking. Foreground Segmentation is done to separate foreground objects from the background scene. Normally, the foreground objects are the objects that are moving in a scene. To track these objects, they are to be separated from the background scene. The following section discusses techniques to track moving objects using segmentation methods.

5.3.1 Background subtraction

Background subtraction [1-3], as discussed earlier, requires regular updating of the background model to adapt to gradual or fast illumination changes and motion changes. In [52], important background models developed up to 2014 are discussed. The authors presented a classification based on mathematical models.

An approach based on a self organizing method without any prior knowledge about the involved patterns is used to generate the background in [53]. The background model is updated by automatically learning background variations using a neural network. To handle effectively both sudden and gradual background changes, in [54], authors used a learning based background subtraction approach. They formulated the background modeling step as a dictionary learning problem, and the foreground detection step as a modified sparse coding problem. In [55], the background is modelled by rectangular regions, described by a color histogram and a texture measure. The current frame in which motion is to be detected is modelled in the same way as the background. The motion is detected by comparing the corresponding rectangular regions computed at the coarsest scale to the finest scale. The Guassian mixture method is finally applied at the finest scale.

Most of the algorithms require a few initial frames to initialize the background. A technique, called “ViBe” (for “VIsual Background Extractor”) presented in [13], requires only a single frame to initialize the background. This frees from the need to wait for several seconds to initialize the background model. The authors modelled each background pixel with a set of samples and the current value of a pixel is compared to its closest samples within the collection of samples. Since a background is non-static, the background model should be updated regularly. For updating, the authors selected a value to be replaced randomly, in contrast to other approaches, where the oldest value is replaced first.

To adapt to sudden or gradual illumination changes, a probabilistic background model based on kernel density estimation is proposed in [56]. In this, the background is modelled as a probabilistic model. The kernel density estimation Guassian model is updated at every frame by controlling the learning rate according to the situation. The updating method automatically adapts to various environments and stochastically deletes non-background information or adds new-background values.

A background modeling and object extraction module to enhance object extraction process is proposed by Kumar et. al.[1]. The method uses a Kalman filter for tracking, and it consists of two stages. The first stage uses the background subtraction and temporal difference mechanism to derive an approximate motion field and calculates regional entropy to get the actual moving pixels that have low entropy. The second stage uses the Kalman filtering for object tracking. Their method is found to be effective to eliminate ghost and aperture distortion.

5.4 Appearance based tracking

In real world, objects are typically complex and articulated in nature and often change their shapes over time. To track such objects, the target object’s appearance change is incorporated into tracking algorithms. Despite extensive research on this topic, no conclusive appearance model has been built till date. Therefore, an effective modeling method of appearance change is required for designing a robust visual tracker. In literature, researchers have proposed many appearance models. A detailed review of various 2D appearance models is presented in [21]. Appearance based tracking methods are broadly classified into the following:

5.4.1 Kernel based tracking

Kernel based tracking methods estimate the motion of the target by masking it with an isotropic kernel. The target object is represented by a primitive geometric shape. The tracking process exploits the constancy of appearance in consecutive video frames and the motion is computed by seeking the mode of the appearance distribution function [57]. Generally, kernel based tracking methods assign a weight to each pixel. These weights are then used to estimate the density gradient in the image coordinates to find the new object position. Mean-shift[15] is a nonparametric density gradient estimator that finds the image region that is most similar to the object's color histogram in the current frame. It maximizes the appearance similarity iteratively by comparing the histograms of regions in the current frame and the previous frame. At each iteration, the region is shifted towards the direction where the Bhattacharyya coefficient is maximum. Yilmaz [57] used an asymmetric kernel over an isotropic kernel to represent the object because isotropic kernels include non-object regions which biases the motion estimation. Babu et. al.,[58] improved the performance of mean-shift tracker when the object undergoes large displacements and in case of partial/full occlusion by combining sum-of-squared differences(SSD) and color-based mean-shift tracker. SSD was used for finding the object's location in successive frames and mean-shift tracker was used to track local object properties.

To handle illumination variation, background clutter and partial occlusion, sparse representation based tracking methods have attracted much attention recently. A new multi-kernel fusion based sparse representation tracking is proposed by Wang et. al.,[59]. In this, first, kernel sparse representation (KSR) is presented by introducing the kernel method into sparse representation. Then KSR algorithm is integrated into particle filter framework. Tang and Feng[60] proposed multi-kernel correlation filter(MKCF) based tracker that incorporates both strengths of multiple channels and multiple kernels.

