

# Survey on Tracking Algorithms

Padmavathi S

CSE Department,  
Amrita School of Engineering,  
Coimbatore.

Divya S

CSE Department,  
Amrita School of Engineering  
Coimbatore .

**Abstract**—Tracking non-rigid bodies like hands and face are of great significance in Sign Language Recognition Systems. (SLR Systems). In the gesture and sign language video sequences, the movement of the hand and face tends to be rapid and involves a lot of occluding scenarios. Thus we understand that naïve color based tracking methodologies are clearly insufficient. To improve the performance of tracking system a predict-update framework may be employed but this requires careful initialization of parameters, since the tracker tends to drift and lose track of these objects in the required sequences. This paper provides the survey of the various methods that have been used for determining the trajectory of the objects.

**Keywords**—Tracking, Point Tracking, Kernel Tracking Silhouette Tracking, Contours, Sign Language (SL), Kalman Filter, Particle Filter.

## I. INTRODUCTION

The aim of the hand and face tracking is to generate the trajectory of the hand and face over time by locating its position in every frame. The tracker should be able to provide complete region of the non-rigid objects. The task of tracking these non-rigid objects across frames involves (1) detecting the objects (2) Associating the Objects across frames. The model selected to represent the non-rigid objects should take into consideration the movement of the objects and also the type of deformation the object can undergo.

These representation may restriction the type of movement the object can undergo. For non-rigid objects, kernel or silhouette based representation can be used. The Tracking Systems have been classified into three types based on the approach used for tracking [1].

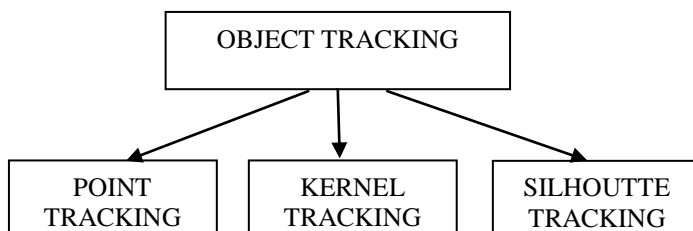


Fig 1 – Taxonomy of Tracking Systems

In point tracking objects are to be detected separately in consecutive frames and then the association algorithm is used to determine the trajectory of the object. Kernel tracking requires a model of the object shape and appearance of the object. The objects are basically tracked by computing the motion of the objects in the consecutive frames. The motion of the object is restricted to parametric transforms such as translation, rotation, scaling and affine transformations.

Silhouette Tracking is often used to track the contours of the non-rigid objects. This approach makes use of the information encoded in the object region to trace the presence of these objects in the consecutive frames. Given object models, the tracking is done either by (1) Shape Matching or (2) Contour Evolution. Both these methods can be assumed to be segmentation algorithms in the temporal domain. Section (II) provides a survey on the point tracking methodologies. Section (III) provides an insight into the various techniques available to perform Kernel Tracking and Section (IV) provides the techniques used in Silhouette Tracking. Section (V) elucidates the Summary and Conclusion.

## II. POINT TRACKING

Objects in point tracking are represented by points and the association of these points are based on the points present on the previous frame. There are a lot of algorithms that used for point tracking. Point tracking methodologies are basically used for objects that can be simply taken as points. Multiple points are used to represent large objects. The association of points may become very complicated in case of occlusions. Point tracking does not take into account entries and exits of the objects in the field of view. The point tracking also becomes complex when there a lot of misdetections.

Point tracking is of two types based on the approach used for association of points (1) Statistical Correspondence (2) Deterministic Correspondence. The association of points is very complex and a lot of constraints on the motion of the object have been taken into consideration to minimize the cost of association. The proximity constraint for example states that the object does not move far off from its location in the previous frame, the uniform acceleration constraint takes into consideration that the object to be tracked has zero acceleration, etc. In a method proposed by Sethi and Jain [2], the point tracking was done by associating the points using greedy approach. To minimize the cost, the method takes into account the rigidity and proximity constraint. This method failed to handle misdetections, entries and exits of objects. In [3] Sethi and Salari modified the method by first associating correspondences for the detected points and then adding hypothetical points to identify missing objects. In a method proposed by Rangarajan in Shah [4], the optical flow method was used to initialize the correspondences between points. When there is a considerable number of points in the field of view decreases, system assumes that the occlusions have occurred. Occlusions are handled by associating correspondences for the detected points in the current frame. For the remaining objects, position is predicted based on the constant velocity assumption. However the method fails to recognize the entries and exit of objects. The method proposed by Intille et al [5], the method used in [4] was modified. The

objects to be tracked were detected using background subtraction method. By examining specific regions in the field of view, the method is capable of identifying the entries and exits of objects.

