Object Detection and Motion Based Tracking Using LSK and Variant Mask Template Matching

Ms. Priyanka S. Bhawale, Mr. Sachin Patil

Abstract— **The aim of object tracking is to find moving object in a frames extracted from video input. A object tracking system consist of three steps of data tracking; object extraction, object recognition and tracking. This framework takes the region of target object as input in the previous frame and a stored feature of the target object, and locates the object in the current frame. It finds the frame region that matches the input. As the object changes its view, the object model is updated. For tracking of the objects Local Steering Kernels can be used and for matching variant mask template matching can be used. The proposed method overcomes the problem of occlusion.**

Keywords—Object tracking; occlusion; Local Steering Kernels; Variant mask template matching

I. INTRODUCTION

Visual object tracking is very important to many computer vision applications like surveillance system and robotics, gesture recognition and motion recognition. Progress in the imaging transformations, online learning, and object detection has increased the approach of tracking by detection. The target object is identified by the user in the first frame and is described by a set of features. A set of feature describes the background. Another binary classifier separates target from background in successive frames. The appearance changes needs to be handled and the classifier can be updated incrementally over time for it. Visual object recognition is an very difficult for computation. The problem is that each object in the world casts number of 2-D images on the retina. The object position, lighting, and backgrounds change as per the viewer position. Object tracking means following the trajectory of the object in the sequence of image frames. For this the object is to be represented first. For object representation appearance based methods can be used. After the representation of the object the object is to be detected and then the tracking of the object can be done. Suspicious behaviour of the object can be detected and tracked in surveillance system using visual object tracking. In traffic flow monitoring object tracking is used to track the vehicles and monitor the flow of the traffic so as to avoid any jams. One of the applications of object tracking is video compression. Example Video object tracking is applied in banks, parking lots, residential areas, malls for monitoring human activities. Object tracking is also used for hand gesture recognition in human-computer interaction applications. It is difficult to project 3D world into 2D image. This may cause loss of information. Various methods are used for tracking purpose. Object tracking is affected by the effect of noise and the changing illumination conditions of the object of interest. Tracking of object can be complex due to the articulated nature of the object. Occlusion can also be a major problem in the object tracking. Motion of the object can be complex and there can be many real time processing requirements for tracking. Hence proper method must be selected according to the where object tracking is being used. In this paper we discuss different the techniques used for object tracking.

A. Problem Statement

A system to implement Object detection and motion based tracking of moving objects. To set up a system for segmentation and tracking of moving objects in stationary as well as moving camera video scenes. To make significant improvements in commonly used algorithms. Finally, the aim is to show how to perform detection and motion tracking of objects in a video from a stationary camera as well as moving camera. To track the object efficiently in case of illumination changes and in the case of full and partial occlusion.

B. Objective

The aim of object tracking is to find moving object in video frames sequence. Normally a video tracking system combines three stages of data tracking; object extraction, object recognition and tracking. This system takes the region target object in the previous frame and stored features of the object as input. It tries to locate the object in the current frame by finding the matching region. The goal of the work: To set up a system for frame extraction and tracking of objects in stationary as well as moving camera. To make improvement in commonly used algorithm. Finally, the aim is to show how to perform detection and tracking based on motion of moving objects in a video from a stationary camera as well as moving camera. To make changes in the already existing algorithms. To handle the case of illumination changes and partial and full occlusion.

ISSN: 2278 – 1323

II. RELATED WORK

A. *Object Representation Methods*

First step of object tracking is the representation of the object of interest. Object can be represented by their shape and appearance. In this section, we will first describe the object shape representations used for tracking. Then address the joint shape and appearance representations [1]

1) Point: The object is represented by a point, the centroid or it is represented by a set of points [2]. This representation can be used for tracking objects that are in smaller region in an image.

2) Primitive geometric shapes: Object shape is represented by a rectangle, ellipse [2], etc. Object motion for such representations is usually modeled by translation, or projective homography transformation.