5.4.2 Patch based tracking

To address appearance changes in target object more robustly, patch based tracking divides the target object into multiple smaller local patches that contain the local information. Each local patch represents one part of the target, hence the local patches when combined together can represent the complete structure of the target as a whole.

Kwon and Lee[61] proposed a model that preserves spatial information of the target object well by representing the object by a number of patches and update the model by adding, deleting or, moving patches to a new position using online updating scheme. Li et. al.,[62] proposed appearance models for both foreground and background and obtained foreground probability map. They constructed the foreground appearance model by selecting only pertinent patches, which appear near the center of the bounding box of the target object, thus excluding the background information. The background is modelled at each block location obtained after dividing the image into non-overlapping blocks. The object position is determined by convolving the foreground probability map with the pertinent mask.

To locate visual objects by directly inferring a set of reliable patches through particle filter under a sequential Monte Carlo framework is proposed by Lee et.al.[63]. They represented the object as a collection of features and used voting scheme to locate the object. Li et. al.,[64] model the target as a whole by exploiting features of individual patches and the relative position relationship, by incorporating spatial information, among patches.

5.5 Learning based methods

Visual tracking problem can also be represented as a decision making process in which a classifier is first trained to distinguish between the target object and the background. The different views and appearances of the object are first learned by a classification function and then decision is made on the test region whether it is the target object or not. Features are extracted from different view and appearances and then different supervised learning approaches, such as boosting, decision trees, neural network and support vector machines are applied.

Support Vector Machines (SVM) classifies data into two separable classes by finding the maximum marginal hyperplane in a multidimensional space. The distance between the hyperplane and the closest data points is referred to as the margin and the data points that lie on the boundary of the margin of the hyperplane are called the support vectors.

In [65] the SVM was integrated with the optic-flow tracker to detect the rear end of the vehicles. The SVM was trained offline to classify between vehicles and non-vehicles by maximizing the SVM classification score. To handle large motions in image, Gaussian pyramid from every support vector was used. Tian[66] proposed an ensemble classifier, the linear SVM with online learning process, that automatically select useful key frame of the target from the history information. The ensemble SVM classifier is made up of several classifiers that help in online updating of the classifier that better handles target appearance variation. Tomasz Malisiewicz[67] proposed Exemplar-SVM framework that trains a separate classifier for each exempler, thus creating a collection of simpler exempler-SVM. Each exempler SVM is trained with a single positive example and millions of negatives. This ensemble of SVMs offers good results. To generate more accurate results, Zhang[68] proposed to use multi view SVM framework with online update scheme. The authors used three distinct features: the grey scale value, HOG, LBP to train the classifier and obtained different confidence scores of the candidate.

Boosting is another learning based method to track objects. It is used to reduce the error of any weak learning algorithm and finds accurate classifier by combining many base classifiers. It repeatedly runs a weak learning algorithm over the training data and then combines the classifiers produced by the weak learner into an accurate classifier. Kim[69] proposed multi classifier boosting algorithm which softly partitions the input space for each classifier and learns multiple appearances of the object. Dan[70] proposed instance transfer boosting framework which transfers knowledge from the first frame and previous two frames to the current frame for robust object tracking, thus alleviating drifting issues in tracking. Binh[71] used online training boosting algorithm that limits drifting and loosing of tracking object.

VI. CONCLUSIONS

Over the past few decades, object tracking has been an active research area in the field of computer vision. For tracking objects, object detection is the first step. Object detection and tracking has various practical applications, such as motion analysis, people and vehicle tracking, crowd analysis, etc. This paper presents a comprehensive review of popular object detection and tracking methods developed till date and discusses the main concept of each method. The paper also proposes a categorization of the methods based on their approaches for tracking.

Some successful applications of different tracking methods have been reported in literature, but tracking in an unconstrained environment is still eluding researchers. Hence, further extensive research is required to handle complex situations, such as real time tracking of multiple objects in natural lighting conditions with a dynamic background, shape changing objects, very fast moving objects, etc., in real time on commonly available computing hardware.

REFERENCES

[1] S. Kumar and J. S. Yadav, "Video Object Extraction and its Tracking using Background Subtraction in Complex Environments", Perspectives in Science vol.8, 317-322, 2016.

