International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

A Study On Object Tracking and its Approaches

Akanksha[1], Nirupma Tiwari[2]
[1,2]Department of CSE/IT
[1,2]SRCEM Banmore, Gwalior, India

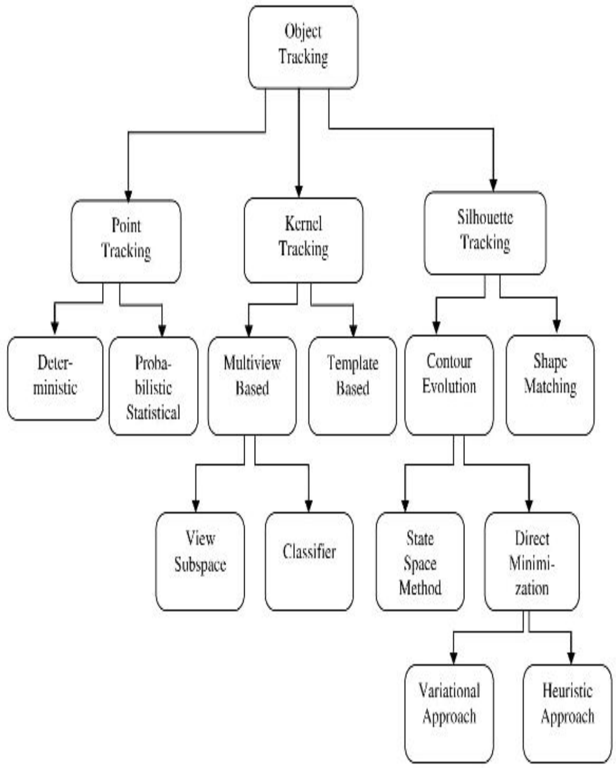
Abstract: *Tracking is the procedure which identifies the position of the moving object within the video. Identifying the position may be a far more challenging task than detecting the moving object during a video. Tracking of objects is applicable for many applications like in robot vision, monitoring the traffic, Video surveillance, Video in-painting and Simulation. Here we are getting to present a quick review of various object detection, object classification and object tracking algorithms available. so as to beat the difficulty of detection, tracking associated with object movement and appearance. Most of the algorithm focuses on the tracking algorithm to smoothen the video sequence. On the opposite hand, few methods use the prior available information about object shape, color, texture then on. Tracking algorithm which mixes above stated parameters of objects is discussed and analyzed during this research. The goal of this paper is to research and review the previous approach towards object tracking.*

INTRODUCTION:

Tracking is that the process of following an object of interest within a series of frames, from its initial appearance to its last. the sort of object and its description within the system depends upon the appliance . During the time that it's present within the scene, it's going to be occluded (either partially or fully) by other objects of interest or fixed obstacles within the scene. A tracking system should be ready to predict the position of any occluded objects through the occlusion, ensuring that the thing isn't temporarily lost and only detected again when the thing appears after the occlusion. the methods are often extended to style an algorithm which may help to beat occlusions. Object tracking systems are typically geared toward surveillance applications where it's desired to watch people and or vehicles moving about a neighborhood . Systems like these get to perform in real time, and be ready to affect world environments and effects like changes in lighting and spurious movement within the background (like trees occupying the wind). Other surveillance applications include data processing applications, where the aim is to annotate video after the event. Target representations are often categorized into two major classes. One is for a set of general objects, like human bodies or faces, computer monitors, motorcycles, and so on. the opposite is for one precise target including a selected person, car, toy, building then on. The targets are often images, concrete objects or maybe abstract feature points.The aim of a target tracker is to make the trail of an object over time by locating its pixel position altogether images of the video. Object tracker can also give the entire region within the image that's engaged by the thing at whenever instant. The tasks of finding the target and fixing correspondence between the thing instances across the frames also can be performed jointly or separately. within the first case, probable object regions in each frame are obtained by means of an object detection technique, then the tracker corresponds to things across frames. Ultimately, the target region and mapping is jointly estimated by iteratively updating object region and site information obtained from previous frames. In either tracking approach, the objects are represented using the appearance and shape models.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

Figure 1.

