A Survey on 2d Object Tracking in Digital Video

A.S.Khobragade[a], C.G.Dethe[b], K.D.Kulat[c]

[a]Yeshwantrao Chavan College Of Engineering
Department Of ECE,WanaDongri, Nagpur, India
Tel: 9970616040
E-mail: atish_khobragade@rediffmail.com

[b]Priyadarshini Institute Of Engineering
Hingna Road, Near CRPF Campus, Nagpur, India
Tel : 9422926170
E-mail : cgdethe@rediffmail.com

[c]Visvesvaraya National Institute of Technology
Nagpur,India,
Tel :9822470387

ABSTRACT

This paper presents object tracking methods in video. Different algorithms based on rigid, non rigid and articulated object tracking are studied. The goal of this article is to review the state-of-the-art tracking methods, classify them into different categories, and identify new trends. It is often the case that tracking objects in consecutive frames is supported by a prediction scheme. Based on information extracted from previous frames and any high level information that can be obtained, the state (location) of the object is predicted. An excellent framework for prediction is kalman filter, which additionally estimates prediction error. In complex scenes, instead of single hypothesis, multiple hypotheses using Particle filter can be used. Different techniques are given for different types of constraints in video.

Keywords: *Motion, Articulated, Occlusion, Object tracking, Bayesian Tracking.*

1.0 INTRODUCTION

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. Given 'm' objects moving in scene, a sequence of 'n' image frames is taken from the scene. The aim of tracking is to automatically find the same object in an adjacent frame from a video sequence once it is initialized. The previous research on object tracking falls into three different categories: appearance modeling, motion modeling, and searching methods. The appearance of an object is either directly represented by the image or some kind of features in a feature space. Object appearances can change over time due to image distortion, illumination changes, object motion, and occlusion. A motion model predicts an object's location in a new frame of an image sequence using its motion history and other known object movement characteristics. Linear models impose constraints that an object can only have translational or affine motions (Lucas Kanade, 1981). Nonlinear models impose less constraint on motion than do linear models, but they are more difficult to estimate and are more sensitive to noise (Davatzikos, Prince & Bryan 1996). Searching methods use various strategies to find an object within an area predicted by a motion model, that is, the object whose appearance is the most similar to the appearance of the tracked object in an adjacent frame of a video sequence. Apart from the location, a searching algorithm may also search for the most proper scale of the tracking target. Early searching algorithms were developed for tracking feature points in video, such as the Lucas-Kanade algorithm and Shi's feature point selection algorithm (Shi Tomasi, 1994). Feature point searching algorithms do not utilize spatial constraints of the feature points of an object and therefore are often used to compute optical flow. Many efficient and robust searching algorithms such as mean-shift (Comanicu, Ramesh & Meer, 2000) have been developed to search for the local best match for a rigid or non rigid object. Probabilistic modeling and sampling techniques are employed to achieve efficient tracking. For example, the Kalman filter has been used to track objects using the randomness generated by a linear dynamic operator perturbed by Gaussian noise. Particle filtering is superior to Kalman filtering without being constrained by the assumptions of linear dynamic and Gaussian observations using nonparametric density estimation and multiple hypotheses (Isard & Blake, 1998). Template methods reserve full spatial information and have been successfully applied in tracking rigid objects (Deutsch, Grass, Bajramovic & Denzler, 2005). However, template methods are not robust for tracking non rigid object motion. Many features have been proposed to characterize both rigid and non rigid object appearances such as color histogram (Heisele, Kressel & Ritter, 1997) contours (Kass, Witkin & Terzopoulos, 1988), (Yilmaz, Li & Shah, 2004) and texture descriptors (Shahrokni, Drummond & Fua, 2004). The object appearances through the temporal dimension can be learned globally using statistical models such as a linear prediction scheme (Yang & Waibel, 1996), the Gaussian mixture model over time (Grimson & Stauffer, 1999), adaptive filter methods, minimal and maximal intensity value methods, the PDE level set, Hidden Markov models

(HMMs), and kernel density estimation techniques (Elgammal, Duraiswami, Harwood, & Davis, 2002). In (Park & Aggarwal, 2002), researchers proposed to segment a person into local regions and track the local regions individually to improve people tracking performance. Trust-region tracking system is more effective than a line-search-based mean-shift tracker.

