OCCLUSION RESISTANT OBJECT TRACKING

E. Loutas† *K. Diamantaras*‡ *I. Pitas*†

†Department of Informatics
University of Thessaloniki
GREECE
(E-mail: pitas@zeus.csd.auth.gr)

‡Department of Informatics
Technological Education Institute
of Thessaloniki GREECE
(E-mail: kdiamant@it.teithe.gr)

ABSTRACT

Object tracking with occlusion prediction using multible feature correspondences is proposed. The tracking region is defined by a set of point features, tracked using Kanade-Lucas-Tomasi algorithm. During total occlusion the region position is estimated using motion prediction based on a Kalman filtering scheme applied to the motion model prior to occlusion. During partial occlusion the displacements of the occluded features are predicted based on the motion of the bounding box of the moving object. Experimental results on real and artificial images have shown that the algorithm behaves well under total and partial occlusion.

1. INTRODUCTION

Occlusion (partial or full) is a serious problem in the object tracking process. It can cause loss of target in many object tracking algorithms, or it can seriously downgrade the tracker performance. Many object tracking algorithms fail to work correctly even in partial occlusion situations, although some methods show robustness to different amounts of occlusion.

In general, object tracking algorithms can be based either on minimizing the sum of squared differences between two regions [1], or on the detection and tracking of a sparse collection of point features (or contours). Examples of algorithms robust to partial occlusion can be found. Robustness in partial occlusion is achieved by using robust statistics [1]. An object tracking algorithm robust to severe partial occlusions based on Kalman snakes has also been introduced [2], while in [3] object tracking is achieved using Deformable Templates. The last algorithm is insensitive to moderate amounts of occlusion. An algorithm that is insensitive on the disappearance and reappearance of point features is described in [4]. A system based on active contours that can automatically determine when occlusion has occurred and is robust to significant occlusion is presented in [5]. In [6], object tracking is performed by tracking a set of point features. Geometric invariants are used for verification. However none of the above methods behaves well under total occlusion.

The tracking algorithm proposed in this paper is based on minimizing the sum of squared differences of a large set of point features generated in the tracking region. The Kanade-Lucas- Tomasi feature tracking algorithm (KLT) is used for feature tracking [8]. Kalman filtering motion prediction is employed to estimate the tracked region position during occlusion. Subsequently, a graph matching based object verification scheme is utilized to allow object tracking after reappearance. Robustness to partial occlusion is achieved by estimating the occluded features' motion using the estimated motion of the bounding box.

One of main characteristics of this algorithm is the fact that it treats occlusion without using depth information acquired from stereo image sequences. Furthermore, experimental results provide evidence that the algorithm can cope well with full occlusion. For most of the existing algorithms this is usually not the case. Nevertheless the proposed algorithm performance deteriorates when the tracked object is fully occluded for very long time periods.

2. ALGORITHM

2.1. Tracking of point features

The algorithm is based on selecting a large number of point features in the specified tracking region. These features are tracked in the next frame. The tracking region in the next frame is specified as the bounding rectangle of all the tracked features. The Kanade Lucas Tomasi (KLT) algorithm [8] is utilized for tracking of the point features. The displacement $\mathbf{d} = [d_x, d_y]^T$ between two feature windows on images I and J is found by minimizing

$$\epsilon = \int\int_W [J(\mathbf{x} + \frac{\mathbf{d}}{2}) - I(\mathbf{x} - \frac{\mathbf{d}}{2})]^2 w(\mathbf{x}) d\mathbf{x} \qquad (1)$$

where $\mathbf{x} = [x, y]^T.W$ is the region of the window and $w(\mathbf{x})$ is a weighting function that can be set to 1 for simplicity. The equation (1) uses $[J(\mathbf{x} + \frac{\mathbf{d}}{2}) - I(\mathbf{x} - \frac{\mathbf{d}}{2})]$ instead of $[J(\mathbf{x}) - I(\mathbf{x} - \mathbf{d})]$ used in [8], because of its symmetry with respect to both images [12]. In order to perform one iteration of the minimization procedure of (1) the equation $Z\mathbf{d} = \mathbf{e}$ must be solved where:

$$Z = \int\int_W \mathbf{g}(\mathbf{x})\mathbf{g}^T(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \qquad (2)$$

$$\mathbf{e} = 2\int\int_W [I(\mathbf{x}) - J(\mathbf{x})]\mathbf{g}(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \qquad (3)$$

and

$$\mathbf{g} = \begin{bmatrix} \frac{\partial(I+J)}{\partial x} \\ \frac{\partial(I+J)}{\partial y} \end{bmatrix}. \qquad (4)$$

Feature occlusion is determined using the process described in [8].