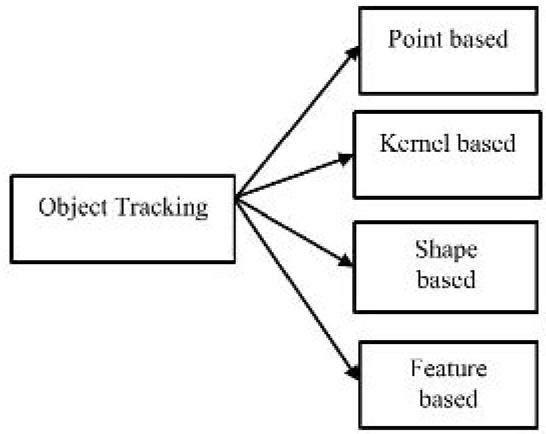


Tracking systems are required to function in a wide variety of conditions. System must be able to function in both indoor and outdoor environments, and need to be able to deal with challenges such as illumination changes and changing weather conditions (i.e. fog, rain, snow).

Other challenges such as occlusions are commonplace in real world scenarios, and systems must be able to maintain an object's identity and a reasonable approximation of its position during these occlusions. In order to overcome some of these problems, algorithms that are able to utilise multi-camera setups have been developed (S. Khan et al. 2003). This however introduces additional challenges such as camera calibration, track handover between views, and how to utilise cameras.

Object Tracking: Object representation is that the process of finding the route of the moving objects during a sequence of the pictures . Object tracking is accomplished to seek out or produce the trail for an entity by discovering its location in each frame. Four major approaches of the tracking are Point based tracking, Kernel based tracking, Shape based tracking, Motion based tracking etc. as shown in figure 6. Some of the foremost widely used tracking techniques are Mean shift tracking, CAM-Shift tracking, KLT tracking etc.

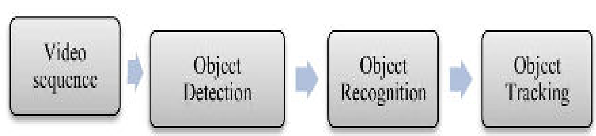
Figure 2.



Subsequently, this tracking is often explained because of the procedure of determining the orientation of an object across the time because the object moves throughout a scene. This is posting importance in the arena of computer vision because of expansion of high powered computers and the growing need for automated surveillance systems, and it is widely implemented for applications namely automated surveillance, robotics monitoring, human-machine interface, motion-based recognition, vehicle navigation, traffic monitoring and video indexing.

In this object tracking application, the target object could be determined as anything which is engaging for analysis. In addition, moving objects tracking is one among the main tasks in computer vision and broadly applied in industrial vision, intelligent transport systems and visual surveillance.

Figure 3.



Studies related to object tracking: Studies associated with object tracking: the thing tracking is that the term which wont to identify the moving object position also as tracking them from video sequences (Balasubramanian et al., 2014). The tracking method is assessed into three types like kernel, point and silhouette based tracking (Yilmaz et al., 2006). Compared to silhouette methods, most of them have focused on kernel-based methods due to high accuracy with less computational cost. Nevertheless, the point tracking method has low computational cost with reduction in accuracy (Weng et al., 2013). The various sorts of object tracking techniques are shown in figure1.

1. **Point based tracking:** Moving objects are usually denoted by their feature point within the image sequence during tracking. Point tracking doesn't suit for occlusion conditions and false object detection. Some of the points tracking algorithms are as described below.

a.) Kalman Filter: The Kalman filter is valuable for tracking distinct types of moving objects. It was initially created by Rudolf Kalman at NASA to track the path of spacecraft. Kalman filters can be utilized by various different types of linear dynamical systems. These are built on the Maximum Periodic Documents Dispensation Algorithm. The Kalman Filter achieves the obstructive probability compactness propagation. It uses a gaggle of scientific equations that gives a cheap computational thanks to evaluate the object's state of occurrence. The Kalman filter measures a procedure through a sort of opinion control. The Kalman filter is a recursive two-stage filter. The two phases are "predict and update". In the predict phase, the current location of the moving object is estimated or predicted to support the previous observation made. For illustration, if an object is moving with consistent dispatch, the current location of the object at Xt often supports its prior location, Xt-1. In the update phase the capacity of the thing ‟s current location is pooled with the anticipated location and acquires the posteriori projected current position of the object.

