International Journal of Electronics and Communication Engineering (IJECE)
ISSN 2278-9901
Vol. 2, Issue 1, Feb 2013, 59-68

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

A LITERATURE SURVEY ON OBJECT TRACKING

SHUBHAM SRIVASTAVA & MS. PRATIBHA SINGH

M.E. (Digital Instrumentation) Institute of Engineering and Technology D.A.V.V Indore, Madhya Pradesh, India

ABSTRACT

The goal of this article is to review the state of the art of tracking methods, classify them into different categories, and identify new useful tracking methods . Many difficulties arise in object tracking due to camera motion, occlusions, non rigid object structures, abrupt changes in the appearance patters of both the object and the scene, therefore object tracking is a challenging problem. In this paper we present different object representations, their detection and categorize different tracking methods on the basis of the object and motion representation used.

KEYWORDS: Contour, Motion Detection, Object Detection, Object Tracking, Shape Features

INTRODUCTION

In its form of tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words a tracker assigns consistent labels to the tracked objects in different frames of a video [1]. Tracking objects can be complex due to:

- loss of information caused by projection of the 3D world on a 2D image,
- noise in images,
- non rigid object structures,
- complex object motion,
- camera motion tracking.

The use of object tracking is pertinent in the task of:

- Motion based recognition,
- Automated surveillance,
- Video indexing,
- Human-computer interaction,
- Traffic monitoring,
- Vehicle navigation.

Object tracking is a process of scanning an image for an object of interest like people, faces, computers, robots or any object.

In this paper, we first describe various object representations, followed by feature selection used for object detection and finally categorize the different methods used for tracking, different methods used for object detection and finally categorize the different tracking methods on the basis of object and motion representation.

OBJECT REPRESENTATION

Objects can be represented by their shapes and appearances. In a tracking scenario, an object can be defined as anything that is of interest for further analysis. In this section we will describe the object shape representations commonly employed for tracking. [35]

Points: The object is represented by a point, or by a set of points. Point representation is normally used for the objects which occupy very small regions in the image.

Figure 1: Object Representation of Multiple Points

Primitive Geometric Shapes: In this case object shape is represented by a rectangle, ellipse. [26]

Figure 2: Object Representation of Rectangular Patch

Object Silhouette: Contour representation defines the boundary of an object. The region inside the contour is called the silhouette of the object. This representation is normally used for tracking complex non rigid shapes [35].

Figure 3: Representation of Object Silhouette

Articulated Shape Models: Articulated objects are composed of body parts that are held together with joints. For example, the human body is an articulated object with legs, hands, head and feet connected by joints.

Figure 4: Representation of Part Based Multiple Patches

Skeletal Models: Object skeleton can be extracted by applying medial axis transform to the object silhouette [2],[6].

Figure 5: Representation of Object Skeleton

FEATURE SELECTION

Selection of right feature is a very important task during object tracking. Feature selection is closely related to object representation. For example colour is used as a feature for histogram based representations, whereas for contour-based representation [35], object edges are usually used as features. The details of common visual features are as follows:

Colour: In image processing, the RGB(red,green,blue) colour space is used to represent colour. However RGB space is not a uniform colour space. These colour spaces are however sensitive to noise [34].

Edges: Edge detection is used mainly to identify the strong changes in the image intensities that are caused by the object boundaries. The most important property of edges is that they are less sensitive to illumination changes compared to colour features. For the tracking of the boundary of the objects many algorithms use edges as a representative feature. Canny edge detector is the most important edge detection technique because of its accuracy and simplicity [4].

Optical Flow: Optical flow is a dense field of displacement vectors which defines the translation of each pixel in a region. Optical flow is commonly used as a feature in motion based segmentation and tracking applications.[8]

Texture: Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. [31]

Mostly features are chosen manually by the user depending on the application domain.

OBJECT DETECTION

Every tracking method requires object detection mechanism. A common approach for object detection is to use information in a single frame, whereas some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detection.This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames.