In order to avoid tracking background features, a clustering procedure is applied. The mean (μ_x, μ_y) and the variance (σ_x, σ_y)

0-7803-6725-1/01/$10.00 ©2001 IEEE

of the feature coordinates are found for each frame's tracked region. Let $[x, y]^T$ be the coordinates of some feature at frame t and (μ_x, μ_y), (σ_x, σ_y) are the mean and variance for the feature locations in that frame. Then the feature is kept in frame $t+1$ if $x \epsilon [\mu_x - \sigma_x, \mu_x + \sigma_x]$, $y \epsilon [\mu_y - \sigma_y, \mu_y + \sigma_y]$, else it is rejected. Assuming that the object features have similar motion patterns, we are enabled to reject stationary or slow moving background features, after a number of frames, while keeping object features. This is particularly useful if the initialized region contains a portion of background.

2.2. Initialization

The region bounding box is used to specify the region to be tracked. Inside the tracking region a large number of point features is generated. Feature generation is based on the algorithm used for the point feature tracking. A good feature is defined as the one whose matrix Z has two large eigenvalues that do not differ by several orders of magnitude [7]. A good feature assures that the equation $Z\mathbf{d} = \mathbf{e}$ is well conditioned. It can be found that the large eigenvalue prerequisite implies that the partial derivatives $\frac{\partial(I+J)}{\partial x}$ and $\frac{\partial(I+J)}{\partial y}$ are large.

2.3. Robustness to Occlusions

The occluded features of the feature set cannot be reliably tracked using the KLT algorithm. Thus occlusion causes many problems during the object tracking procedure.

In order to cope with partial occlusion, a prediction scheme is utilized. The lost features are not tracked by the tracker but their coordinates are updated using the estimated movement of the upper left and the lower right point of the bounding rectangle. The procedure is stopped if the occlusion is total that is when none of the features comprising the feature set can be tracked correctly due to occlusion.

In order to cope with total occlusion the position of occluded region is updated using the velocity estimates of the region corners obtained from the measurements before total occlusion with the help of a Kalman filtering scheme. The Kalman filtering prediction process is applied on the upper left corner and the lower right corner of the region bounding rectangle before total occlusion. A constant acceleration model is used.

Let $\mathbf{d}(k)$, $\mathbf{u}(k)$, and $\mathbf{a}(k)$ denote the displacement velocity and acceleration for each corner of the bounding box at time k respectively. The state-transition equation for each corner is, [9]:

$$\mathbf{s}(k) = \mathbf{C}\mathbf{s}(k-1) + \mathbf{w}(k), k = 1, \ldots, N \quad (5)$$

where $\mathbf{w}(k)$ is a zero mean, white random sequence and $\mathbf{s}$ is a 6x1 vector containing the coordinates of displacement velocity and acceleration, for each corner of the bounding box

$$\mathbf{s} = \left[\begin{array}{cccccc} d_x & d_y & u_x & u_y & a_x & a_y \end{array} \right]^T . \quad (6)$$

The measurements $\mathbf{d}(k)$ are related to the state variables $\mathbf{s}(k)$ with

$$\mathbf{d}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{v}(k), k = 1, \ldots, N \quad (7)$$

where $\mathbf{v}(k)$ denotes a zero-mean, white observation noise sequence. The matrices describing the model are given below:
The 2x1 observation vector is:

$$\mathbf{d} = \left[\begin{array}{c} d_x \\ d_y \end{array} \right] . \quad (8)$$

The 2x6 measurement matrix is:

$$\mathbf{H} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array} \right] . \quad (9)$$

The observation equation states that the noisy displacement coordinates of each corner of the bounding box can be observed.
The 6x6 state transition matrix describing the model is [9]:

$$\mathbf{C} = \left[\begin{array}{cccccc} 1 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right] . \quad (10)$$

2.4. Reappearance Prediction

Reappearance prediction is made possible by estimating and tracking the *occluding* region. To estimate the occluding region bounding box a simple region growing segmentation algorithm is utilized. The seed is determined by the last position of the occluded region before total occlusion. The occluded object is considered to be disoccluded in its entirety when the intersection of the occluding region and the predicted occluded region is found to be zero.

$$area1 \bigcap area2 = \emptyset,$$

where $area1$ is the occluding region and $area2$ is the predicted occluded region. The occluded region appears again provided that the above condition is satisfied, if it is similar enough to the occluded region. The reappearance is made possible by regenerating of a set of point features. The regeneration procedure is the same as initialization. The selected features are those that have large eigenvalues of matrix Z.