b.) Particle Filtering: Particle Filter is the most widely used filter to track single and multiple moving objects. It is a hypothesis tracker, filtered subsequent distributions are estimated by a set of weighted particles. One restriction imposed by the Kalman filter is that predefined state variables are normally dispersed (Gaussian). Usually this algorithm uses texture feature mapping which contours, color features etc. The particle filter is a Bayesian chronological Trial procedure, which uses a recursive approach. It also contains two phases almost like that of Kalman filter, predict and update phase.

2. **Kernel based tracking:** Kernel based tracking is usually performed to work out objects in motion, which is symbolized by the elementary object area, from frame to border . The object movement is usually within the method of parametric movement like transformation, affine, etc. This type of tracking can be applied for both rigid and non-rigid moving objects. Kernel based tracking usually supports the thing's representation, appearance, silhouette of the thing .

a.) Mean shift tracking: In mean shift tracking the position of the nonstationary object is tracked centred on the histogram. In this method the region of the object can be traced constructed on the similarity computation of the object. An inclined rise scheme is introduced to transfer the tracker to a position that exploits a similarity value within the model and then the current frame region. Target regions are usually selected in rectangle form or elliptical form. This tracking procedure consists of target model and candidate model. Color histogram is typically chosen to characterize the target. Probability Density Function is employed to represent the target model. If the similarity scores match then the object is tracked.

b.) Simple Template Matching: Template matching is a heuristic search process of scrutinizing the important areas in the video. A background image equalises with the sequence of images that are seperated from a video. This algorithm is applicable for single moving object tracking. Template matching is the method for handling sequences of images to discover minor portions of an image that matches the model with a template image in every frame. The matching technique encompasses a picture pattern for all probable locations within the basic image and computes a geometrical key that identifies how accurately the model suits the image position. This process is widely used when a single object or particle occluded object is to be tracked.

c.) Layering based tracking: This is often the foremost common mechanism for tracking multiple non-stationary objects during a video. Every layer comprises shape illustration, movement like rotation and translation, and layer presence, depending upon power. Layering is obtained mainly by recompensing the background movement, such that the entity"s movement can be calculated by the content image via 2D parametric gesture. All pixel"s credibility are estimated to support the entity"s previous gesture and shape structures. It can also track multiple images and complete the occlusion condition of the item.

3. **Shape based tracking:** Some objects would have complex shapes like hand, finger, shoulder, nose etc. which cannot be demonstrated by exact geometric shapes. The goal of this system is to spot the moving object state in each and every one frame through the object model produced within the prior frame. This technique is proficient in tracking complex shape objects. Occlusion condition, split and merge condition also can be handled during this method. Information of the thing region is employed by this system to trace the objects.

a.) Contour Tracking: Contour tracking methods, repeatedly process a principal contour of the prior structure to its new area within the present image sequence. This silhouette procedure needs specific measures regarding items present within the current frame intersection with the entity district within the prior frame. Contour Tracking could also be implemented utilizing two distinct methodologies. The principal methodology utilizes formal space models to demonstrate the shape , shape and movement. The ensuing approach straightforwardly develops the form by reducing the contour vitality using direct reduction methods, likewise gradient descent. It is flexile to handle huge variation in shape of items.

b.) Shape Based Matching: These are the techniques which are used to introspect the object model in the surviving frame. Shape identical algorithms work analogously to the template based tracking method. This method is employed to spot the matching shapes detected in 2 consecutive frames. Shape matching is often measured like point matching. Background subtraction mechanism is employed for detecting the form of the thing . Procreation of objects is executed within the structure of density gatherings, shape, frontier, edges of the objects. Shape based matching is proficient in handling solitary objects and Hough renovate technique is employed to handle occlusion conditions.