Some of the common object detection methods are as follows:

- Point Detectors
- Background subtraction
- Segmentation
- Supervised learning

Point Detectors: Point detectors are used to locate the points in images which have an expreesive feature in their respective localities. Interest points are being used in the context of motion, stereo, and tracking problems. A desireable

quality of an interest point is its invariance to changes in illumination and camera viewpoint. Interest points detectors include Moravec's interest operator, Harris interest point detectors, KLT detector and Tomasi and SIFT detectors.

To find the interest points, Moravec's operator computes the variation of the image intensities in a 4x4 patch in a horizontal, vertical, diagonal and anti diagonal directions and selects the minimum of the four variations as representative values for the window.

Harris detectors computes the interest points by computing the first order image derivatives,(Ix, Iy) in x and y directions to highlight the directional intensity variations, and a second moment matrix which encodes this variation is evaluated for each pixel in a small neighbourhood.

$$M = \begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix}$$

An interest point is identified using the determinant and the trace M which measures the variation in a local neighbourhood $R=\det(M)-k.tr(M)^2$, where k is a constant. The moment matrix is used in the interest point detection step of the KLT tracking method. Interest point confidence, R, is computed using the minimum eigenvalue of M, λ_{min}. Quantatively both harris and KLT emphasize the intensity variations using very similar measures. For instance, R in Harris is related to the characterize polynomial used for finding the eigenvalues of M: $\lambda^2 + \det(M)-\lambda.tr(M)=0$, while KLT computes the eigenvalues directly. In practice both methods find the same interest points, the only difference is the additional KLT criterion that enforces a predefined spatial distance between detected interest points.

The matrix M is invariant to both the translation and rotation, but is not invariant to projective transformation. In order to introduce robust detection of interest points under transformations SIFT method is introduced [35] .

Background Subtraction: In background subtraction a representation of the scene called the background model is incoming frame. Any significant change in an image region from the background model signifies the moving object. The regions of the pixels undergoing a change are marked for further processing [30].

A connected component algorithm is applied to obtain connected regions corresponding to the objects. This process is reffered to as Background Subtraction. The model parameters, the mean $\mu(x,y)$ and the covariance $\sum(x,y)$, are learned from the colour observation in several consecutive frames. Once the background model is derived for every pixel (x,y) in the input frame, the likelihood of its colour coming from $N(\mu(x,y),\sum(x,y))$ is computed and the pixels that deviate from the background model by comparing it with every Gaussian in the model until a matching Gaussian is found. If a match is found, the mean and variance of the matched Gaussian is updated, otherwise a new Gaussian with amean equal to the current pixel colour and some initial variance is introduced into a mixture. Each pixel is classified based on whether the matched distribution represents the background process.

Segmentation: The aim of image segmentation algorithms is to partition the image into perceptually similar regions. Every segmentation algorithm addresses two problems, the criteria for good partitioning and the method for achieving efficient partitioning. The recent segmentation techniques that are relevant to object tracking are as follows:

- Mean shift clustering
- Image segmentation using graph cuts
- Active contours

Supervised Learning: With the help of supervised learning mechanism object detection can be performed by learning different object views automatically from a set of examples. Given a set of learning examples supervised learning methods generate a function that maps input to desired outputs. Learning of different object views waves the requirement of storing a complete set of templates. A standard formulation of supervised learning classification problem where the learner approximates the behavior of a function by generating an output in form of either a continouous value, which is called regression, or a class label which is called classification. In the context of object detection the learning examples are composed of object features and an associate object class where both of these quantities are manually defined. Supervised learning methods require a large collection of samples from each object class. It has been found that starting from a set of labelled data with two sets of statistically independent features, contraning has been used to reduce the amount of manual interaction required for training in the context of adaboost and support vector machines [14].