2.5. Reappearance verification

When the tracker predicts that reappearance has taken place it also has to decide if the region to reappear is similar enough to the tracking region before occlusion. This can be made possible either by graph matching or by calculating region difference.

2.5.1. Graph matching

To verify the reappearance of the occluded object, a modified graph matching algorithm is implemented. We assume the object is rigid and undergoes simple motion that does not distort the graph. A set of point features is generated on the region being tracked in the last frame of the image sequence, before total occlusion takes place, forming the initial grid. The tracker is used to track the set of point features between the initial and the current frame. To accomplish a successful tracking, a modification of the tracking algorithm is performed. The search range of the algorithm is updated using the prediction results. The predicted region is allowed to slightly change (by little motion) and the resulting grid is allowed to deform. The grid deformation is controlled by the tracker search region, (the allowed deformations do not exceed the tracker search range). The initial and the resulting grid are usually different. The difference can be caused by the grid deformation and by the disappearance of certain features. To measure the difference a cost function J is utilized.

$$J = \frac{\sum_i \sqrt{(x_{i1} - x_{i2})^2 + (y_{i1} - y_{i2})^2}}{N_{RF}} + \lambda N_{LF}, \qquad (11)$$

where $(x_{i1}, y_{i1}), (x_{i2}, y_{i2})$ are feature coordinate pairs relative to the upper left corner of the region bounding box, N_{RF} is the number of the point features that are not lost and N_{LF} is the number of the lost point features. It is assumed that the result graph resembles the initial graph if the cost function (5) is smaller than a predefined threshold.

2.5.2. *Estimating Region difference*

As an alternative, a less computationally intensive method for verifying the reappearance of the occluded object is to calculate the region mean square error between the reference and the target image. To avoid excessive noise influence image presmoothing is necessary. If the region mean square error is below a certain threshold then the reappearance of the occluded object is verified. To ensure reliability, an appropriate threshold must be chosen. This method does not have the limitations of the graph matching method. Nevertheless the computational cost can be greatly affected by the region size. Moreover the reliability of the method is decreased if it is applied to large regions. Our experience shows that this approach has poorer performance compared to the graph matching methods, but it is less computationally expensive.

3. RESULTS

The algorithm was tested on real and artificial images. In Figure 1 we show results on the artificial image sequences. The circle shown in Figure 1a is moving slowly and is fully occluded by a faster moving ellipse (Figure 1b) the tracking region bounding rectangle is reborn after the total circle disocclusion. It is clearly seen that the algorithm works even if the occluding object reappears suddenly and was not present in the first frame of the image sequence.

Results on real image sequences are shown in Figures 2 and 3. In Figure 2 the tracking region (head of the athlete) is occluded by the foot of another athlete. The tracking region reappears after three frames. In Figure 3 results showing robustness to partial occlusions are presented. The face is partially occluded. After the end of partial occlusion the tracking region remains intact.

4. CONCLUSIONS

In this paper, object tracking resistant to partial and full occlusion is presented. Tracking is accomplished by creating a set of point features and subsequently using the Kanade Lucas Tomasi feature tracking algorithm. Resistance to partial and total occlusion is achieved by using a Kalman filtering scheme prior to total occlusion. During partial occlusion the coordinates of lost features are updated using the filter estimates, while in case of total occlusion the estimated position of the occluded region is updated using the filter output.

Experimental results show that the algorithm can well handle partial occlusion and can cope with total occlusion situations. Future work includes the creation of a multi object tracker capable of handling "inter-object" and "external" occlusion.

5. REFERENCES

[1] G. D. Hager and P. N. Belhumeur, "Efficient Region Tracking With Parametric Models of Geometry and Illumination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.

[2] Natan Peterfreund, "Robust Tracking of Position and Velocity with Kalman Snakes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 564–569, 1999.