2.0 OBJECT REPRESENTATION

(Marr, 1982) proposed five criteria for evaluating object representations: Accessibility - needed information should be directly available from the model rather than derivable through heavy computation, Scope - a wide range of objects should be representable, Uniqueness - an object should have a unique representation, Stability – small variations in an object should not cause large variations in the model and Sensitivity - detailed features should be represented as needed. Object Representation Modeling schemes in computer vision mainly belong to two families: Property representations that define objects by properties or constraints, without recourse to an explicit geometric model, the satisfaction of which should lead to unique identification, and Geometric representations that represent object shape and structure.

A typical property representation associates lists of expected properties with each object. Some examples of this are: Property and relationship representations often take the form of a graph. Here, object features become nodes in the graph, relationships between the features become the arcs and properties of the features are labels on the nodes.

Graph representations have the advantage of adding some structure to the object properties, and providing a common representation method for many problems. One problem is all object details tend to be represented at the same level, so the graphs can become large without benefit. Adding more detail increases the computational difficulties of matching rather than easing them. Hierarchical graph representations are investigated in matching to try to overcome the computational complexity.

The factors taken into consideration for Property Representation are: color, size and height for image regions, rough object sizes, colors and edge shapes for desk top objects, face shape, edge lengths and two dimensional edge angles for identifying polyhedra.

Early geometric models were based on three dimensional point or line descriptions.

Points: The object is represented by a point, that is, the centroid (Figure 1(a)) or by a set of points (Figure 1(b)) Primitive geometric shapes: Object shape is represented by a rectangle, ellipse Figure 1(c), (d). Object silhouette and contour: Contour representation defines the boundary of an object Figure 1(g), (h). The region inside the contour is called the silhouette of the object (see Figure1 (i)). Silhouette and contour representations are suitable for tracking complex non-rigid shapes

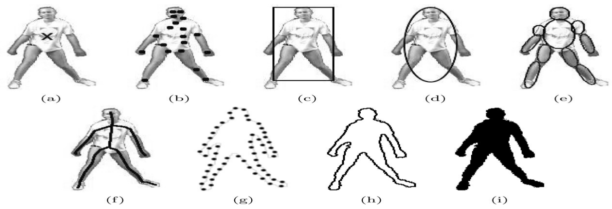

Figure 1: Different Object Representation

Articulated shape models: Articulated objects are composed of body parts that are held together with joints. The relationship between the parts is governed by kinematics motion models, for example, joint angle, etc. Skeletal models: Object skeleton can be extracted by applying medial axis transform to the object silhouette. Skeleton representation can be used to model both articulated and rigid objects (see Figure 1(f). There are a number of ways to represent the appearance features of objects. Some common appearance representations in the context of object tracking are:

—Probability densities of object appearance. Parametric & Non Parametric.

—Templates are formed using simple geometric shapes or silhouettes. An advantage of a template is that it carries both spatial and appearance information.

—Active appearance models. Active appearance models are generated by simultaneously modeling the object shape and appearance. For each landmark, an appearance vector is stored which is in the form of color, texture, or gradient magnitude.

3.0 OBJECT DETECTION

Detection refers to anything from identifying a location to identifying and registering components of a particular object class at various levels of detail. For example, finding the faces in an image, finding the eyes and mouths of the faces. One could require a precise outline of the object in the image, or the detection of a certain number of well-defined landmarks on the object, or a deformation from a prototype of the object into the image. The deformation could be a simple 2D affine map or a more detailed nonlinear map. The object itself may have different degrees of variability. It may be a rigid 2D object, such as a fixed computer font or a 2D view of a 3D object, or it may be a highly deformable object, such as the left ventricle of the heart. All these are considered object-detection problems, where detection implies identifying some aspects of the particular way the object is present in the image—namely, some partial description of the object instantiation. Issues related with object detection are shape variation and illumination variation. Using feature extraction, feature transform and machine learning we can detect objects. The various classifiers used are SVM, Bayesian network, Neural Network, Adaboost etc. Different objects, due to their distinct structure and texture properties, might results in different graph models, for example, Composition model, Constellation Model, and Pictorial Model. Computational speed is another issue in the object detection algorithms. Simple graph model usually results in faster computation