OBJECT TRACKING

The goal of object tracking is to estimate the locations and motion parametres of a target in an image sequence given the initialized position in the first frame. Research in tracking plays a key role in understanding motion and structure of objects. It finds numerous applications including surveillance, human computer interaction, traffic pattern analysis recognition, medical image processing. Since there exists no single tracking method that can be successfully applied to all tasks and situations. A typical tracking system consists of three components:

- Object representation
- Dynamic model
- Search mechanism

Object tracking algorithms can be classified as either determin istic or stochastic based on their search mechanism. With the target of interest represnted in some feature space, object tracking can always be reduced to a search task and formulated as an optimization problem. Taxonomy of tracking methods is shown below:

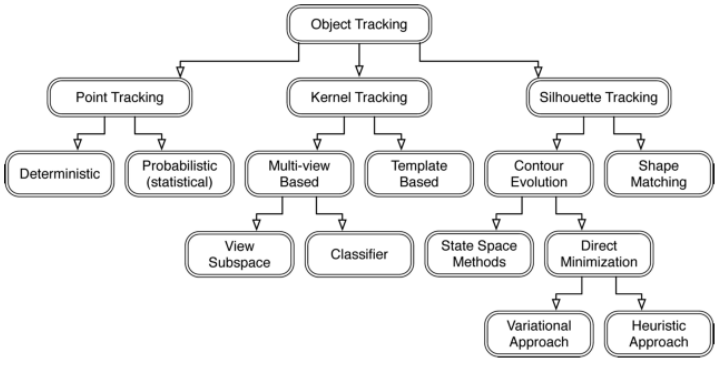

Figure 6: Taxonomy of Tracking Methods

Now not going much into the details we will discuss few lines about three main methods of object tracking and they are

- Point Tracking
- Kernel Tracking
- Silhouette Tracking

Point Tracking: In point tracking the objects which are detected in consecutive frames are represented in points and the points association is based on the previous state which can include object position and motion. This approach requires an external mechanism to detect the objects in every frame. Overall point correspondence methods can be divided into two broad categories, they are:

- Deterministic Method
- Statistical Method

Deterministic Method: The deterministic methods use qualitative motion heuristics to constrain the correspondence problem. This method defines a cost of associating each object in frame t-1 to a single object in frame t using a set of motion constraints. Minimization of the correspondence cost is formulated as a combinatorial optimization problem [33]. A solution to the above problem can be given by Hungarian algorithm, which consists of one-to-one correspondences among all possible associations and can be obtained by optimal assignment methods. The correspondence cost is usually defined by using a combination of the following constraints:

- Proximity
- Maximum velocity
- Small velocity change
- Common motion
- Rigidity
- Proximal uniformity

Statistical Method: Measurements obtained from sensors invariably contain noise. The tracking problems can be solved by taking the measurements and the model uncertainities into account during object state estimation. The statistical correspondence methods use the state space approach to model the object properties such as position, velocity, and acceleration[32]. Mesurements generally consist of the object position in the image, which can be obtained by a detection mechanism. In cases where the measurements arise due to the presence of a single object in the scene measurements need to be associated with the corresponding object states. The two steps are:

- Single object state estimation
- Multiobject data association and state estimation

Kernel Tracking: The word kernel refers to the object shape and appearance. Kernel tracking is performed by computing the motion of the object which is represented by a primitive object region, from one frame to the next. These algorithms differs in terms of the appearance representation used, the number of objects tracked and the method used to estimate the object motion [27]. A rectangular template or an elliptical shape can be an example of the kernel. The motion in the kernel tracking is in the form of parametric transformation such as translation, affine and rotation. We divide these tracking methods into two sub categories, they are:

- Templates and density based appearance models
- Multiview appearance models

Silhouette Tracking: In its form, tracking is done by estimating the object region in each frame. This form of tracking makes the use of the information encoded inside the object region. This information can be in the form of density and shape models which are usually in the form of edge maps. Silhouette tracking based methods provide an accurate shape description for objects having complex shapes.The goal of a silhouette based object tracker is to find the object region in each frame by means of an object model generated using the previous frames. The goal of a silhouette based object tracker is to find the object region in each frame by means of an object model generated using the previous frames. This model can be in the form of a colour histogram object edges or the object contour. Silhouette tracking is divided into two categories:

- Shape Matching
- Contour Tracking

Shape Matching: Shape matching makes the use of current frame in the search of object silhouette. Shape matching can be done for object tracking on the basis of template matching. The search can be performed by computing the similarity of the object with the model generated from the hypothesized object silhouette based on previous frame. In this approach non rigid object motion is not handled [29].