[3] Yu Zhong, Anil K. Jain, and M.-P. Dubuisson-Jolly, "Object Tracking Using deformable Templates", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 5, pp. 544–549, 2000.

[4] Singh Manku, Pankaj Jain, Amit Aggarwal, Lalit Kumar, and Subhashis Banerjee, "Object Tracking Using Affine Structure for Point Correspondences", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , 1997.

[5] B. Galvin, B. McCane, K. Novins, "Virtual Snakes for Occlusion Analysis", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* , 1999.

[6] U. Uenohara and T. Kanade, "Geometric Invariants for Verification in 3-D Object Tracking", *International Conference on Intelligent Robots and Systems '96 IROS 96 Proceedings of the 1996 IEEE/RS* ,Vol. 2 pp. 785-790 1996.

[7] J. Shi and C. Tomasi, "Good Features to Track", *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR94)*,Seattle, June 1994.

[8] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams: a Factorization Method - Part 3 Detection and Tracking of Point Features", *tech. report CMU-CS-91-132, Computer Science Department, Carnegie Mellon University.*, April 1991.

[9] A. Murat Tekalp, "Digital Video Processing", *Prentice Hall*, 1995.

[10] Andrew Blake and Michael Isard, "Active Contours", *Springer-Verlag*, 1998.

[11] F. Ziliani, F. Moscheni, "Kalman filtering motion prediction for recursive spatio-temporal segmentation and object tracking", *Proc. of the workshop on Image Analysis for Multimedia Interactive Services, Louvain-la-Neuve, Belgium*,pp. 63-68 1997.

[12] Stan Birchfield, "Depth and Motion Discontinuities", *Ph.D. Thesis, Stanford University*, June 1999.

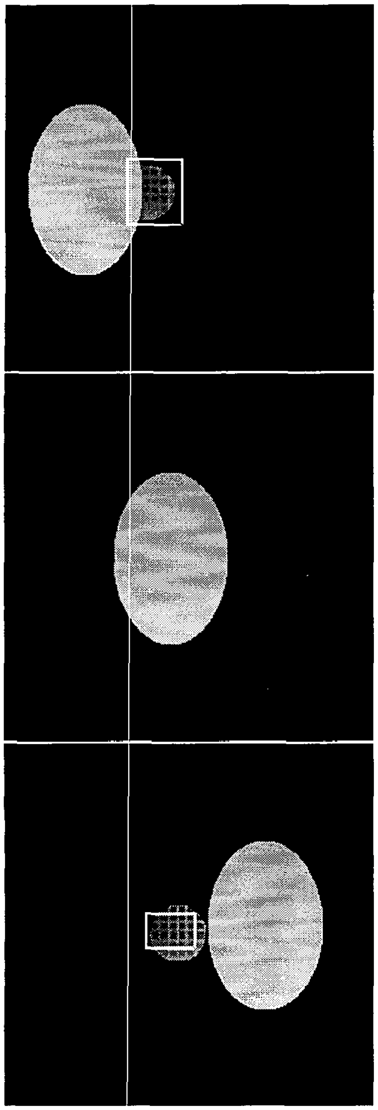

Fig. 1. Artificial image sequence (a) Before total occlusion (b) Occlusion (c) Region reappearance

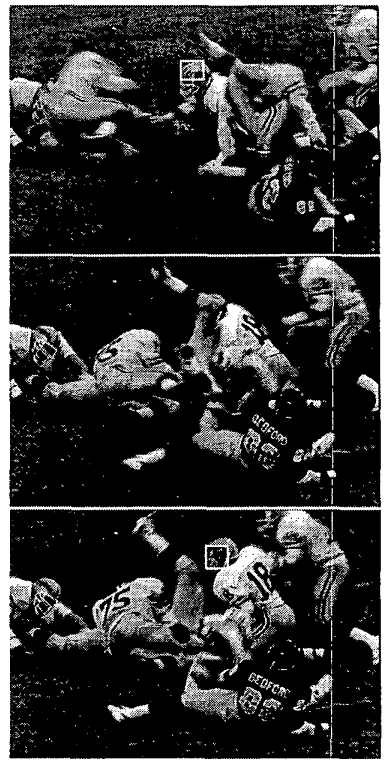

Fig. 2. Football image sequence (a) Before total Occlusion (b) Occlusion (c) Region reappearance

Fig. 3. Lab image sequence (a) Tracking Region initial frame (b) Partial Occlusion (c) Region after occlusion