Entropy - based metrics for the analysis of partial and total occlusion in video object tracking

Evangelos Loutas Christophoros Nikou Ioannis Pitas

University of Thessaloniki
Department of Informatics
Box 451, Thessaloniki 54124, Greece
Tel.+30 2310 996 361, Fax.+30 2310 996 304
E-mail: pitas@zeus.csd.auth.gr

Address for correspondence :
Professor Ioannis Pitas
University of Thessaloniki
Department of Informatics
BOX 451, 54124 Thessaloniki, Greece
Tel. ++ 30 2310 99 63 04
Fax. ++ 30 2310 99 63 04
Email: pitas@zeus.csd.auth.gr

Abstract

Metrics measuring tracking reliability under occlusion that are based on mutual information and do not resort to ground truth data are proposed in this paper. Metrics for both the initialization of the region to be tracked as well as for measuring the performance of the tracking algorithm are presented. The metrics variations may be interpreted as a quantitative estimate of changes in the tracking region due to occlusion, sudden movement or deformation of the tracked object. Performance metrics based on the Kullback-Leibler distance and normalized correlation were also added for comparison purposes. The proposed approach was tested on an object tracking scheme using multiple feature point correspondences. Experimental results have shown that mutual information can effectively characterize object appearance and reappearance in many computer vision applications.

1 Introduction

Partial or full occlusion is an important issue in an object tracking process. A variety of algorithms handling occlusion exist [1, 2, 3, 4]. However, they do not handle total occlusion properly. Tracking is performed in [1] using Sum of square differences (SSD). Tracking does not rely on feature point sets. Partial occlusion and illumination changes are handled. Nevertheless, the proposed algorithm does not handle full occlusion. A contour tracking algorithm is proposed in [2]. The resulting scheme is reliable in image clutter and partial occlusion. Nevertheless, it is not reliable to large amounts of occlusion. The algorithm presented in [3] relies on deformable templates and can handle moderate amounts of partial occlusion. In [4] the role of geometric invariants in tracking is examined. Feature point tracking verification using geometric invariants is presented. The aim of the method is to compute the target feature point set using geometric invariants. An algorithm insensitive to the disappearance and reappearance of feature points is described in [5]. Although the above mentioned methods handle partial occlusion, only few of them behave well under total occlusion.

A model based tracking scheme performing object tracking using edge information and capable of handling partial and total occlusion is proposed in [6]. The proposed method can handle partial and total occlusion events. Nevertheless, it is computationally expensive. A new approach on occlusion resistant object tracking, using Kalman filtering and robust statistics was proposed [7]. This method can handle full occlusion for short time periods. The way the tracking system recovers after total occlusion implies that the position of the disoccluded object lies within the tracker's search range. Another approach for tracking multiple articulated objects in the presence of occlusion based on a Kalman filtering mechanism is presented in [8]. This system was tested in a surveillance scheme used to track moving people. The algorithm has shown good results in severe partial occlusion caused by inter object and object-environment interferences. Finally, a probabilistic multiple object tracking approach working under inter object occlusions is presented in [9].

The performance measure of a tracking algorithm is also an open issue. Although most of the proposed techniques apply subjective evaluation methods, some of them use quantitative approaches based on ground truth [10]. Therefore, implementation of reliability measures not resorting to ground truth data is particularly important. Several metrics for performance evaluation of tracking algorithms without ground-truth, based on color and motion were introduced in [11]. A more recent work on those metrics provides their incorporation in a tracking scheme in order to perform better tracking [12]. A variety of confidence measures for the analysis of optical flow techniques was presented in [13]. However, the confidence measures analyzed in [13] are only used for the evaluation of the velocity field and are application oriented.

The use of mutual information in object tracking as a tool for extracting information concerning the condition of a tracking object is assessed in this paper. The proposed scheme is efficient in extracting information under partial and total occlusion. Mutual information was first introduced in computer vision in [14] for medical image registration applications. In [15, 16] it was applied to combine the outputs of multiple tracking algorithms in order to improve the overall tracker performance.