Contour Tracking: Contour tracking approaches on the other hand evolve an initial contour to its new position in the current frame by either using the state space models or direct minimization of some energy functional. This contour evolution requires that some part of the object in the current frame overlap with the object region in the previous frame. Tracking by evolving a contour can be performed using two different approaches. The first approach uses state space models to model the contour shape and motion. The second approach directly evolves the contour by minimization the contour energy using direct minimization techniques such as gradient descent.

RESULTS AND DISCUSSIONS

Point Tracking

Table 1

Qualitative Comparison of Point Trackers (**#**: number of objects, **M**: multiple objects, **S** single object. Symbols √ and × denote whether the tracker can or cannot handle occlusions, object entries object exits, and provide the optimal solution.)

	#	Entry	Exit	Occlusion	Optimal
GE [Sethi and Jain 1987]	M	×	×	×	×
MGE [Salari and Sethi 1990]	M	√	√	√	×
GOA [Veenman et al. 2001]	M	×	×	√	√
MFT [Shafique and Shah 2003]	M	√	√	√	×
Kalman [Bar-Shalom and Foreman 1988]	S	×	×	×	√
JPDAF [Bar-Shalom and Foreman 1988]	M	×	×	×	×
MHT [Cox and Hingorani 1996]	M	√	√	√	√

Point tracking methods can be evaluated on the basis whether they generate correct point trajectories [35]. The performance can be evaluated by computing precision and recall measures. Precision and recall measures can be defined as:

$$\text{Precision} = \frac{\#\text{of correct correspondences}}{\#\text{of established correspondences}}$$

$$\text{Recall} = \frac{\#\text{of correct correspondence}}{\#\text{of correct correspondences}}$$

Kernel Tracking

Table 2

Qualitative Comparison of Geometric Model-Based Trackers (Init. denotes initialization. **#**: number of objects, **M**: multiple objects, **S**: single object respectively, **A**: affine or homography, **T**: translational motion, **S**: scaling, **R**: rotation, **P**: partial occlusion, **F**: full occlusion. Symbols √ and × respectively denote if the tracker requires or does not require training or initialization.)

	#	Motion	Training	Occ.	Init.
Simple template matching	S	T	×	P	√
Mean-shift [Comaniciu et al. 2003]	S	T + S	×	P	√
KLT [Shi and Tomasi 1994]	S	A	×	P	√
Appearance Tracking [Jepson et al. 2003]	S	T + S + R	×	P	√
Layering [Tao et al. 2002]	M	T + S + R	×	F	×
Bramble [Isard and MacCormick 2001]	M	T + S + R	√	F	×
EigenTracker [Black and Jepson 1998]	S	A	√	P	√
SVM [Avidan 2001]	S	T	√	P	√

In this category the main goal of the trackers is to estimate the object motion. In the case of analyzing the object behaviour based on the object trajectory, only the motion is adequate. In order to analyze the performance of the trackers in this category, one can define measures based on what is expected to provide only object motion, the evaluation can be performed by computing a distance measure between the estimated and actual motion parameters [35]. An example of a distance measure can be the angular distance, d=**A.B/|A||B|,** between the motion vectors, A and B. For applications when the tracker is required to provide the correct object region in addition to its trajectory, the tracker performance can be evaluated the precision and recall measures. Precision is the ratio of the intersection to the hypothesized and correct object region, whereas the recall is the ratio of the intersection to the ground.

Silhouette Tracking

Table 3

Qualitative Comparison of Silhouette Trackers (Occ. denotes occlusion handling and Trn. denotes training. **#**: number of objects, **S**: single, **M**: multiple, **P**: partial, **F**: full. Symbols √ and × denote whether the tracker can or cannot handle occlusions, and requires or does not require training.)