#### A. Kalman Filter [6]

A statistical approach to point tracking is done using the method of Kalman Filter based tracking. It is an optimal estimator of the objects position across frames and the Kalman filter works on the concept of recursion. By principle the Kalman Filter [7] is used to predict the state of a linear system whose state is assumed to be Gaussian distributed. The Kalman Filter is composed of two steps (1) Prediction (2) Correction Step. The prediction step uses the state model to predict the next state of the system [1]

$$\bar{X}^t = DX^{t-1} + W \quad \text{Eq. (1)}$$

$$\bar{\Sigma}^t = D\Sigma^{t-1}D^T + Q^t \quad \text{Eq. (2)}$$

$\bar{X}^t$  and  $\bar{\Sigma}^t$  are the state and covariance prediction at time t. D represents the state transition matrix which define the relationship between the state variable at time t and t-1. The matrix Q is the covariance of the white noise W.

The correction step uses the actual observations of the object's current position  $Z^t$  to update the state of the object.

$$K^t = \bar{\Sigma}^t M^t [M\bar{\Sigma}^t M^t + R^t]^{-1} \quad \text{Eq. (3)}$$

$$X^t = \bar{X}^t + K^t [Z^t - M\bar{X}^t] \quad \text{Eq. (4)}$$

$$\Sigma^t = \bar{\Sigma}^t - K^t M \bar{\Sigma}^t \quad \text{Eq. (5)}$$

K is the Kalman Gain and Eq. (17) is the Ricatti equation used for the propagation of the state models.  $X^t$  is the update state. In Eq. (18), the term  $[Z^t - M\bar{X}^t]$  is called the innovation.

The object whose trajectory is to be determined is capable of movement in the 2 dimensional space and hence each region at the (x, y) position with velocities ( $v_x, v_y$ ) follow the equations:

$$x^p = x + v_x t + \frac{1}{2} a t^2 \quad \text{Eq. (6)}$$

$$y^p = y + v_y t + \frac{1}{2} a t^2 \quad \text{Eq. (7)}$$

$$v_x = v_{x|t-1} + a(t-1) \quad \text{Eq. (8)}$$

$$v_y = v_{y|t-1} + a(t-1) \quad \text{Eq. (9)}$$

$x^p$  and  $y^p$  represent the change in the positions in the X and Y directions respectively. The terms  $v_{x|t-1}$  and  $v_{y|t-1}$  represent the 'x' and 'y' velocities at time (t-1). The regions are represented by the state vector which holds information about the x position, y position, horizontal and vertical velocities.  $v_x$  &  $v_y$  represent the velocities in the x and y direction at time 't'.

The state vector  $\hat{x}$  is represented as follows

$$\hat{x} = (x^p, y^p, v_x, v_y) \quad \text{Eq. (10)}$$

Since we use a only a four dimensional vector for representing the state, the transition matrix can be simply given by

$$D = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. (11)}$$

The major requirement for Kalman Filter to work properly is the assumption that the states of system are required to be Gaussian distributed. The Filter gives wrong estimations for state variables that aren't Gaussian distributed.

#### B. Particle Filter

In order to model accurately the dynamics of a moving object it necessary to take into account the elements of non-linearity and also elements which are non-Gaussian [9], we use the particle filter [8]. Particle Filter also works on the predict-update framework with an inclusion of another method called select. The filter does not require the state variables to be Gaussian Distributed. By principle, the particle approximates the posterior distribution by set a weighted particles. The particles are weighted based on the likelihood score and then propagates these particles based on the motion model. For the state estimation, the Particle Filter assumes to be a Markov Model. By principle the Markov Model states that past and future states are independent of each other. Hence the particle filter estimates are only based on the present state. Particle Filter makes use of weights to define the importance of every sample. The most common sampling scheme used for particle filters (1) Selection (2) Prediction (3) Correction. [1]

#### C. Joint Probability Density Association (JPDA)

There are several data association methods available today [10]. The easiest approach is the nearest neighbor approach. The JPDA is an extension of the Probability Data Association (PDA) algorithm for multiple targets. [11] Rasmussen and Hager [12] used JPDA algorithm to track regions. A track is assumed to sequence of measurements that arises from an objects. Considering that the object at time "t" gives rise to "m" measurements. The JPDA associates all these measurements to the tracks. The weighted innovation used in JPDA can be plugged into innovation step in the Kalman Filter. [1] In the experiments results of [13], it can be understood that JPDA is reliable and also can be modified to be characterized as low complexity. JPDA can handle partial occlusions. The major limitation of JPDA is its inability to handle entries of new objects into the field of view and the

exits of already tracked objects from the field of view. Since JPDA does the association based on the number of objects tracked during the threshold number of frames, a change in the number of the objects in the field of view will cause serious errors.