[2] S.H Shaikh, K. Saeed, N Chaki, "Chapter-2, Moving Object Detection Approaches, Challenges and Object Tracking", Moving Object Detection using Background Subtraction, springerBriefs in Computer Science, , ISBN 978-3-319-07386-6, 2014.

[3] O Barnich and M. V. Droogenbroeck," ViBe: A Universal Background Subtraction Algorithm for Video Sequences", IEEE Transactions on Image Processing, 20(6), 1709-1724, 2011.

[4] S. Chen, L. K. Wang, J. Lan "Moving Object tracking based on Background Subtraction Combined Temporal Difference", International Conference on Emerging Trends in Computer and Image Processing (ICETCIP'2011) Bangkok Dec., 2011.

[5] S. Aslani and H. Mahdavi-Nasab, "Optical Flow Based Moving Object Detection and Tracking for Traffic Surveillance", International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol. 7, No:9, 2013.

[6] K.Kale, S.Pawar, P.Dhulekar, "Moving Object Tracking Using Optical Flow And Motion Vector Estimation", 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015.

[7] X. Sun , H. Yao, X. Lu, "Dynamic multi-cue tracking using particle filter", SIVIP 8 (Suppl 1):S95–S101, 2014.

[8] K.Nummiaro, E. Koller-Meier, L.V.Gool, "An adaptive color-based particle filter",Image and Vision Computing, pp 1-12, 2002.

[9] M. Yin, J. Zhang, H. Sun, W. Gu, "Multi-cue-based CamShift guided particle filter tracking", Expert Systems with Applications 38, 6313–6318, 2011.

[10] P. Dollar, C. Wojek, B. Schiele, P. Perona, "Pedestrian Detection: An Evaluation of The State of The Art', IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 34, No. 4, 2012.

[11] R. A. Hadi, G. Sulong, L. E. George, "Vehicle Detection and Tracking Techniques: A Concise Review", Signal & Image Processing : An International Journal (SIPIJ) Vol.5, No.1, 2014.

[12] M.Andriluka, S Roth, B. Schiele, "People-Tracking-by-Detection and People-Detection-by-Tracking", IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[13] O. Barnich and M. V. Droogenbroeck, "ViBe: A Universal Background Subtraction Algorithm for Video Sequences", IEEE Transactions on Image Processing, 1709-1724, 2011.

[14] D. P. Chau, F. Bremond, M. Thonnat, "Object Tracking in Videos: Approaches and Issues. The International Workshop "Rencontres UNS-UD", Danang, Vietnam, 2013.

[15] D. Comaniciu, V. Ramesh, P. Meer, "Kernel-Based Object Tracking", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 25, NO. 5, 2003.

[16] A. Yilmaz, O. Javed, M. Shah, " Object Tracking: A Survey", ACM Computing Surveys (CSUR), 38(4):13, 2006.

[17] A. S. Jalal and V. Singh, "The State-of-the-Art in Visual Object Tracking", Informatica 36, pp 227-248, 2012.

[18] Z. Pan, S. Liu, W. Fu, "A Review of Visual Moving Target Tracking", Multimedia Tools Appl, Vol. 76, Issue 16, pp 16989–17018, 2016.

[19] R. A. Hadi, G. Sulong, L. E. George, "Vehicle Detection and Tracking Techniques: A Concise Review", Signal & Image Processing : An International Journal , vol.5, No.1, February 2014.

[20] A. P. Shukla and Mona Saini, "Moving Object Tracking of Vehicle Detection: A Concise Review", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.8, No.3, pp.169-176, 2015.

[21] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, A. V. Henge, "A Survey of Appearance Models in Visual Object Tracking", ACM Transactions on Intelligent Systems and Technology ,Vol. 4 Issue 4, 2013.