3) Object silhouette and contour: The boundary of a region is defined by the contour representation. The region inside the contour is known as silhouette of the object. These can be used in the tracking of complex and non rigid shapes [3].

4) Articulated shape models: Articulated objects are composed of body parts that are held together with joints. For example, the human body can be called as an articulated object with torso, legs, hands, head, and feet connected by joints. The relationship between the parts is governed by kinematic motion models, for example, joint angle, etc. Articulated objects are represented by modeling the constituent parts using cylinders or ellipses.

5) Skeletal models: Medial axis transform is applied to the object silhouette to extract the object skeleton[4]. This model can be used as a shape representation for reorganization of the objects.

B. *Object Detection Methods*

Tracking mechanism requires an object detection mechanism when the object first appear in the video. Once the objects are represented using any of the mentioned models next step is to detect the object in the frame. This is done when the object first appears in the frame or video. The temporal information of the object in the first frame is extracted to detect it. Some models use more than one frame to extract the information; this is done by frame differencing. Some of the object detection methods are as follows:

1) Frame differencing: The moving object is determined by calculating difference between two consecutive images. It has strong adaptability for the variety of dynamic environments[5].

2) Optical Flow : In this method image optical flow field is calculated. And clustering processing is done in accordance to optical flow distribution characteristic of image. This method gets the complete movement information and detects the moving object. This method is sensitive to noise, poor anti-noise performance [6].

3) Background subtraction : First step for background subtraction is background modeling. Background Modeling should be sensitive so as to recognize moving objects. Background Modeling outputs a reference model. In background subtraction each video sequence is compared to the reference model to determine possible Variation. The variations between current video frame to that of the reference frame in the terms of pixel signify the existence of moving objects[7]. Currently, mean filter and median filter are widely used to realize background modeling. The background subtraction method is to use the difference method of the current image and background image to detect moving objects. This is a simple algorithm, but very sensitive to change in the external environment. This method has poor anti- interference ability. It provides the complete object information in the case background is known [8]. Various background subtraction models
are MOG (Mixture of Gaussians), Bayesian decision rules, the Codebook-based model, Kernel density estimation [9]. The Codebook algorithm [10] constructs a background model based on a quantization/clustering method. Firstly for each pixel a background model is constructed. This model contains one or more codeword. A codeword is a data structure which
contains information about color, brightness and frequency [10]. Stauffer and Grimson et al. [1] proposed a Gaussian mixture model based on background model to detect the object. Mixture of Gaussians was used to model the pixel color [1]. The pixel in the frame is compared with the background model. While checking the pixel is compared with every Gaussian in the model till a match is found. If found, the mean and variance of the math is updated [1]. Non parametric Kernel density estimation can be used to model the per-pixel background. The pixel is matched with the pixel in the background model and with the nearby pixels [11].

C. *Object Tracking Methods*

1) Point Tracking : In image, moving objects are represented by their feature points. In the incidence of occlusions there is the problem of false detection of object occurs. Point correspondence methods are divided into two categories, namely, deterministic and statistical methods. The deterministic method use qualitative motion heuristics to oblige the problem of correspondence. Probabilistic methods take the object measurement and uncertainties into account to establish correspondence [1]. Kalman filters can be used for object tracking. Kalman filters are based on optimal recursive data processing algorithms. Kalman filter consists of two phase prediction and correction phases. The next state is predicted using the current set of observation. And then the current set is updated. The second step gives updates the predicted values and gives approximation of the next state

[12]. Particle filter uses contours, color features or texture mapping for object tracking. Variable which is not sampled is selected. Particle filter samples the variable according to proposal distribution [12]. MHT (Multiple Hypothesis Tracking) algorithm is an iterative algorithm. Prediction about the position of the object in the frame is made. Then distance measure is used to compare the prediction.

2) Kernel Tracking : Kernel tracking computes the motion of object form frame to frame. The motion of the object is in form of parametric motion or dense flow field computed in subsequent frames. There are two subcategories density-based appearance models, and multi view appearance models.