In our method, the tracking process is modeled as a communication task between a transmitter and

a receiver through a channel. Information theory based metrics are introduced. The mutual information is used as a quantitative measure of the tracking process. Its variations can improve the understanding of the tracked region characteristics and are closely related to changes in the tracked region. These changes are caused by partial or total occlusion, movement of the occluding object and abrupt movements or deformations of the occluded (tracked) object. Determining and understanding these changes may improve tracker performance and assist an event detection scheme. Measures based on the Kullback-Leibler distance and the normalized correlation are also implemented for comparison purposes. The entropy is used as a measure of the initialization efficiency of the tracking process and is closely related to the first metric. The proposed metrics were tested on a feature point based tracking algorithm [17]. The algorithm is enhanced with an occlusion handling scheme, while an object reappearance verification scheme is also designed to allow tracking continuation after object reappearance. It relies on a mutual information-based metric measuring the similarity between a reference and a target region. The modified tracking algorithm performs better than [17] in partial and total occlusion situations.

The main contribution of the current work is the introduction of information theory based metrics as measures of tracking reliability. The use of the metrics does not impose the utilization of ground truth data and is extended to the analysis of partial and total occlusion in object tracking. Moreover, occlusion is processed without resorting to multiple camera systems fusing the outputs of different tracking cues.

The remainder of the paper is organized as follows: The feature point generation and tracking is presented in section 2. The information theory based metrics are described in section 3. Tracking algorithm enhancement is presented in section 4. Experimental results are presented in section 5 and conclusions are drawn in section 6.

2 Feature point generation and tracking

Object tracking is performed by minimizing the sum of squared differences of a large set of feature points generated in the tracking region. The algorithm presented in [17] is used for feature point tracking. Kalman filtering motion prediction is employed to estimate the tracked region position during occlusion. The tracked region in the subsequent video frame is specified as the bounding rectangle of all the tracked feature points. Robustness to partial occlusion is achieved by estimating the motion of the lost feature points, using the estimated motion of the bounding box of the tracked object.

The displacement $\mathbf{d} = [d_x, d_y]^T$ between two feature point windows on images J_2 and J_1 is obtained by minimizing

$$\epsilon = \int\int_W [J_2(\mathbf{x} + \frac{\mathbf{d}}{2}) - J_1(\mathbf{x} - \frac{\mathbf{d}}{2})]^2 w(\mathbf{x}) d\mathbf{x} \tag{1}$$

where $\mathbf{x} = [x, y]^T$, W is the region of the intergration window and $w(\mathbf{x})$ is a weighting function that can be set to 1 for simplicity. Equation (1) uses $[J_2(\mathbf{x} + \frac{\mathbf{d}}{2}) - J_1(\mathbf{x} - \frac{\mathbf{d}}{2})]$ instead of $[J_2(\mathbf{x}) - J_1(\mathbf{x} - \mathbf{d})]$ used in [17], because of its symmetry with respect to both images [18]. In order to perform one iteration of the minimization procedure of (1), the equation $\mathbf{Zd} = \mathbf{e}$ must be solved where:

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

$$\mathbf{e} = 2 \int \int_W [J_1(\mathbf{x}) - J_2(\mathbf{x})]\mathbf{g}(\mathbf{x})w(\mathbf{x})d\mathbf{x} \tag{3}$$

and

$$\mathbf{g} = \begin{bmatrix} \frac{\partial(J_1+J_2)}{\partial x} \\ \frac{\partial(J_1+J_2)}{\partial y} \end{bmatrix}. \tag{4}$$

Feature point occlusion is determined using the process described in [17] and is essentially controlled by the residue ϵ. Large values of ϵ when compared to a predefined threshold imply that the feature point of interest should be rejected. The object tracking occlusion handling is based on feature point occlusion handling as presented in [17] That is, an object part is considered lost when the feature points "belonging" to that part are lost. Other methods of occlusion handling involve the use of constraints based on articulation [19] and layer representation [20]. The approach used in the context of present work is general and can be used in a variety of applications. The layer representation method is very useful in coding and compression, while the role of articulation constraints in determining self occlusions in human body part tracking is vital.

In order to avoid tracking stationary or slowly moving background feature points, we have introduced a clustering procedure. The mean (μ_x, μ_y) and the variance (σ_x, σ_y) of the feature point coordinates are computed for the tracked region in each frame. Let $[x, y]^T$ be the coordinates of a feature point at video frame t and (μ_x, μ_y), (σ_x, σ_y) their mean and variance. A feature point is retained in video frame $t+1$, if $x \in [\mu_x - \sigma_x, \mu_x + \sigma_x]$ and $y \in [\mu_y - \sigma_y, \mu_y + \sigma_y]$, otherwise it is rejected. Assuming that the object feature points have similar motion patterns, we can reject stationary or slow moving background features, after a number of frames, while retaining the moving object feature points. This procedure is particularly useful, if the initialized region to be tracked contains some portions of background regions.