#### D. Multiple Hypothesis Tracking (MHT)

MHT is an iterative algorithm that not only takes into consideration the first few frames to determine the object hypothesis but a specific set of frames to learn the object hypothesis. Each hypothesis is a set of tracks of disjoint, for which, a prediction of the next state is made. Hence the MHT overcomes the issue posed by JPDA, in that MHT does not make hypothesis based on the first few frames. The MHT Algorithm can clearly handle the entries and exits of the objects from the field of view. [1] MHT is intensively computational algorithm and hence the Probabilistic MHT (PMHT) [14] was a better option.

Point Trackers are evaluated based on whether they generate the right trajectories. To solve the issues of high computational costs in deterministic approach, a combination of motion based constraints are taken into consideration. Statistical Approaches solve the issue by modelling uncertainties and taking them into consideration.

### III. KERNEL TRACKING

Kernel Tracking is the most commonly used category of technique when the object in motion follows the parametric motion model. The algorithms differ in terms of the appearance of the object to be tracked, the number of objects to be tracked and most significantly the motion model used to estimate the next state of the object along the path. Kernel based tracking methodologies can be further categorized into (1) Appearance Based Tracked (2) Multi-View Models.

#### A. Simple Template Matching

One of the most common approach in the appearance based methods is the simple template matching techniques. Most of the template matching techniques involves the search of given template in the frame to obtain the trajectory of the object. The search can be a intensively exhaustive brute force search [15] or it can be optimized by optimization algorithms [16]. In addition to the use of mere representations of the objects as templates, the histogram or mixture model of the template region can be computed and used for tracking. To reduce the computational complexity, the search can be limited to the nearest 8 or 4 neighbors depending on the type of motion the object follows. In [17] Fieguth and Terzopoulos determined the average intensity of the region to track the object in the successive frames and also to reduce the complexity they limited the search to the nearest 8 neighborhood.

#### B. MeanShift Procedure

MeanShift Algorithm is one of the most commonly used approaches to tracking whose similarity is verified using Histograms. Instead of just performing a brute force search, a mean shift procedure is applied.

An algorithm that iteratively shifts a data point to its average point in the neighborhood. Considering a set  $S$  with 'n' data points  $x_i$  in 'd-D' Euclidean space,  $X$ . [18][19]

$K(x)$  denotes the kernel function and also indicates how much the 'x' contributes to the estimate of the mean. Therefore the sample mean, 'm', at 'x' with the Kernel 'K' is given by

$$m(x) = \frac{\sum_{i=1}^n K((x-x_i)x_i)}{\sum_{i=1}^n K(x-x_i)} \quad \text{Eq. (12)}$$

The difference  $m(x)-x$  yields the mean shift. Mean Shift Algorithms iteratively moves the data point to its mean.

The colors of a region of interest are held by a histogram and the tracking is done by just comparing the histograms of the regions (i.e.) to find the regions with the most similar color distribution. Ideally it is essential to know if the pixel intensity is from the region of interest or not, so the computations of the likelihood maps are a must. The likelihood maps are computed based on (1) Color (2) Texture (3) Shape (4) Predicted Location. Histogram Similarity can be obtained from the Bhattacharya Coefficient.

$$\sum_{u=1}^b P(u)Q(u) \quad \text{Eq. (13)}$$

In the above equation, 'b' represents the number of bins. At each iteration, the mean shift vector is computed and the similarity measure is increased. The process is repeated until convergence is achieved [1]. One of the biggest advantages of Mean Shift Tracker over the naive template matching approach is the elimination of the brute force.

#### C. CamShift

CAMShift – Continuously Adaptive Mean Shift developed on the basis of Mean Shift Algorithm. It encapsulates Mean Shift in a loop by continuously varying the window size until convergence is achieved. At each iteration, the mean shift procedure is applied with window of the given size. MeanShift Algorithm worked well on statically changing probability distributions and failed on dynamically changing probability distributions. This tracker doesn't handle entries of new objects if the new object entry is similar to one that is being tracked. For better results the proposed will require the frames to be converted to HSV space to compute the histogram. The window size is automatically adjusted to the scaling movement of the object with computation of the zeroth moment. [20].