Density-based models are simple and have relative low computational cost. Templates are formed using image intensity or color feature. There are three methods in kernel tracking approach Simple template matching, mean shift method, simple vector machine (SVM) and layering based tracking [12]. Simple Template matching can track only single object. The object of interest is verified with the frame form the video. It can deal with the partial occlusion of the object. In Mean shift method the object of interest is defined using rectangular frame. Then the tracked object is separated from the background. It uses translation and scaling to track the object motion. This method can deal with partially occluded objects. The simple vector machine uses training set of values. These training values are positive or negative, positive values contain tracked object and the negative sample contains the values which are not tracked. In Layering based tracking multiple objects can be tracked. Ellipse is used to represent the shape of the object and uses layer appearance based on the intensity. The background motion of the object is compensated and then the each pixels probability based on the foreground motion is estimated. This method can deal with full occlusion problem [13].

3) Silhouette Tracking : This method generates an object model based on the previous frame. Using these object model the object from each frame is find out. This model can be in the form of a color histogram, object edges or the object contour. Histograms of color and edges can be used as the object models. This method models the object appearance by the edge information obtained inside the object silhouette to match silhouettes in consecutive frames [1].

i. Contour Tracking: Contour tracking method progress a contour in the previous frame to new position in the current frame. This requires that a contour in the current frame must match the object region in the previous frame. It can be performed using two different methods. The first method uses state space models to model the contour shape and motion. The second method changes the contour by minimizing the contour energy. This is done by using direct minimization techniques like gradient descent. It can handle a large variety of object shape [12].

ii. Shape Matching: These approaches checks for the object model in the existing frame. Shape matching performance is similar to the template based tracking in kernel approach. Another approach to Shape matching is to find matching silhouettes which are detected in two successive frames. Silhouette matching, can be considered similar to point matching. Detection based on Silhouette is carried out by background subtraction. Models object are in the form of density functions, silhouette boundary, object edges. Capable of dealing with single object and Occlusion handling will be performed in with Hough transform techniques [12].

III. IMPLEMENTATION DETAILS

The input of the system is a video file. The frames are generated from the input video. The region of interest (ROI) is marked on one selected frame. Region of interest is the region which we have to track . Once the ROI is marked, features of that ROI is extracted. Using these feature values and other frames tracking is done using Local steering Kernal.

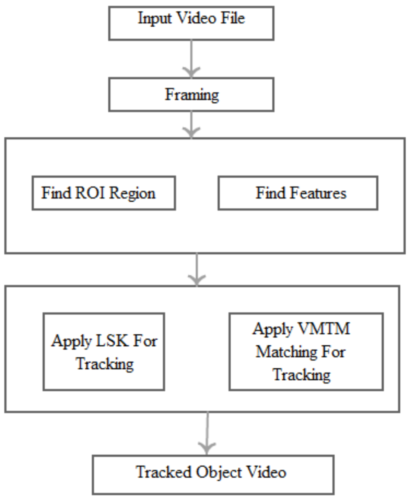


Fig. 1. Architecture Diagram

Kernel tracking computes the motion of object form frame to frame. The motion of the object is in form of parametric motion or dense flow field computed in subsequent frames. There are two subcategories density-based appearance models, and multi view appearance models. Density-based models are simple and have relative low computational cost. Templates are formed using image intensity or color feature. Once the object is detected accurately then the tracking process can be done accurately. For tracking of object Local Steering Kernel method can be used. Color histograms and density values of object can be used for that. Object is tracked form frame to frame. The object information from the current frame is compared with the information from the next frame. If match then object is detected and the new position and values of the object are put on the stack. For further comparison this new values on stack are used The Object is tested for occlusion

ISSN: 2278 – 1323

A. Mathematical Model

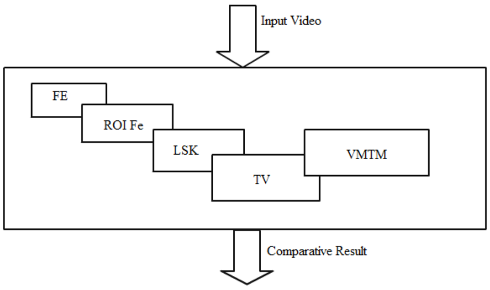


Fig. 2. Mathematical Model.