2.1 Initialization

The region bounding rectangle is used to specify the region to be tracked. A large number of feature points is generated inside the tracked region using the process described in [17, 18, 21]. A good feature point is defined as the one whose matrix $\mathbf{Z}$ has two large eigenvalues that do not differ by several orders of magnitude [21]. In order to avoid loss of target, caused by too many lost feature points, the feature point set is periodically regenerated. Different strategies for the periodic feature point regeneration can be applied. It can be thorough (the entire feature point set is regenerated), periodic (it occurs after a fixed number of frames) or asyncronus (its occurrence is based on the tracking process metric value). Feature point generation and tracking are transparent to the observer.

The number of the generated feature points is essentially user controlled. The user controls the number of feature points by selecting their number and the minimum allowed distance between the feature points. Let N_s be the desired number of feature points selected by the user. The number N_k of feature points generated in the region to be tracked depends essentially on the minimal allowed distance between the feature points ($N_k \leq N_s$). Therefore, a set of the possible configurations of the ensemble of the possible feature point sets can be defined. Large minimum allowed distances between the feature points may lead to a small N_k and a poor tracker performance.

3 Robustness to partial and total occlusion

The previously described tracking process can be modeled as a communication between a transmitter (reference frame) and a receiver (target frame) with an N_{max} symbol alphabet (the maximal number of grayscale levels). The tracking process is characterized by loss of information caused by feature point rejection and wrong feature point correspondences. Mutual information is a well known measure of the amount of information transmitted through the communication channel [22, 23]. Therefore, it can be used as a quantitative measure of tracking performance.

3.1 Mutual information as tracker evaluation metric

Let $\mathbf{x}_i^r$ and $\mathbf{x}_i^c$ represent the coordinate vectors of feature point i in the reference and current frame, respectively. During the tracking process, a feature point set of the initial video frame

$$S_1 = [\mathbf{x}_1^r, \ldots, \mathbf{x}_{N_k}^r]^T \tag{5}$$

is tracked to a feature point set

$$S_2 = [\mathbf{x}_1^c, \ldots, \mathbf{x}_N^c]^T, \tag{6}$$

of the target video frame, with $N \leq N_k, N_k \leq N_s$, where N_s is the initial user preference for the number of the feature points

Let U, V be two random variables with marginal probability mass functions $p(u), p(v)$ and $u_i = J_1(\mathbf{x}_k^r)$, $v_j = J_2(\mathbf{x}_k^c)$ their possible outcomes, where J_1 and J_2 are the reference and target image respectively and $\mathbf{x}_k^r \in S_1, \mathbf{x}_k^c \in S_2$. The mutual information of the two random variables U, V with a joint probability mass function $p(u, v)$ is defined as:

$$I(U,V) = \sum_{i=1}^{N_{max}} \sum_{j=1}^{N_{max}} p(u_i, v_j) \log_2 \frac{p(u_i, v_j)}{p(u_i)p(v_j)}, \tag{7}$$

where N_{max} is the maximum number of the available grayscale levels. In order to take into account the lost feature points during the tracking process a cost function E_m is defined:

$$E_m(U,V,N,N_k) = c_1\left(\frac{I(U,V)}{I_{max}(U,V)} - \lambda_1 \frac{N_k - N}{N_k} + c_2\right) \tag{8}$$

The term $\frac{I(U,V)}{I_{max}(U,V)}$ is the mutual information part of the cost function. The maximum mutual information $I_{max}(U,V)$ is [24]:

$$I_{max}(U,V) = -\sum_{i=1}^{N_{max}} p(u_i) \log_2 p(u_i) \tag{9}$$

The term $\frac{N_k - N}{N_k}$ is a penalizing quantity depending on the number of the lost feature points during the tracking process. The use of the penalizing term is necessary, because the mutual information part of the metric measures only the matching efficiency between the feature points that have not been lost. In the context of present work $c_1 = 0.5, \lambda_1 = 1, c_2 = 1$. The constants c_1, c_2, λ_1 are chosen to satisfy:

$$0 \leq E_m \leq 1. \tag{10}$$

In the case of total occlusion:

$$\frac{I(U,V)}{I_{max}(U,V)} = 0 \quad and \quad \frac{N_k - N}{N_k} = 1 \tag{11}$$

leading to the minimum value of E_m. The maximum value of E_m occurs when:

$$I(U,V) = I_{max}(U,V) \quad and \quad N = N_k \tag{12}$$

The metric E_m is a measure of the information flow during the tracking process. Large values of E_m represent large amounts of information carried from the reference region to the target output region. In this case, the similarity between the reference region and the target region and, consequently, the reliability of the tracker output are high. Small values of E_m are an indication that the tracking process is unreliable.

3.2 Kullback-Leibler distance based tracking metric

The Kullback-Leibler distance is defined as [25]:

$$D(p(u)||p(v)) = \sum_{i=1}^{N_{max}} p(u_i) \log_2 \frac{p(u_i)}{p(v_i)} \tag{13}$$

and measures the similarity between $p(u_i)$ and $p(v_i)$. It is not symmetric, i.e. in general $D(p(u)||p(v)) \neq D(p(v)||p(u))$. An upper bound of the Kullback-Leibler distance can be easily found as follows, since:

$$D(p(u)||p(v)) = \sum_{i=1}^{N_{max}} p(u_i) \log_2 \frac{p(u_i)}{p(v_i)} = \sum_{i=1}^{N_{max}} p(u_i) \log_2 p(u_i) - \sum_{i=1}^{N_{max}} p(u_i) \log_2 p(v_i) \tag{14}$$

The first term is negative or zero, while the second is positive. Therefore, an upper bound of the Kullback-Leibler distance is:

$$D(p(u)||p(v)) \leq - \sum_{i=1}^{N_{max}} p(u_i) \log_2 p(v_i). \tag{15}$$

A similar metric to $E_m(U,V,N,N_k)$ based on the Kullback-Leibler distance can be defined as:

$$E_K(U,V,N,N_k) = c_1(1 - \frac{D(p(u)||p(v))}{D_{max}(p(u)||p(v))} - \lambda_1 \frac{N_k - N}{N_k} + c_2) \tag{16}$$

and by construction is expected to behave similarly to E_m. Large values of E_K imply a better matching between the reference and the target region. Both Mutual information and Kullback-Leibler tracking metrics are expected to perform best when we have planar object motion with partial and total occlusions.

3.3 Normalized correlation based metric

The normalized correlation between the reference and the target feature point sets can be defined as: [26]:

$$C_n = \frac{\sum_{i=1}^{N} J_1(\mathbf{x}_i^r) J_2(\mathbf{x}_i^c)}{\sqrt{\sum_{i=1}^{N} J_1^2(\mathbf{x}_i^r) \sum_{i=1}^{N} J_2^2(\mathbf{x}_i^c)}} \tag{17}$$

since a one by one correspondence exists between the feature point sets. Equation (17) expresses the similarity between J_1 and J_2 and can be used to construct a metric similar to those already presented in the context of present work (Eq.8,16). The metric constructed is of the form:

$$C_n^{'} = c_1(C_n - \lambda_1 \frac{N_k - N}{N_k} + c_2) \tag{18}$$

and was also tested under similar tracking conditions with the other two. It stands that:

$$0 \leq C_n^{'} \leq 1 \tag{19}$$

The values of the constants c_1,c_2,λ_1 are the same as in equations (8), (16).

3.4 Tracker initialization evaluation metric

Since the feature point set S_1 generated on the initial frame belongs to the power set of the possible feature point set configurations, a metric measuring the reliability of S_1 can be defined. It can characterize the efficiency of the initially selected region for tracking. Each feature point set S_k is characterized by its entropy:

$$H_{S_k} = -\sum_{i=1}^{N_{max}} p_k(u_i) log_2 p_k(u_i), \tag{20}$$

where $u = J(\mathbf{x})$ are the image luminances at feature point locations on the initial frame. Let N_k be the number of feature points generated in the tracked region. In general, $N_k \leq N_s$. The maximal value of H_{S_k} depends on N_k, if $N_k \leq N_{max}$, since in that case the number of grayscale levels, belonging to the feature point set, cannot reach N_{max}. Then the distribution $p_k(u_i) = \frac{1}{N_k}$ can create an upper bound H_{S_k} if $N_k \leq N_{max}$. Therefore:

$$p_k(u_i) = \begin{cases} \frac{1}{N_k} & N_k \leq N_{max} \\ \frac{1}{N_{max}} & N_k > N_{max} \end{cases} \tag{21}$$