	#	Occ.	Trn.	Features	Technique
Shape Matching					
[Huttenlocher et al. 1993]	S	×	×	Edge template	Template matching
[Li et al. 2001]	S	×	√	Edge template	Template matching
[Kang et al. 2004]	S	×	×	Color histogram	Histogram matching
[Sato and Aggarwal 2004]	S	√	×	Silhouette	Hough transform
Contour Evolution using State Space Models					
[Terzopoulos and Szeliski 1992]	S	×	√	Gradient mag.	Kalman filtering
[Isard and Blake 1998]	S	×	√	Gradient mag.	Particle filtering
[MacCormick and Blake 2000]	M	F	√	Gradient mag.	Particle filtering
[Chen et al. 2001]	S	×	√	Gradient mag.	JPDAF
Contour Evolution by Direct Minimization					
[Bertalmio et al. 2000]	S	×	×	Temporal gradient	Gradient descent
[Mansouri 2002]	S	×	×	Temporal gradient	Gradient descent
[Paragios and Deriche 2002]	S	×	×	Temporal gradient	Gradient descent
[Cremers et al. 2002]	S	P	√	Region statistics	Gradient descent
[Yilmaz et al. 2004]	M	F	×	Region statistics	Gradient descent

The silhouette object trackers choose the representations in the form of motion models, appearance models or a combination of these. Object appearances are modelled by parametric or non parametric density functions such as mixture of Gaussians or histograms. Silhouette tracking is used when tracking of the complete region of the object is required [35]. For region tracking, the precision and recall measures are defined in terms of the intersection of the hypothesized and correct object regions. The precision is the ratio of the hypothesized region and recall is the ratio of the intersection to the ground truth. Some algorithms only use the information about the silhouette boundary for tracking, while others use the complete region inside the silhouette. The main advantage of the silhouette tracking is their flexibility to handle a large variety of object shapes.

FUTURE WORK

During the last few years a significant progress has been made in the field of object tracking. Many robust trackers have been developed which can track objects in real time in simple scenarios. Thus, tracking and associated problems of feature selection, object representation, dynamic shape, and motion estimation are very active areas of research and new solutions are continuously being proposed. The main challenge in tracking is to develop algorithms for

tracking objects in unconstrained videos. These videos usually contain multiple people in a small field of view. Therefore there is severe occlusion, and people are only partially visible. One solution to this problem is to employ audio in addition to video for object tracking. There are also some methods being developed for estimating the point of location of audio source, for example, a person's mouth, based on four or six microphones. This audio based localization of the speaker provides additional information which then can be used in conjunction with a video based tracker to solve problems like severe occlusions.

CONCLUSIONS

In this article, we present a survey on the object tracking methods; we divide the tracking methods into three methods on the basis of object representations. Recognizing the importance of object detection for tracking systems, we include a short discussion on various object detection methods. A detailed summary of object trackers, motion models and the parameter estimation schemes employed by the tracking algorithms has been provided. We hope that this article will give a valuable insight into the important research topic and encourage new research.

REFERENCES

1. Aggarwal, J.K. and Cal, Q. 1999. Human motion analysis: A review Comput. Vision Image Understand. 73, 3, 428-440.
2. Ali, A. And Aggarwal,J. 2001.Segmentation and recognition of continuous human activity. In IEEE Workshop on Detection and Recognition of Events in Video.28-35
3. Avidan, S. 2001. Support vector tracking. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)184-191.
4. Bowyer, K. Kranenburg, C., and Dougherty, S. 2001.Edge detector evaluation using empherical roc curve. Comput. Vision Image Understand. 10, 77-103.
5. Baddeley, A. 1992.Errors in binary images and an l version of the haus-dorff metric. Nieuw Archief voor Wiskunde 10, 157-183
6. Ballard, D. and Brown, C. 1982.Computer Vision. Prentice-Hall.
7. Bar-Shalom, Y. and Foreman, T. 1988. Tracking and Data Association. Academic Press Inc.
8. Barron, J., Fleet, D., and Beauchemin, S. 1994.Performance of optical flow techniques. Int. J. Comput. Vision 12, 43-77.
9. Bertalmio, M., Sapiro, G., and Randall. , G. 2000.Morphing active contours. IEEE Trans. Patt. Analy. Mach. Intell. 22, 7, 733-737.
10. Beymer, D. And Konolige, K.1999. Real-time tracking of multiple people using continuous detection.. In IEEE Conference on Computer Vision (ICCV) Frame -Rate Workshop.
11. Black, M. And Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. Comput. Vision Image Understand. 63, 1, 75-104.
12. Black, M. And Jepson, A. 1998. Eigentracking: Robust matching and tracking of articulated objects using a view based representation. Int.J Comput. Vision 26, 1, 63-84.
13. Blake, A. And Isard, M. 2000. Active Contours: The Application Of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion. Springer.
14. Blum, A. And Mitchell, T. 1998. Combining labelled and unlabelled data with co-training. In 11th Annual Conference on Computational Learning Theory. 92-100.
15. Blum, A. And Langley, P. 1997. Selection of relevant features and examples in machine learning. Artific. Intell. 97, 1-2, 245-271.