Let S be a system that describes object tracking system

S={FE,ROI Fe, LSK, VMTM, TV}

Where,

FE: frame extraction

ROI Fe: Region of interest feature extraction

LSK: LOCAL STEERING KERNELS

VMTM: Varient Mask Template Matching

TV: Tracked Video

1. FE(Frame Extraction)

2. ROI Fe:

Region of interest is firstly find out manually by cropping it manually from extracted frame. This is done by cropping region which we want to track. After that we find out its salient features by using LSK. It's useful for matching ROI with other frames of video.

3. LSK(Local steering kernels):

Local steering kernels represent the local structures in images, and give us a measure of local pixel similarities. Given a query image Q, target image T, overall threshold τ0, and a window size P2 , the generic object detection algorithm proposed by involves the following: First, the LSKs for the (grayscale) target and query images must be computed. Let these be denoted by KQ(xi − x; Hi) and K j T (xi−x; Hi), where the subscripts Q and T denote the LSKs for the query and target, respectively, and the superscript j denotes that the kernels were computed for the j th patch in T that is the same size as Q. Now, the LSKs are too dense to use as effective descriptors, and so the next step would be to reduce the dimensionality of these vectors. Before this, however, we first need to normalize our data. Normalization of KQ(xi − x; Hi) and K j T (xi − x; Hi) is given by the following formulas:

$$W_Q(x_i - x) = \frac{K_Q(x_i - x;\, H_i)}{\sum_{l=1}^{P^2} K_Q(x_l - x;\, H_l)}$$

$$W_T^j(x_i - x) = \frac{K_T^j(x_i - x;\, H_i)}{\sum_{l=1}^{P^2} K_T^j(x_l - x;\, H_l)}$$

Let WQ and W j T denote the collection of normalized LSKs for all the xi's in the query, and all the xi's in the j th patch in the target, respectively. After normalization, Principal Component Analysis (PCA) can then be used to reduce the dimensionality. Applying PCA to WQ and extracting the top d Eigen images gives us the collection of Eigen images, FQ, and the projection space AQ that was used to obtain these Eigen images. We then project W j T onto AQ to obtain FT , which is a a collection of Eigen images for the target that are in the same space as that of the query. With these reduced descriptors, we can now compute the similarity between two patches. For this, the Cosine Similarity measure is used. The similarity of the j th patch in the target to the query is given by

$$\rho_j = < \frac{F_Q}{\|F_Q\|}, \frac{F_T}{\|F_T\|} >$$

From this measure, we generate the resemblance map by calculating the resemblance

$$f(\rho_j) = \frac{\rho_j^2}{1 - \rho_j^2}$$

for all patches in the target. Finally, significance tests and non-maximum suppression are applied to find the objects. First, the resemblance map is threshold by the overall threshold τ0 (ideally set to 0.96; see for details), to determine if there are any objects present in the target. If no values are above τ0, then no objects are present. Then, the resemblance map is threshold by a second threshold, τ , which is extracted from the PDF of f(ρj) and is set so that only 1% of the resemblance values are above it. This gives a 99% confidence level in the produced data. The last step is to apply non-maximum suppression to find the locations of the objects

4. Varient Mask Template Matching:

i. As the target location yielded by the first template matching is inaccurate, part of the target might stay outside the ROI (Region of interest)

ii. The portion of the target that lies within the ROI is still correctly identified. We can therefore utilize this information to align the target to its precise location using the VMTM algorithm.

$$\hat{a}_A = \arg\min_a \frac{1}{\mathrm{sum}(M_A)} \sum_{x \in \Omega_T} \left(\frac{I[\phi(x;a)] - \hat{T}(x)}{\sigma_E(x)} \cdot M_A(x;a) \right)^2$$

iii. The accurate target location is acquired by performing the following parameter search:

Where is the rectified transformation parameter vector which defines the rectified ROI, MA is a dynamic template mask which varies with the a under test, and is generated from the preliminary outlier map by

$$M_A(x;a) = 1 - U_{prim}\{\mathrm{round}[\phi(x;a)]\}$$

Where Uprim denotes the preliminary outlier map.

iv. The unmasked part of the template and unmasked part of the ROI are always dissimilar unless the ROI is located exactly where the target is.