4. **Feature Based tracking:** Feature based tracking is the most widely used tracking technique nowadays. It is mainly divided into 2 steps. First phase is to extract the features of the entity like centroid, shape, color etc. Another phase is to equalise those features in every frame. One or more features can be combined to obtain better results or outcome.

a.) Color feature based tracking: The data about color of all objects in a video frame is tacked group-by-group. Specific weights are assigned for each group for further evaluation. The color evidence is obtained from the movement chunks within the current frame that are divided into regions of nearly alike color as a group. Color information can be either RGB or HSV. The next step is to acknowledge matching color information. This is accomplished by matching the group of color data of the motion chunk within the existing frame with the cluster color evidence of motion chunks within the prior frames using weighted matching. The maximum evaluation score of each group within the existing frame is achieved and therefore the tracker is initialized thereto position. This process repeats until the end of the video sequence.

b.) Centroid feature tracking: The key feature centroid of each object is extracted. The centroid information is accessed from the gesture chunks in the current frame to classify identical centroid information among gesture blocks in the existing frame and prior frames. Accordingly, a bounding box with centroid is consigned for the gesture blocks within the existing frame. New centroid is detected by averaging the sum of previous centroid as in the equation.

$$Cen_n = \frac{\sum_{i=1}^{n-1} cen_i}{n}$$

where Cen_n is the new centroid point and Ceni prior centroid and n is the complete number of frames processed. Results of the centroid feature based tracking is as shown below.

Figure : Tracking results for centroid based object tracking

c.) Edge feature tracking: Edge feature tracking is similar to the color based feature tracking. Edge of the moving object can be determined by various edge detection techniques such as canny edge detection, sobel edge operator etc. in every frame. The next step is to spot matching edge information. This is obtained by matching the load information of the gesture block of this frame with the load data of the gesture chunks within the prior frames. The highest score comparison of every group in the current frame is achieved and the tracker is initialized to that position. This method is repeated until the last frame of the video sequence.

Performance Evaluation:

Calculation of performance of the video analysis could be achieved by mining a manually non-stationary object as a ground-truth image and comparing it with the output obtained. Some parameters like True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FP) are estimated. TP refers to the total amount of pixels where both ground-truth and proposed outcome overlap. TN refers to the total number of pixels that does not comprise an object, both in ground truth and proposed. FP is the sum of pixels in which the proposed system contains objects but ground truth does not. FN refers to the total amount of pixels where ground-truth contains an entity and the projected system doesn‟t have objects.

Precision and Recall are some of the important factors for performance evaluation. Recall provides the share of truth positive rate that's almost like ground truth image. It is given by the equation.

Recall= TP/ (TP+FN)

Precision offers the proportion of truth positive that doesn't match with the ground-truth image.The precision is found by using the equation

Precision = TP / (TP+FP)

A) MOTP (Multiple object tracking precision)

Total number of faults, in identified area for comparing object hypothesis within all frames, averaged via sum of comparisons made. It illustrates the capacity of a tracker to determine a specific object area, liberated of its proficiency at identifying objects formation.

$$MOTP = \frac{\sum_{i,t} d_i^t}{\sum_t c_t}$$

B) MOTA (Multiple object tracking accuracy)

$$MOTA = 1 - \frac{\sum_i (m_i + fp_i + mme_i)}{\sum_i g_i}$$

Where mi, fpi indicates the number of misses related to false positives. Time mismatches at an interval of „i" is represented by mmei.

MOTA provides various error percentages that are

$$m' = \frac{\sum_i m_i}{\sum_i g_i}$$

Is the proportion of errors within the group, and is calculated round the sum of quantity of object existing in every frame.

Percentage of false positive and percentage of disparities is given by the equation (1) and (2) respectively

$$fp' = \frac{\sum_i fp_i}{\sum_i g_i} \quad (1)$$

$$mme' = \frac{\sum_i mme_i}{\sum_i g_i} \quad (2)$$

LITERATURE SURVEY:

Sarkar et al. (2012) developed a color information method for identification of skin in images that occurred from FERET especially for mouth and eye region detection. This technique requires less computational cost also as applied to video sequences. However, this method is incapable of identifying the small face images which are away from the camera.