[22] H. Yang, L. Shao, F. Zheng, L. Wang, Z. Song, “Recent Advances and Trends in Visual Tracking: A Review”, Neurocomputing, 74, pp. 3823–3831, 2011.
[23] F. Porikli and A.yilmaz, “chapter 1 - Object Detection & Tracking”, Video Analytics for Business Intelligence, Springer, e-ISSN 1860-9503, 2012.
[24] P.Perez, C.Hue, J, Vermaak, M. Gangnet, “Color –Based Probabilistic Tracking” European Conference on Computer Vision ,2002.
[25] X. Li, N. Zheng, “Adaptive Target Color Model Updating for Visual Tracking Using Particle Filter”, IEEE International Conference on Systems, Man and Cybernetics, pp 3105-3109, 2004.
[26] M. Fotouhi, A. R. Gholami, S. Kasaei, “Particle Filter Based Object Tracking using Adaptive Histogram” Iranian Conference on Machine Vision and Image Processing, 2011.
[27] A. Zhao, “Robust Histogram based Object Tracking in Image Sequences”, Digital Image Computing Techniques and Applications, 2007.
[28] N. Naik, S. Patil, M. Joshi, “A Scale Adaptive Tracker using Hybrid Color Histogram Matching Scheme”, ICETET, IEEE Computer Society, pp 279-284, 2009.
[29] S. He, Q. Yang, R.W. Lau, J.Wang, M.H. Yang. “Visual Tracking via Locality Sensitive Histograms”, IEEE conf on Computer Vision and Pattern Recognition, 2013.
[30] D.G. Lowe,” Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision 60(2):91–110, 2004.
[31] A. E. Abdel-Hakim and A. A. Farag, “CSIFT: A SIFT Descriptor With Color Invariant Characteristics” IEEE Conf. on Computer Vision and Pattern Recognition, 2006.
[32] Z. Wang , K. Hong “A New Method For Robust Object Tracking System Based on Scale Invariant Feature Transform and Camshift” Research in Applied Computation Symposium, ACM, 132–136, 2012.
[33] Y. Yan, J. Wang, C. Li, Z. Wu. “Object Tracking Using Sift Features in a Particle Filter” IEEE International Conference on Communication Software and Networks, 2011.
[34] H. Bay, T. Tuytelaars, L. V. Gool, “Surf: Speeded Up Robust Features”, European Conference on Computer Vision, pp. 404-417, 2006.
[35] Z. Zhou, X. Out, J. Xu,” Surf Feature Detection Method Used in Object Tracking”, International Conference on Machine Learning and Cybernetics, Tianjin, 1865-1868, 2013.
[36] H. Shuo, W.Na, S.Huajun, “Object Tracking Method Based on Surf”, Aasri Conference on Modelling, Identification and Control, 2012.
[37] N. Dalal and B. Triggs, Histograms of Oriented Gradients for Human Detection , IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
[38] S. Sun, Q. Guo, F. Dong, B. Lei,” On-Line Boosting Based Real-Time Tracking with Efficient Hog”, IEEE International Conf. on Acoustics, Speech and Signal Processing, 2013.
[39] K. Yucai, L. Shirong, W. Jiangping, Z. Botao, “Person following Based on Haar-like Feature and HOG Feature in Indoor Environment”, 35th Chinese Control Conference, China, 2016.
[40] D. A. Wagenaar, W. B. Kristan Jr, “Automated Video Analysis of Animal Movements Using Gabor Orientation Filters” Neuroinformatics 8:33–42, 2010.
[41] Z. Zhao, S. Yu , X. Wu, C. Wang, Y. Xu, “A Multi-Target Tracking Algorithm Using Texture for Real-Time Surveillance”, IEEE International Conference on Robotics and Biomimetics Bangkok, 2009.
[42] G. Yuan, Y. Gao, D. Xu, “A Moving Objects Tracking Method Based on a Combination of Local Binary Pattern Texture and Hue”, Advanced In Control Engineering And Information Science, Procedia Engineering 15, 3964 – 3968, 2011.
[43] X. Tan and B. Triggs, “Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions”, IEEE Trans. On Image Processing, Vol 19, Issue 6, 2010.
[44] J. Hariyono, V. Hoang, K. Jo, “Moving Object Localization using Optical Flow for Pedestrian Detection from a Moving Vehicle”, The Scientific World Journal, 2014.
[45] D. Kim, J. Kwon, “Moving Object Detection on a Vehicle Mounted Back-Up Camera”, Sensors, 16(1), 2015.
[46] Z. Chen, “Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond”,2003.
[47] N. Najafzadeh, M.Fotouhi, S. Kasaei, “Object Tracking using Kalman Filter with Adaptive Sampled Histogram”, 23rd Iranian Conf. on Electrical Engineering, 2015.
[48] A. E. Nordsjo, “A Constrained Extended Kalman Filter for Target Tracking”, IEEE Radar Conference, 2004.
[49] E.A. Wan, R. V. Merwe, “The Unscented Kalman Filter for Nonlinear Estimation”, IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000.
[50] C. Musso, N. Oudjane, F. LeGland, “Improving Regularized Particle Filters” , chapter 12 in Sequential Monte Carlo Methods in Practice, Springer-Verlag, 2001.
[51] C. Chang and R. Ansari, “Kernel Particle Filter for Visual Tracking”, IEEE Signal Processing Letters, Vol. 12(3), 2005.
[52] T. Bouwmans, “Traditional and Recent Approaches In Back-Ground Modeling for Foreground Detection: An