#### D. Kanas Lucade Tracker (KLT)

In this method the object tracking is done with the help of feature points. The major constraint of the KLT system is that it requires the system to hold on the constant illumination condition. The KLT tracker also requires that the feature points to be tracked lie on the same surface and also there feature should show no sudden acceleration. [1] Feature points used for the purpose of tracking can be obtained by a variety of methods. The most commonly used are the Harris Corner Detector, SURF Features, Eigen Features, etc. [21] Once the location of the feature point is obtained, the KLT then evaluates the correctness of the tracked path by computing the affine transformation between the corresponding patches in the consecutive frames. The method is robust to occlusion,

highly accurate and also time efficient. This also handles effectively the entry and exit of objects from the field of view. Through optical flow estimation, motion parameters of the moving objects can be obtained and at the same time, the phenomena of occlusion of objects can also be avoided [22]. Though it provides high accuracy for tracking, the tracking of multiple objects under natural circumstances is complex task to be implemented KLT.

The use of primitive geometric shapes for the representation of objects is quite common. The tracking complexity can be minimized by analyzing the characteristics of the object's motion. One of the limitation while using primitive Euclidean shapes for representations is that the object's region of interest might not totally lie within the enclosed Euclidean space and hence there is a possibility of the background pixels residing inside the bounded region. The phenomenon is observed both in case of tracking rigid and non-rigid bodies. To overcome this limitation, one way is to force the kernel to reside inside the object region rather than having background pixels inside the kernel. [1]

#### IV. SILHOUETTE TRACKING

Silhouette tracking approaches are basically used for non-rigid object tracking like hands, face, etc. These objects tend to have a complex shape and structure. The goal of the silhouette tracker is identify the object region in the successive frame by computing the model obtained from the previous frames. Silhouette trackers are categorized based on the approaches used into (1) Shape Matching (2) Contour Evolution.

##### A. Shape Matching

The technique is similar to the template matching technique described in Section (III). The shape and its associated model are exhaustively searched to find a match in the current frame. In the method [23] proposed by Huttenlocher et al. performed shape matching using an edge-based representation. They used the Hausdroff distance metric to construct the correlation surface from which the minimum is selected as the new object position. Hausdroff metric is a mathematical measure used to determine the least similar members.

Let  $A$  and  $B$  represent two sets of points.

$$A = \{a_1, a_2, a_3 \dots a_n\}$$

$$B = \{b_1, b_2, b_3 \dots b_n\}$$

$$H(A, B) = \max\{h(A, B), h(b, A)\} \quad \text{Eq. (14)}$$

Where  $h(A, B) = \sup_{a \in A} \inf_{b \in B} \|a - b\|$  and  $\|\cdot\|$  is the norm of choice [1].

##### B. Tracking by Direct Minimization of Contour Energy

In the context of contour evolution, there is an analogy between the segmentation methods and the contour evolution techniques. Both the segmentation and tracking methods minimize the energy functional either by greedy methods or by gradient descent. The contour energy is defined in terms of temporal information in the form of either the temporal

gradient (optical flow). Silhouette tracking is employed when tracking of the complete region of an object is required. The most important advantage of tracking silhouettes is their flexibility to handle a large variety of object shapes. Silhouettes can be represented in different ways. The most common silhouette representation is in the form of a binary indicator function, which marks the object region by ones and the non-object regions by zeros. For Contour-based methods, the silhouette is represented either explicitly or implicitly. Tracking of a complete object can be obtained by employing active contours [23], which were introduced by Kass et al [24]. The objective of these active contours is to obtain a tight contour around the object of interest by minimizing the energy function.

$$E(T) = \int_0^1 E_{internal}(v) + E_{image}(v) \cdot ds$$

In the above function's' represents the arc length.  $E_{image}$  energy based on the image observations and  $E_{internal}$  prevents gaps and rapid bending. In [23], it was stated that in most of the tracking systems, during occlusions either the centroids of the objects in occlusion have to be predicted during the occlusion time interval or the occlusion scenarios should be left as such. [25][26]. But the proposed method by Yilmaz et al [23] handles the occluding scenarios.

The representations of the Silhouette trackers can be motion models, appearance models and shape based models or even a combination of the three. These category of trackers do not handle occlusion scenarios explicitly, hence it is necessary to taken into considerations like motion constraints (i.e.) uniform velocity, uniform acceleration, etc. Another important advantage of Silhouette trackers is that they can implicitly handle topology changes like region split and merge. [1].