Clearly H_{S_k} is maximized when $N_k \geq N_{max}$ and $p_k(u_i) = \frac{1}{N_{max}}$. The maximal symbol value of the communication alphabet is N_{max} (maximum number of grayscale levels). In order to handle degenerative cases, where the number of the generated feature points N_k is much smaller than the initial user preference N_s, a penalizing term depending on the number of not generated feature points is added. Such cases occur when the minimum allowed distance between feature points is large, compared to the region size. Therefore, the metric, measuring the efficiency of the feature point sets produced during the initialization step, is defined as:

$$E_i(H_{S_k}, N_k, N_s) = \begin{cases} \frac{H_{S_k}}{log_2 N_k} & N_T \leq N_k < N_{max} & \\ \lambda_H \frac{H_{S_k}}{log_2 N_k} + \lambda_F \frac{N_k}{N_s} & N_k < N_{max}, & N_k < N_T \\ \frac{H_{S_k}}{log_2 N_{max}} & N_T \leq N_k, & N_{max} \leq N_k \\ \lambda_H \frac{H_{S_k}}{log_2 N_{max}} + \lambda_F \frac{N_k}{N_s} & N_{max} \leq N_k, & N_k < N_T \end{cases} \tag{22}$$

Threshold N_T is usually a fraction of the user specified feature point number N_s. In the context of present work we have chosen: $N_s = 180$, $N_T = \frac{N_s}{4}$, $\lambda_H = 0.5$, $\lambda_F = 0.5$. The penalizing term is introduced only when $N_k < N_T$. In such cases the number of the feature points N_k is small and the penalizing term

of equation (22) has to be added. The metric E_i is a measure of efficiency of the initial feature point set configuration. It imposes a feature point selection based on the feature point set entropy. The initialization metric imposes large feature point set luminance variation by using entropy maximization. The most effective way of controlling the feature point set configuration is by changing the minimum distance between the feature points. Small distances lead to a feature point concentration in certain parts of the object being tracked. Larger distances usually help at providing feature point sets with better coverage of the object being tracked and at attaining better tracking results. Ideally, the average feature point distance should be greater than the texture cell or grain size.

The entropy based selection criterion aims at imposing a large feature point intensity dispersion in order to provide better tracking results. The penalizing term is introduced to prevent a feature point set choice with too large distances between the feature points that contains a small number of feature points. The initialization criterion can also be applied to untextured objects with limited success. The choice of the initial feature point set configuration is important for the success of the object tracking process.

4 Tracking algorithm enhancement

The tracking algorithm presented in section 2 is enhanced by using an occlusion handling scheme. It is capable of handling partial and total occlusion in a variety of cases. The Occlusion handling scheme is assisted by an object verification scheme, applied to total occlusion situations. The object verification scheme is based on the metric E_m, in the context of present work. Nevertheless, other techniques like elastic graph matching can also be used.

4.1 Occlusion handling

In order to cope with partial occlusion, a prediction scheme is applied. The lost features are not tracked. However, their coordinates are updated using the estimated movement of the upper left and the lower right corner of the bounding rectangle of the tracked object. The procedure is stopped if the occlusion is total, that is, when none of the feature points comprising the feature point set can be further tracked correctly due to occlusion. In order to handle large variations of the bounding box size, caused by the feature point loss, the area of the tracked region is introduced as a reliability measure of the update of the upper left and lower right bounding box coordinates. The feature points, whose coordinates are updated, are considered lost if the bounding box area exceeds a threshold T_{max} or is smaller than a threshold T_{min}. Periodical regeneration of the lost feature points during the tracking process using the procedure presented in section 2 is also a useful tool in order to handle partial occlusion and allow tracking for long time periods. The feature points not lost in the tracking process are not regenerated. In order to cope with total occlusion, the position of the 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 [27]. 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, [27]:

$$\mathbf{s}(k) = \mathbf{C}\mathbf{s}(k-1) + \mathbf{w}(k), k = 1, \ldots, N \tag{23}$$