while complicated graph models need optimization in the computations. Dynamic program, belief propagation algorithms can be applied to accelerate the computation.
The difficulty of articulated object detection lies in two aspects: the shape variance and the self-occlusion. Because the large number of degrees of freedom of articulated objects, it is hard to build a shape model to model all possible shapes of articulated objects, although some researchers did build such models. The other factor is the self-occlusion of articulated objects. Previous articulated object detecting systems, in order to deal with the large shape variance, either take the "pose-based" approaches or part-based approaches.
Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. Object detection is done through Point Detectors, Background Subtraction, Segmentation, and Supervised Learning. Motion segmentation based on an adaptive background subtraction method models each pixel as a mixture of Gaussians and uses an on-line approximation to update the model. This yields a stable, real-time outdoor tracker that reliably deals with lighting changes, repetitive motions from clutter, and long-term scene changes. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint. Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. This process is referred to as the background subtraction (Fig 2). Blob trackers have become increasingly powerful in recent years largely due to the adoption of statistical appearance models which allow effective background subtraction and robust tracking of deforming foreground objects. Labeling and Object passing in front are the drawbacks of Blob trackers.
For Gaussian Modeling of stationary background the model parameters, the mean $\mu(x, y)$ and the covariance $\Sigma(x, y)$, are learned from the color observations in several consecutive frames. Every pixel (x, y) in the input frame, the likelihood of its color coming from $N(\mu(x, y), (x, y))$ is computed, and the pixels that deviate from the background model are labeled as the foreground pixels. Adaptive Gaussian method is used for object detection. The most important limitation of background subtraction is the requirement of stationary cameras. Camera motion usually distorts the background models.

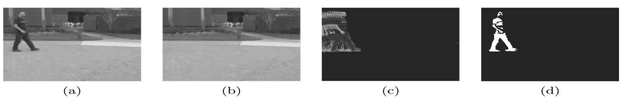

Figure 2: Mixture of Gaussian modeling for background subtraction. *(a)Image from a sequence in which a person is walking across the scene (b) The mean of the highest-weighted Gaussians at each pixel position. These means represent the most temporally persistent per-pixel color and hence should represent the stationary background. (c)The means of the Gaussian with the second-highest weight; these means represent colors that are observed less frequently. (d)Background subtraction result. The foreground consists of the pixels in the current frame that matched a low-weighted Gaussian.*

4.0 OBJECT TRACKING

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video. Tracking objects can be complex due to: Loss of information caused by projection of the 3D world on a 2D image, Noise in images, Complex object motion, Non-rigid or articulated nature of objects, Partial and full object occlusions, Complex object shapes, Scene illumination changes, and Real-time processing requirements.
Tracking algorithms can be classified into Single Object & Single Camera, Multiple Object & Single Camera, Multiple Objects & Multiple Cameras, and Single Object & Multiple Cameras. Single Object & Single Camera require accurate camera calibration and scene model, suffers from Occlusions, it is not robust and it is object dependant. Single Object & Multiple Camera give accurate point correspondence between scenes. Occlusions can be minimized or even avoided, redundant information for better estimation, it suffers from multiple camera Communication problems. Multiple camera views resolve object occlusion, but camera calibration & synchronization is to be done accurately.
Main tracking categories are: Point Tracking, Kernel Tracking, and Silhouette Tracking (Fig.3).

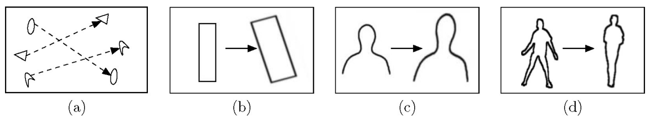

Figure 3: Different tracking approaches
(a) Multipoint correspondence (b) Parametric transformation of a rectangular patches (c, d)Two examples of contour evolution.

Point correspondence is a complicated problem-specially in the presence of occlusions, misdetections, entries, and exits of objects. Different motion constraints are: Proximity, Maximum velocity, Small velocity-change, Common motion, Rigidity constraints. Point tracking methods can be evaluated on the basis of whether they generate correct point trajectories. Rigid object tracking can be divided into four parts:

4.1 Object Tracking Approaches Using Region Based

In the initialization step, color histograms of all the objects of interest in the scene are computed from a frames of video sequence, stored in a database as reference color histograms, and are used later in the matching process. In each new frame of the video sequence for each of the tracked objects a color histogram is calculated for every candidate object position. Each derived histogram i.e. target histogram is compared against the reference color histogram of the object in order to determine the best match and find the position of the tracked object in current frame.
The current frame is searched for a region, namely a window of variable size but fixed shape whose color content best