#### V. CONCLUSION

This paper has provided a survey of the various techniques available for tracking. To determine the best tracker for an object, it essential to analyze the motion characteristics of the object. Based on these obtained characteristics and motion dynamics, the appropriate model and tracker can be determined. Table -1 tabulates the characteristics of techniques outlined in the paper. (A- Affine Motion , T- Translation , S- Scaling, Partial – Partial Occlusion Handled )

Table -1

Name	Motion Type	Occlusion	Optimal
Kalman Filter	A	No	Yes
Particle Filter	A	Yes	Yes
MeanShift	T	Partial	No
CAMShift	S+T	No	No
Shape Matching	T	No	No
Tracking by minimization of Contour Energy - [23]	A	Yes	No
Simple Template Matching	T	Partial	No
KLT	A	Yes	Yes

Table -1 Characteristics of the types of Tracking Systems

## REFERENCES

- [1] Yilmaz, A., Javed, O., and Shah, M. 2006. "Object tracking: A survey". ACM Computing. Survey. 38, 4, Article 13 (Dec. 2006), 45 pages. DOI =10.1145/1177352.1177355 <http://doi.acm.org/10.1145/1177352.1177355>
- [2] Sethi, I. and Jain, R. 1987. "Finding trajectories of feature points in a monocular image sequence". IEEE Trans. Patt. Anal. Mach. Intell. 9, 1, 56-73.
- [3] Salari, V. and Sethi, I. K. 1990. Feature point correspondence in the presence of occlusion. IEEE Trans. Patt. Anal. Mach. Intell. 12, 1, 87-91.
- [4] Rangarajan, K. & Shah 1991. "Establishing motion correspondence". Conference Vision Graphics Image Process 54, 1, 56-73.
- [5] Intille, S., Davis, J., & Bobick A. 1997. "Real-time closed-world tracking". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 697-703.
- [6] R.E Kalman, "A new approach to linear filtering and prediction problems", Transactions of the ASME, Ser. D., Journal of Basic Engineering, 82, 34-45. (1960).
- [7] Welch, G., Bishop, G.: "An Introduction to the Kalman Filter", ACM SIGGRAPH 2001, Course 8, available at <http://www.cs.unc.edu/~welch/Kalman/> (2001).
- [8] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174-188, Feb. 2002.
- [9] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings F*, vol. 140, no. 2, pp. 107-113, 1993.
- [10] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," *Proc. of the IEEE*, vol. 92, no. 3, pp. 536-557, 2004.
- [11] Y. Bar-Shalom and X.R. Li, "Estimation and Tracking: Principles, Techniques and Software", Artech House, Boston, MA, 1993.
- [12] Rasmussen, C. and Hager, G. 2001. Probabilistic data association methods for tracking complex visual objects. IEEE Trans. Patt. Anal. Mach. Intell. 23, 6, 560-576.
- [13] M. Jaward, L. Mihaylova, N. Canagarajah and D. Bull, "Multiple Object Tracking Using Particle Filters", *IEEEAC paper # 1280*.
- [14] Streit, R. L. and Luginbuhl, T. E. 1994. Maximum likelihood method for probabilistic multi-hypothesis tracking. In Proceedings of the International Society for Optical Engineering (SPIE.) vol. 2235. 394-405.
- [15] H.T. Nguyen, M. Worring, and R. van den Boomgaard. "Occlusion robust adaptive template tracking." In ICCV01, pages I: 678-683, 2001.
- [16] M.J. Black and A.D. Jepson. "Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation", January 1998.
- [17] Fieguth, P. and Terzopoulos, D. 1997. "Color-based tracking of heads and other mobile objects at video frame rates". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 21-27.
- [18] D. Comaniciu, V. Ramesh, and P. Meer. Mean shift: A robust approach towards feature space analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(5):603-619, 2002.
- [19] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In IEEE Proc. on Computer Vision and Pattern Recognition, pages 673-678, 2000.
- [20] Bradski, G. and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media Inc.: Sebastopol, CA, 2008.
- [21] Carlo Tomasi and Takeo Kanade. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.
- [22] Shahzad Malik, "Real-time Hand Tracking and Finger Tracking for Interaction", CSC2503F Project Report, [www.cs.toronto.edu](http://www.cs.toronto.edu), Paper for Vision Course. University of Toronto, December 2003.
- I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.
- [23] Alper Yilmaz, Member, IEEE, Xin Li, and Mubarak Shah, Fellow, IEEE "Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 11, NOVEMBER 2004
- [24] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision*, vol. 1, 1988.
- [25] N. Paragios and R. Deriche, "Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 3, pp. 266-280, 2000.
- [26] A. Mansouri, "Region Tracking via Level Set PDEs Without Motion Computation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 947-961, July 2002.