16. Boser, B., Guyon, I. M., and Vapnik, V. 1992. A training algorithm for optimal margin classifiers. In ACM workshop on Conference on Computational Learning Theory (COLT)142-152.
17. Bregler, C., Hertzmann, A., and Biermann, H. 2000. Recovering nonrigid 3d shape from image streams. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)690-696.
18. Broida, T. And Chellappa, R. 1986.Estimation of object motion parameters from noisy images. IEEE Trans. Patt. Analy. Mach. Intell. 8, 1, 90-99.
19. Cai,Q. and Aggarwal, J. 1999.Tracking human motion in structured environments using a distributed camera system. IEEE Trans. Patt. Analy. Mach. Intell. 2, 11, 1241-1247.
20. Canny, J. 1986.A computational approach to edge detection. IEEE Trans. Patt. Analy. Mach. Intell. 8, 6, 679-698.
21. Caselles, V., Kimmel, R., and Sapiro,G. 1995. Geodesic active contours. In IEEE International Conference on Computer Vision (ICCV) 694-699.
22. Cham,T. and Rehg, J.M. 1999. A multiple hypothesis approach to figure tracking. In IEEE International Conference on Computer Vision and Patterm Recognition . 239-245.
23. Chang, Y.l. and Aggarwal, J.K. 1991. 3D structure reconstruction from an ego motion sequence using statistical estimation and detection theory.In Workshop on Visual Motion. 268-273.
24. Chen, Y., Rui, Y., and Huang, T. 2001. Jpdaf based hmm for real-time contour tracking. In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) 543-550.
25. Collins ,R., Lipton. A., Fujiyoshi, H., and Kanade, T. 2001. Algorithms for cooperative multisensor surveillance. Proceedings of IEEE 89, 10, 1456-1477.
26. Comaniciu, D. 2002. Bayesian kernel tracking . In Annual Conference of the German Society for Pattern Recognition. 438-445.
27. Comaniciu, D., Ramesh ,V., and Meer, P. 2003. Kernel-based object tracking. IEEE Trans. Patt. Analy. Mach. Intell. 25, 564-575.
28. Comaniciu, D. and Meer, P. 1999. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Analy. Mach. Intell. 24, 5, 603-619.
29. Comanciu, D. and Meer, P. 2002. Mean shift: A robust approach toward feature space analysis. IEEE Trans. Patt. Analy. Mach. Intell. 24,5,603-619.
30. Elgammal, A., Harwood, D., and Davis, L. 2000. Non-parametric model for background subtraction. In European Conference on Computer Vision (ECCV) 751-767.
31. Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit,S., and Anderson,C. 1994. Overcomplete steerable pyramid filters and rotation invariance. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 222-228.
32. Isard, M. and Blake, A. 1998. Condensation-conditional density propagation for visual tracking, Int. J. Comput. Vision 29, 1, 5-28.
33. Sethi,I. and Jain, R. 1987. Finding trajectories of feature points in a monocular image sequence. IEEE Trans. Patt. Analy. Mach. Intell. 9,1,56-73.
34. Song, K.Y., Kittler, J., and Petrou,M. 1996. Defect detection in random color textures. Israel Verj. Cap.J. 14,9,667-683.
35. Yilmaz,A., Javed,O., and Shah,M.2006. Object tracking: A survey. ACM Comput Surv.38, 4, Article 13 (Dec.2006).