IV. Result And Analysis

In order to perform and implement our work data in the form of image is used. A set of images are captured from the video camera. Input is set of images. The images are extracted from the video input. These image frames are further sent for processing. Interactive watershed method can be used to divide the 2D video into sequence of images. Our experimental datasets were extracted from the video

camera which are placed in shopping malls, residential areas, surveillance system. The image frames are then applied to the first algorithm for the detection of the interest region. The result of the first module is the detection of the interest
region from the image frame. The initial step is for the detection of the interest region from the input image frames. The second step takes the output of first step and tracks the object in each frame. The third step acts along with the second for the detection of occlusion and tracks the object in that case.

A. Result comparison with base system

Comparison of different algorithms in handling a challenging long-term occlusion. The two rows from top display the results of the algorithms of and our proposed algorithm, respectively. The five columns from left show frames 47, 147, 163, 184 and 234, respectively. In LSK tracker it fails to detect ROI object when occlusion occurs in frame 163 and 184.while by using VMTM Tracker it successfully track ROI object when occlusion occur.

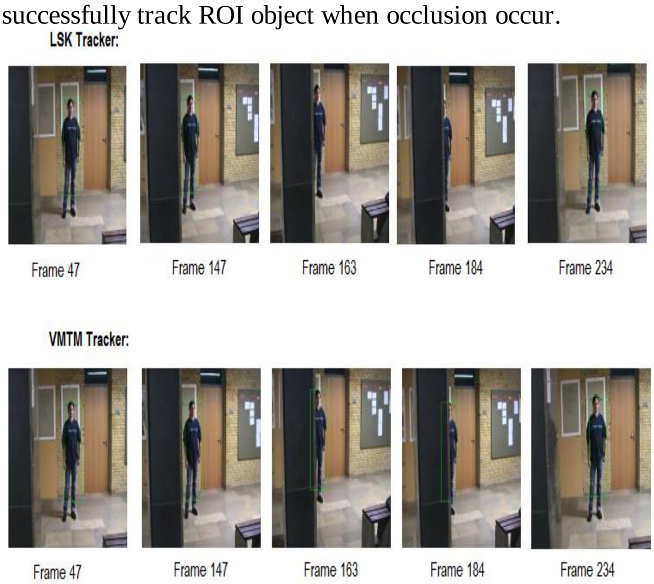
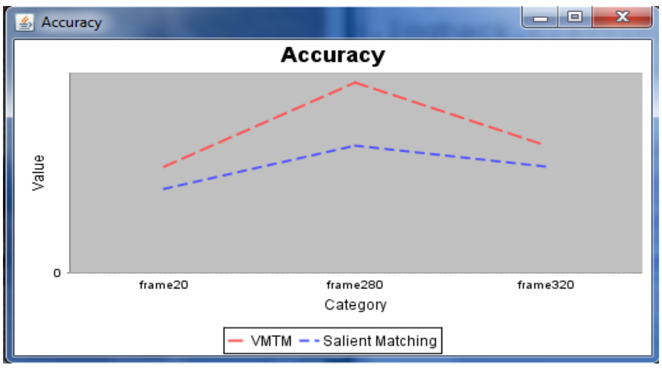


Fig. 2. Result

B. Accuracy Formula and graph:

TP (True positive): ROI object recognized
TN (False Positive): ROI object not recognized
Accuracy = (TP+TN/Total frames)