Coşkun and Ünal (2016) suggested a camshift technique towards tracking the object from the video sequences. Further, they demonstrated this approach successfully carried out in mobile platforms even with the change in object size and illumination. The drawback observed during this proposed technique fails to perform for input video with full occlusion. The performance of the obtained framework has been demonstrated for camera calibration on simulated with real data, underlying the problems of single-object localization and tracking, as well as for multi-object tracking. Further, they planned to reinforce the proposed method towards other multi-object filters and a comparative study of those approaches for camera-based tracking also as for camera calibration.

Zhang et al. (2016) proposed an approach by combining frame difference and nonparametric method for video analysis traceability. The simulation result proved this approach performance was better than the normal frame difference and GMM. Further, it's ready to eliminate the noise from a background which gives us the power to spot the moving object more accurately within the applications like food and agriculture related product traceability analysis. However, this study needs to enhance the capability of the traceability system and supply the visual supply chain for the common user to ensure the safety.

Oiwa et al. (2016) suggested a probabilistic background model towards tracking the object from video sequences. The imitation results display the accuracy and effectiveness of this method highly compared to previous techniques. However, this study needs to concentrate on higher speed as well as improve the accuracy of object tracking.

Mohammed and Morris (2014) suggested a color-based technique which was the combination of accruing and normalizing histograms towards object tracking under different conditions based on a mobile device. This technique was easy to use and robust against varying illuminations. However, this system fails to spot the whole region of symbols thanks to the acute camera view.

Aggarwal et al. (2006) suggested a novel technique which was the combination of motion estimation and background subtraction for object tracking using video sequences. The system mainly focused on four scenarios like interpolation, identify the thing , subtract the background and object selection. However, this research must concentrate more about complete occlusions of video sequences, multiple object tracking, fast camera motion and unsupervised object tracking.

Yoav Rosenberg and Michael Werman describe an object tracking algorithm using non stationary cameras. The algorithm is predicated on domain knowledge and motion modelling. Displacement of each point is assigned a discrete probability distribution matrix. Based on the model, an image registration step is administered . The registered image is then compared with the background to trace the moving object.

Ci gdem Ero glu Erdem and Bulent San have detailed a feedback –dependent approach for object tracking in presence of occlusion. In this method several performance evaluation measures for tracking are placed during a feedback circuit to trace non rigid contours during a video sequence

With the help of tracking the object in motion occlusion, Alok K.Watve and Shamik Sural provide a compatibility of outcomes with other algorithms like Kalman Filtering for foreground extraction and camera modelling for Background Subtraction with multiple cameras for both stationary and moving objects. By mining the feature of tracking the objects in Block matching computes the frame difference; in exploiting the domain information the author evaluates the motion parameters with the displacement of scale parameters and displacement vector. The compressed domain object tracking method is computed using the limiting rectangle of an object with unmatched frames by using histogram.

P. Subashini, M. Krishnaveni,Vijay Singh illustrates the word tracking in Frame Rate Display and Color transformation with background subtraction using different techniques like Estimating median overtime, Computing median overtime, and Estimating moving objects. With the comparison of various segmentation algorithms like Sobel Operator, Canny operator and Roberts operator the object is segmented through edge detection and its derivatives are calculated. With the help of region filtering in color segmentation color samples for skin are prosecuted and are calculated to mean and covariance over color channels. Object tracking for non stationary objects over motion vectors is calculated using an optical flow algorithm and Blob analysis for binary features of a picture is calculated. Tracking of objects is measured by the position done by tracking in region filtering and the information of the object is created as an estimation of the new object.

Feris (2000) suggested a statistical skin-color model towards detecting the various regions of a candidate's face as well as applied template matching techniques to define presence or absence of a face in a particular region. They tracked the face of the candidate by lip corners, the location of pupils and nostrils location in the tracked image sequence. However, this study must improve the system robustness and accuracy of detection also as tracking the countenance .

Kim (2007) said that objects are randomly selected by the users and are tracked using SIFT features and a Kalman filter. Specifically, they target tracking human, car, or pre learned objects. The objects are accumulated, exploiting the learning to successfully track objects even when the objects are missing for some frames. However, this study needs to focus on higher resolution with finding the location of stationary objects.