matches a reference color model. In this case the color distribution of the object is considered as multimodal and, as such, is approximated by a number of Gaussian functions in some color space. Starting from the object location in the previous frame, the method proceeds iteratively at each frame so as to minimize a distance measure to the reference color model. Since object color can often change due to illumination conditions, the model is adapted to reflect the changing appearance of the tracked object. A statistical approach is used in which color distributions are estimated over time by sampling from the object pixels to obtain a new pixel set that is used to update the Gaussian mixture model. An easily implemented algorithm for color changes due to illumination is to use normalized color space.
Using Blob analysis, the object in complex background can be tracked. For each pixel in a blob, its spatial coordinates, along with its textural components are used to form a feature vector. The statistics of each blob are updated with the new information coming from the recently acquired images.
Background subtraction using erosion and dilation is used for outdoor scenes, where the illumination is not constant. The basic idea is that pixels that have been erroneously assigned to the foreground (i.e. outliers) can be eliminated by erosion, while a combination of dilations and erosion can smooth the image regions corresponding to the foreground objects. The scene model can also be dynamically updated.

4.2 Contour Based Object Tracking

Active contours, also known as snakes have been extensively used by researchers to perform contour delineation and tracking. Snakes consist of an elastic parametric curve that can be dynamically deformed to match object shapes. The deformation is subject to internal forces due to contour elastic forces and external forces due to image content and other constraints. Since tracking based on snakes is sensitive to initialization i.e., the snakes need to be initially placed close to the contour that needs to be tracked, otherwise it will fail, a temporal prediction module capable of estimating the position of the object outline in the next frame is often used in conjunction with the snake algorithm, thus achieving at the same time, a reduction in the computational complexity of the snake algorithm.
Objects are modeled as a curve or a set of curves and represented at time t by an image curve parameterized in terms of B-spline. Tracking object using Graph cuts based on active contour tracking is an iterative method that can deform itself to match the desired object boundary. The algorithm uses both the intensity information within the current frame and the intensity difference between the current and the previous frame to find the next position of the object contour.

4.3 Feature Based Tracking

In general, the most desirable property of a visual feature is its uniqueness so that the objects can be easily distinguished in the feature space. Color is used as a feature for histogram-based appearance representations. For contour-based representation, object edges are usually used as features. There are four types of features Color, Edges, Optical flow, Texture. Among all features, color is one of the most widely used features for tracking. Color based Probabilistic tracker based on the principle of histogram difference can effectively handle clutter in the background as well as occlusion (P´erez, Hue, Vermaak, & Gangnet, & Heyden et al. (Eds.), 2002). Optical flow as a feature for contour tracking is shown in (Fig.4). Optical flow is commonly used as a feature in motion-based segmentation and tracking applications. Local features of the detected objects are extracted using the Scale Invariant Feature Transform (SIFT), Harris Operator and, KLT operators.

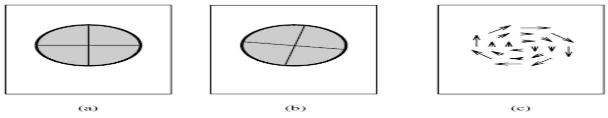

Figure 4: (a) Time t1 (b) Time t2 (c) Optical Flow

Feature Selection using Gabor wavelets gives scaling and rotation invariant object tracking, using Gabor functions with different frequency centers and orientations. If we use S' different frequencies and T different orientations, each image pixel will be associated with S'xT coefficients, the amplitudes of which form a feature vector at the specific pixel.
Texture is another feature which measures the intensity variation of a surface which quantifies properties such as smoothness and regularity. Compared to color, texture requires a processing step to generate the descriptors. Texture calculates local range of an image, local standard deviation of an image, local entropy of a grayscale image. Entropy is a statistical measure of randomness. Gray-Level Co occurrence Matrices is another texture descriptor used.

4.4 Template Based Tracking

It was assumed that template tracking is performed solely on rigid objects. However due to the non-rigid nature of many natural objects or due to viewpoint changes, template tracking fails to provide satisfactory results in a number of real world scenes. For that reason deformable template tracking methods have been introduced in which prior knowledge of the object shape is used in an energy minimization scheme. Deformable templates are specified by a set of parameters which enable a priori knowledge about the expected shape of the object to guide the template matching process. The deformable templates interact with the image in a dynamic manner. An energy function is defined for the template, consisting of terms attracting the template to salient features like intensity, edges. The template parameters are obtained by a minimization of the energy function. In essence the deformable template is obtained by allowing an original template to deform using any appropriate deformation function. The result should cover the various instances of the deformable object as much as possible, with minimum computational overhead while maintaining the attributes of the template like smoothness, connectivity, etc.
Object tracking based on region matching using gradient histogram matching and template matching through

normalized Cross correlation for rotation robustness is presented in (David & Ebrahimi, 2007). There is no need of tunning of classifier and much reduced number of paramters is required.