V. Conclusion And Future Scope

In this paper we have proposed the system for object detection and motion based tracking of moving objects. This work presented a computationally simple, robust tracking method for general unconstrained scenes using a set of simple trackers with appropriate data association. We have divided our system into three parts i.e. first is object detection, second object tracking, third occlusion detection and object tracking. The object detection is done through manually selection of ROI. We have used the Local Steering Kernel method for object tracking for matching the object information from the current frame is compared with the information from the next frame. If match then object is detected and the new position and values of the object are put on the stack. Here our proposed work overcomes the previous limitations. This will detect object in case of partial and full occlusion. Even if object hidden under any other object the VMTM method can detect the object when it again comes in frame. Experimental results showed the effectiveness of our proposed method in object tracking in appearance, affine transformations, and partial and full occlusion. Our results shows that proposed approach gives more accuracy in object tracking as compared to previous existing methods. Videos of any file formats and images of any file formats can be used as input to the system. Tracking is also possible in case of High Definition (HD) videos frames. As the accuracy increases the speed of tracking decreases. The reduction of time for tracking is open for future work. Tracking of fast object can be done by expanding the search region.

Acknowledgment

I express great many thanks to Prof. Kabra R.R. for her great effort of supervising and leading me, to accomplish this fine work. Also to college and department staff, they were a great source of support and encouragement. To my friends and family, for their warm, kind encourages and loves. To every person gave us something too light my pathway, I thanks for believing in me.

REFERENCES

[1] " Object tracking: A survey." , Acm Computing Surveys (CSUR), 38(4):13, 2006.

[2] COMANICIU, D., RAMESH, V., AND MEER, P. " Kernel-based object tracking ", IEEE Trans. on Pattern Analysis and Machine Intelligence
25, 5, 564577.

[3] YU, S. X. AND SHI, J. 2004. " Segmentation given partial grouping constraints." IEEE Trans. Patt. Analy. Mach. Intell. 26, 2, 173183.

[4] "Visual Object Tracking Based On Local Steering Kernels And Color Histogram", Olga Zoidi, Anastasios Tefas, *Member, IEEE,* and Ioannis Pitas, *Fellow,* IEEE, IEEE Transactions on Circuits and Systems for Video Technology, vol:25, no:3, Year 2013

[5] Rupali S.Rakibe, Bharati D.Patil, " Background Subtraction Algorithm Based Human Motion Detection", International Journal of Scientific and Research Publications, May 2013

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

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

ISSN: 2278 – 1323

[8] Himani S. Parekh1, Darshak G. Thakore 2, Udesang K. Jaliya 3, "A Survey on Object Detection and Tracking Methods." International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 2, February 2014
[9] Enrique J. Fernandez-Sanchez, Javier Diaz and Eduardo Ros, " Back-ground Subtraction Based on Color and Depth Using Active Sensors." Sensors 2013, July13
[10] Kim, K.; Chalidabhongse, T.H.; Harwood, D.; Davis, L. " Real-time foreground background segmentation using codebook model." Real Time Imag. 2005, 11, 172185.
[11] Elgammal, A.,Duraiswami, R.,Harwood, D., ANDDAVIS, L. 2002. " Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. "Proceedings of IEEE 90, 7, 11511163.
[12] J.Joshan Athanesious, P.Suresh, " Systematic Survey on Object Tracking Methods in Video",International Journal of Advanced Research in Computer Engineering Technology (IJARCET) October 2012, 242-247.
[13] Ms. Swati P. Baviskar,Mr. Nitin S. Ujgare, " Kernel Based Object Tracking Using Mean Shift Method".

Ms. Bhawale Priyanka S. received the B.E. Degree in Information Technology and is pursuing M.E. degrees in Computer Engineering GHRCOE, Chas, Ahmednagar. She is working at Dr. N.J.P.I.T. College of Polytechnic.

Mr. Sachin Patil Department of Computer Engineering, GHRCOE, Wagholi, Pune, Savitribai Phule Pune University.