Conclusion: Various feature extraction techniques are discussed and compared within the paper. There are trade-offs between various techniques. We have some different kinds of methods which can increase the accuracy as compared to other techniques but their time complexity is more. Whereas other techniques provide acceptable results by doing calculations relatively fasert. Some techniques are storage efficient while others are time efficient. Hence different techniques suit different needs. The accuracy of results also depends upon the dimensions of the window and spatial distribution of pixels. The results obtained show that the Fuzzy Color Moments is relatively good technique among techniques explained above for visual perception . The work is often further extended by fusion of two or more techniques at the feature level to urge better results. This fusion of those techniques is recommendable because it increases the accuracy. Although, it'll increase the accuracy but also will increase the computational load. The utilization of particular techniques or the fusion of varied techniques depends on the applications that these are getting used . These various techniques are getting used in several applications and areas like visual perception , Image Classification, Content Based Image Retrieval (CBIR), Robotics, and AI based System, Knowledge Based System or Expert System, Computer Vision, Learning System etc. The framework discussed during this paper will help to develop these systems further.

REFERENCES:

- N. K. Kanhere, "Vision-based detection, tracking and classification of vehicles using stable features with automatic camera calibration," ed, 2008, p. 105.
- R. Rad and M. Jamzad, "Real time classification and tracking of multiple vehicles in highways," Pattern Recognition Letters, vol. 26, pp. 1597-1607, 2005.
- K. H. Lim, et al., "Lane-Vehicle Detection and Tracking," Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2009), vol. 2, pp. 5–10, 2009.
- S. Gupte, et al., "Detection and classification of vehicles," Intelligent Transportation Systems, IEEE Transactions on, vol. 3, pp. 37-47, 2002.
- C. R. Wren, et al., "Pfinder: real-time tracking of the human body," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, pp. 780-785, 1997.
- N. A. Mandellos, et al., "A background subtraction algorithm for detecting and tracking vehicles," Expert Systems with Applications, vol. 38, pp. 1619-1631, 2011.
- F. Porikli and A. Yilmaz, "Object Detection and Tracking," in Video Analytics for Business Intelligence. vol. 409, C. Shan, et al., Eds., ed: Springer Berlin Heidelberg, 2012, pp. 3-41.
- L. Jin-Cyuan, et al., "Image-based vehicle tracking and classification on the highway," in Green Circuits and Systems (ICGCS), 2010 International Conference on, 2010, pp. 666-670.
- Bagherpour, P., Cheraghi, S.A. & Bin Mohd Mokji, M., 2012. Upper body tracking using KLT and Kalman filter. Procedia Comput. Sci. 13, 185–191. doi:10.1016/j.procs.2012.09.127
- Blackman, S.S., 2004. Multiple hypotheses tracking for multiple target tracking. IEEE Aerosp. Electron. Syst. Mag. 19, 5–18. doi:10.1109/MAES.2004.1263228.
- Chate, M., Amudha, S., & Gohokar, V., 2012. Object Detection and tracking in Video Sequences. ACEEE Int. J. Signal Image Process. 3.
- Comaniciu, D., Ramesh, V. & Meer, P., 2003. Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 25, 564–577. doi:10.1109/TPAMI.2003.1195991
- Coşkun, M. & Ünal, S., 2016. Implementation of Tracking of a Moving Object Based on Camshift Approach with a UAV. Procedia Technol. 22, 556–561. doi:10.1016/j.protcy.2016.01.116
- Du, W. & Piater, J., 2008. A Probabilistic Approach to Integrating Multiple Cues in Visual Tracking, in: Computer Vision – ECCV

2008. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 225–238. doi:10.1007/978-3-540-886884_17

- Fatima, H., Shah, S.I.A., Jamil, M., Mustafa, F. & Nadir, I., 2013. Object Recognition, Tracking and Trajectory Generation in Real-Time Video Sequence. Int. J. Inf. Electron. Eng. 3, 639–642. doi:10.7763/IJIEE.2013.V3.395