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:

$$\mathbf{s} = \begin{bmatrix} d_x & d_y & u_x & u_y & a_x & a_y \end{bmatrix}^T. \tag{24}$$

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

$$\mathbf{d}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{v}(k), k = 1, \ldots, N \tag{25}$$

where $\mathbf{v}(k)$ denotes a zero-mean, white observation noise sequence. The matrices describing the model are given below. The 2x1 observation vector and the 2x6 measurement matrix are given by:

$$\mathbf{d} = \begin{bmatrix} d_x \\ d_y \end{bmatrix} \tag{26}$$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{27}$$

The observation equation states that the noisy displacement coordinates of each bounding box corner can be observed.
The 6x6 state transition matrix describing the model is [27]:

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{28}$$

4.2 Object reappearance prediction and verification

Reappearance prediction is obtained by estimating and tracking the *occluding* region. To estimate the occluding region bounding box, a simple region growing segmentation algorithm is used. A seed is determined by the last position of the occluded region before total occlusion. The occluded object is considered to be entirely disoccluded when:

$$A_1 \bigcap A_2 = \emptyset, \tag{29}$$

where A_1 is the occluding region and A_2 is the predicted occluded region. The occluded region reappears provided that the above condition is satisfied. Reappearance is associated with the regeneration of a set

of feature points, as in the initialization step. Again, the selected feature points for object reappearance are those that have large eigenvalues of matrix $\mathbf{Z}$. The feature point regeneration is thorough after total occlusion, that is the entire feature point set is regenerated inside the bounding rectangle specified by A_2. Object tracking continues after the feature point set regeneration.

When the tracker predicts that reappearance has taken place, it has to decide if the reappearing region is similar enough to the tracked region before occlusion. This can be achieved by using the mutual information metric E_m (8). A feature point set is generated in the tracked region belonging to the frame before total occlusion. The position of this feature point set on the current frame is predicted. The metric E_m is calculated using the feature point sets of the reference and target frames, while the predicted tracked region is allowed to change slightly. The maximum of the E_m value is compared with a threshold. The threshold value can be chosen according to the value of E_m before total occlusion.

Graph matching is an alternative technique that can be used for object reappearance verification. Nevertheless, the use of E_m as previously described is preferred, in the context of present work for simplicity and uniformity.

5 Experimental results

The proposed tracking algorithm was tested on both real and artificially generated image sequences. In order to evaluate the efficiency of the proposed scheme, image sequences containing total occlusion and partial occlusion were used. Curves showing the variations of the metrics E_m, E_K and $Corr$ during the tracking process were calculated for different occlusion cases. The metric E_i of the tracking algorithm initialization efficiency, was tested both on artificial and real image sequences.

The algorithm involves the choice of system parameters in order to work. The parameters' values are kept constant during the experiments. The choice of N_s is left to the user and depends mainly on the tracked object size. N_T is a fraction of N_s acquired by experience. The choice of c_1, c_2, λ_1,$\lambda_F$and λ_H is imposed by the requirement $0 \leq E_m \leq 1$ and $0 \leq E_i \leq 1$. Their value is kept throughout the entirety of experiments. The choice of the minimum distance between feature point is crucial to the tracking process and is obtained by using the initialization metric E_i.

Results on an artificial image sequence are presented in Figure 1. A small circular object (Fig. 1a) moves slowly from right to left and is fully occluded by a faster moving elliptical object that moves in the opposite direction (Figure 1b). The tracked region bounding rectangle is recalculated after total disocclusion. The algorithm performs well, even when the occluding object reappears suddenly without previous appearance in the image sequence. The cost function E_m for various image sequence frames is shown in Figure 2. A decrease in E_m begins after frame 15, marking the beginning of partial occlusion. The minimal value of $E_m = 0$ marks the beginning of total occlusion. Object reappearance is marked by an abrupt increase of E_m. The frames corresponding to the start of partial occlusion and the start of total occlusion are shown in Figure 3.

In Figure 4, the tracked region (head of the football player) is occluded by the foot of another football player. The cost function E_m for each frame of the image sequence is shown in Figure 5. E_m drops

at its minimum value $E_m = 0$ during total occlusion. The results on the real and the artificial image sequences show that E_m can be useful in the analysis of partial and total occlusion in object tracking. Partial occlusion is accompanied by a drop of E_m, while total occlusion is characterized by a zero E_m value. The sudden increase of cost function E_m after the object reappearance in the example of Figure 4 is caused by the generation of a feature point set during the object reappearance described in section 3.6. An increase of E_m is possible, whenever a feature point set regeneration occurs.