5. 0 BAYESIAN OBJECT TRACKING

The tracking scheme first extracts a set of measurements, observations for the estimation of state vector. Then a predictor-corrector filter is exploited to filter the results. Finally, a Bayesian network performs spatial data integration using triangulation, perspective projections, and Bayesian inference. The input to the network is the set of measurements from the previous time step and the states. Kalman filter is a special case of the Bayesian filters, and is the best possible estimator. Two conditions should be satisfied: Functions should be linear & known. The distributions of the process & measurement noises are again Gaussian. Kalman filter is a set of mathematical equation that provides a computationally efficient recursive solution to the least square method. Kalman filter operates in a two step. The current estimate along with an estimate of the error covariance is propagated forward in time. The second stage incorporates a new measurement to modify the propagated current state and error covariance. If the posterior pdf is not Gaussian, kalman filters will not perform adequately. In such a case, particle filters can be used. They are sequential Monte Carlo methods that can be used for object tracking within the Bayesian framework. They come in a variety of names, such as conditional density propagation or the condensation algorithm. The main concept behind particle filters is to represent the probability distribution of alternative solutions as a set of samples i.e. particles, each of which carries a weight. Estimates of the posterior probability distribution are calculated based on the samples and their associated weights. As the number of sample grows the filter approaches to optimal Bayesian estimate. Variation Particle filter is proposed for multi-object tracking, where the proposal distribution is based on the approximated posterior from variational inference rather than using the prior as the proposal distribution in Sampling Importance Re-Sampling particle filter (Jin & Mokhtarian, 2007).
Tracking curves in dense visual clutter is a challenging one. One very effective approach is to use random sampling. The condensation algorithm combines random sampling with learned dynamical models to propagate an entire probability distribution for object position and shape over time. The result is accurate tracking of agile motion in clutter, decidedly more robust than what has previously been attainable by kalman filtering. Particle filters have attracted much attention due to their robust tracking performance in cluttered environments. Particle filters maintain multiple hypotheses simultaneously and use a probabilistic motion model to predict the position of the moving object, and this constitutes a bottleneck to the use of particle filtering in real-time systems due to the expensive computations required.

6.0 2D-ARTICULATED OBJECT TRACKING

2D-Arrticulated object tracking can be model free or model based. Model free methods proceed by exploiting image information like edges, intensity, etc in order to create coherent structures that correspond to the rigid parts of the articulated structure. Occlusion can create problems in such methods, due to lack of visibility or partial visibility of one or more of the rigid parts. Alternatively model based approaches use 2D model depending on the application & precision required. The rigid parts of a 2D model can be represented using geometric primitives, such as sticks, circles, rectangles, and ellipses or by curves and snakes if the object parts are allowed to deform. These can be used to represent the shape of the objects. Additionally, the appearance of the object can be modeled by including texture on the above mentioned geometric primitives. The displacement vectors and the joint angles are the parameters that describe the motion of the articulated object. The dynamic model is a stochastic linear equation that rules these parameters and is used to predict the model configuration. A template matching technique is used to detect the specific feature points for each configuration. The largest matching score is selected as the final result for the frame. A technique based on decentralized scheme and modeling the interpart interaction density within an Bayesian framework (Yang & Waibel, 1996). It exploits decentralized framework using graphical model analysis which gives robustness and speed compared with other articulated object tracking methods.
A model-free articulated object tracking system is presented in (Jesus, Abrantes & Marques, 2002) avoiding the use of prior models by exploiting the pyramid-based Kanade-Lucas feature tracking algorithm and a foreground color distribution obtained through training.

7.0 CONCLUSION

Significant progress has been made in object tracking during the last few years. The assumptions used to make the tracking problem tractable, for example, smoothness of motion, minimal amount of occlusion, illumination constancy, high contrast with respect to background, etc., are violated in many realistic 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. Occlusion can be classified into three categories: self occlusion, inter-object occlusion, and occlusion by the background scene structure. Partial occlusion of an object by a scene structure is hard to detect since it is difficult to differentiate between the object changing its shape and the object getting occluded. The SIFT method is capable of tracking objects under partial or severe occlusions. A common approach to handle complete occlusion during tracking is to model the object motion by linear dynamic models or by nonlinear dynamics and, in the case of occlusion, to keep on predicting the object location until the object reappears. Multiple cameras viewing the same scene can also be used to resolve object occlusions during tracking. The first reason for using multiple cameras is the use of depth information for tracking and occlusion resolution. The second reason is to increase the area under view since it is not possible for a single camera to observe large areas because of a finite sensor field-of-view. One challenge in tracking is to develop algorithms for tracking objects in unconstrained videos, for example, videos obtained from broadcast news networks or home videos. Another related video domain is of formal and informal meetings. Thus, there is severe occlusion, and people are only partially visible. One interesting solution in this context is to employ audio in addition to video for object tracking.