Overview" Computer Science Review, II-I2 31-66, 2014.

[53] L. Maddalena and A. Petrosino, A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications, IEEE Transactions on Image Processing, Vol. 17, No. 7, 2008.

[54] C. Zhao, X. Wang, W. Cham, 'Background Subtraction via Robust Dictionary Learning", EURASIP Journal on Image and Video Processing, 2011: 972961,2011.

[55] P.D Z. Varcheie, M. Sills-Lavoie G. Bilodeau, "A Multiscale Region-Based Motion Detection and Background Subtraction Algorithm", Sensors, 10, 1041-1061, 2010.

[56] J. Lee, M. Park, "An Adaptive Background Subtraction Method Based on Kernel Density Estimation", 12, 12279-12300, Sensors 2012.

[57] A. Yilmaz, "Kernel-Based Object Tracking Using Asymmetric Kernels With Adaptive Scale And Orientation Selection", Machine Vision and Applications, 22:255–268, 2011.

[58] R. V. Babu, P. Pe´rez, P. Bouthemy, "Robust Tracking with Motion Estimation and Local Kernel-Based Color Modeling", Image and Vision Computing, Vol. 25, 1205–1216, 2007.

[59] L. Wang, H. Yan, K. Lv, C. Pan, "Visual Tracking via Kernel Sparse Representation with Multikernel Fusion", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 24, No. 7, 2014.

[60] M. Tang and Jiayi Feng, "Multi-kernel Correlation Filter for Visual Tracking, IEEE International Conf. on Computer Vision, 2015.

[61] J. Kwon and K. Mu Lee "Highly Non-Rigid Object Tracking via Patch-based Dynamic Appearance Modeling", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 35, No. 10, 2013.

[62] Y. Li, J. Zhu, S.C.H. Hoi, "Reliable Patch Trackers: Robust Visual Tracking by Exploiting Reliable Patches", IEEE Conf. on Computer Vision and Pattern Recognition, 2015.

[63] D. Y. Lee, J. Y. Sim, C. S. Kim, "Visual Tracking using Pertinent Patch Selection and Masking", IEEE conf on Computer Vision and Pattern Recognition, pp.3486-3493 ,2014.

[64] F. Li, X. Jia, C. Xiang, H. Lu, "Visual Tracking with Structured Patch-based Model", Image and Vision Computing, Vol 60, pp 124-133, 2017.

[65] S. Avidan, "Support vector tracking," IEEE Trans. Pattern Anal.Mach. Intell., vol. 26, no. 8, pp. 1064–1072, Aug. 2004.

[66] M. Tian, W. Zhang, and F. Liu," On-Line Ensemble SVM for Robust Object Tracking", ACCV 2007, Y. Yagi et al. (Eds.) Part I, LNCS 4843, pp. 355–364, 2007.

[67] T. Malisiewicz, A. Gupta, A. Efros, "Ensemble of Exemplar-SVMs for Object Detection and Beyond", IEEE International Conference on Computer Vision, pp 89-96,2011.

[68] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object Tracking With Multi-View Support Vector Machines", IEEE Transactions on Multimedia, Vol. 17, No. 3, pp 265-278, 2015

[69] T. Kim, T. Woodley, B. Stenger, R. Cipolla, "Online Multiple Classifier Boosting for Object Tracking", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2010.

[70] Z. Dan, N. Sang, R. Huang, S. Sun, "Instance Transfer Boosting For Object Tracking", Optik 124(2013), pp 3446–3450, 2013.

[71] N. D. Binh, "Online Boosting-Based Object Tracking", International Conference on Advances in Mobile Computing and Multimedia, ACM 978-1-4503-3008-4/14/12, 2014.