REFERENCES

Comanicu, D. Ramesh, V. & Meer, P. (2000). Real-Time Tracking of Non-Rigid Objects Using Mean Shift, *Proc. Int'l Conf. Computer Vision and Pattern Recognition, 142-149.*

Davatzikos, C. Prince, & J. Bryan, R.(1996). Image Registration Based on Boundary Mapping, *IEEE Trans. Medical Imaging, vol. 15, no. 1, 112-115.*

David, M. & Ebrahimi, T. (2007). Orientation histogram based matching for region tracking, *WIAMIS'07*

Deutsch, Grassl, C. Bajramovic, F. & Denzler, J. (2005). A Comparative Evaluation of Template and Histogram Based 2d Tracking Algorithms, *German Pattern Recognition Symposium. pp. 269 -276.*

Elgammal, A. Duraiswami, R. Harwood, D. & Davis, L. (2002). Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, *Proceedings IEEE, vol. 7, no. 90, pp. 1151-1163.*

Grimson, W.E.L. & Stauffer, C. (1999). Adaptive Background Mixture Models for Real-Time Tracking, *Proceedings Int'l Conf. Computer Vision and Pattern Recognition, pp. 22-29.*

Heisele, B. Kressel, U. & Ritter, W. (1997). Tracking Non-Rigid Moving Objects Based on Color Cluster Flow, *Proc. Int'l Conf Computer Vision and Pattern Recognition, pp. 253-257.*

Isard, M. & Blake, A. (1998). Condensation Conditional Density Propagation for Visual Tracking, *Int'l J. Computer Vision, vol. 29 no. 1, pp. 5-28.*
Jesus, M. Abrantes, A.J. Marques, J.S. (2002). Tracking the human body using multiple predictors, *Proceedings of the Second International Workshop on articulated motion and Deformable Objects (London, UK, 2002), 155-164.*

Jin, Y. & Mokhtarian, F. (2007). Variational Particle Filter for Multi-Object Tracking, 2007 *IEEE.*

Kass, M. Witkin, A. & Terzopoulos, D. (1988). *Snakes: Active* Contour Models, *Int'l J. Computer Vision, vol. 1, no. 4, pp. 321-331.*

Lucas, B. & Kanade, T. (1981). An Iterative Image Registration Technique with an Application to Stereo Vision, *Proc. DARPA Image Understanding Workshop, pp. 121-130.*

Marr, D., (1982) Vision, pubs: W.H. Freeman and Co., San Francisco.

Park, S. & Aggarwal, J. (2002). Segmentation and Tracking of Interacting Human Body Parts under Occlusion and Shadowing, *Proc. IEEE Workshop Motion and Video Computing.*

P´erez1, P. Hue2, C. Vermaak1, J. & Gangnet1A, M. & Heyden et al. (Eds.). (2002). Color-Based Probabilistic Tracking, *ECCV 2002, LNCS 2350, pp. 661–675, 2002.Springer-Verlag Berlin Heidelberg*

Qu, W. & Schonfeld, D. (2007) Real-Time Decentralized Articulated Motion Analysis and Object Tracking From Videos, *IEEE Transactions On Image Processing, Vol. 16, No. 8, pp.2129-2138.*

Shahrokni, A. Drummond, T. & Fua, P. (2004) Texture Boundary Detection for Real-Time Tracking, *Proc. European Conf. Compute Vision, pp. 566-577.*

Shi, J. & Tomasi, C. (1994). Good Features to Track, *Proceedings Int'l Conf Computer Vision and Pattern Recognition, pp. 593-600.*

Yang, J. & Waibel, A. (1996). A Real-Time Face Tracker, *Proc Workshop Computer Vision, pp. 142-147.*

Yilmaz, A. Li, X. & Shah, B. (2004). Contour Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras, *IEEE Trans. Pattern Analysis and Machine Intelligence vol. 26, no. 11, pp. 1531-1536.*