In Figure 6, results showing robustness to partial occlusions are presented. A person face is partially occluded and, at the end of partial occlusion, the tracked face reappears completely. The beginning of partial occlusion in frame 34, (Figure 7) is marked by a sudden drop in E_m (Figure 12). The mutual information does not increase after face disocclusion, since many feature points were lost during partial occlusion that have not been regenerated after disocclusion. Two frames showing the feature point sets before and after partial occlusion are presented in Figure 8. Notice the loss of feature points, which is caused by partial occlusion. The tracked region size is computed correctly with the help of partial occlusion handling scheme.

The object tracking algorithm containing the occlusion handling scheme and the object reappearance prediction and verification scheme performs better than an object tracking algorithm based on [17] without these new additions. In Figure 9 results of [17] without the new additions on the football image sequence are presented. Notice the performance degradation before total occlusion and the loss of target after total occlusion versus the results shown in Figure 4. Similar results on the artificial image sequence are presented in Figure 10. Performance degradation before total occlusion and loss of target after total occlusion is also noticed, when compared with the results shown in Figure 3 . Results on the lab image sequence are also presented in Figure 11. Partial occlusion affects the tracking performance. One part of the tracked object is lost during and after partial occlusion, as can be seen in Figure 11, in contrast to what is shown in Figure 6.

The variations of the metric E_m for the three sequences are presented in Figures 2,5,12 respectively, while the metric E_k variations are presented in Figures 13,14,15. Finally, the variations of the $Corr$ metric are presented in Figures 17, 18 and 19. As it can be seen metric E_k performs similarly to E_m. Further tests have shown that no significant change in E_k behavior was caused by its asymmetry (Fig. 16). The normalized correlation based metric $Corr$ does not behave as well as the information theory based metrics in partial occlusion situations (Figures 17,18 and 19). The authors believe that the information theory based metrics should be preferred over the normalized correlation one. Mutual information can be very useful as it provides spatial information and is symmetrical. The Kullback-Leibler distance can provide a variety of metrics with similar performance.

The variations of the initialization performance metric E_i (22) with respect to the minimum allowed distance in pixels between feature points in the reference frame are presented in Figures 20 and 21 for the artificial image sequence and the football image sequence respectively. The cost function values are generally bigger in the football image sequence than in the artificial image sequence case due to the fact that the initialized region in the artificial image sequence is uniformly textured. The value of E_i increases when the minimum allowed distance between features increases, provided that $N_k \cong N_s$. A rapid decrease

in the E_i value is noticed when the minimum allowed distance between feature points increase causes the number of feature points generated in the tracking region be much smaller than the initial feature point number user preference ($N_k \ll N_s$).

The effectiveness of the proposed tracker initialization metric was tested by performing object tracking in the football and artificial image sequences under different minimum feature points distances. The algorithm performs well when the minimum between feature points distance in the football image sequence case (Figure 22) is less than 5 pixels. A rapid decrease in performance was noticed when the feature points distance increased above 5 pixels. Tests performed on the artificial image sequence case have shown no significant change in the algorithm performance for feature point distances in the range $[3, \ldots, 7]$ pixels. A decrease in the algorithm performance was noticed for feature point distance arround 10 pixels. It can be noticed in the artificial image sequence that the "best" 5 pixel value is equal to the texture grain size. This exhibits a possible relationship between the texture grain size and the feature point distance.

The effectiveness of the partial occlusion handling scheme during the tracking process is shown in Figures 23 and 24. In Figure 23, the loss of feature points is caused by partial occlusion. Figure 24 demonstrates the usefulness of the updating procedure in a case not containing partial occlusion, since loss of feature points can be caused by illumination changes, deformations of the tracked objects, abrupt motion or a combination of them.

6 Conclusions

In this paper, an object tracking algorithm that is robust to partial and full occlusion was presented. Information theory based metrics were used as a reliability measure to the algorithm initialization and tracking procedures. The mutual information and Kullback-Leibler based metrics provide the means to detect abrupt changes, (partial occlusion, full occlusion or movement of the occluding object). Furthermore, motion detection of the tracked object is also possible in static scenes. Finally, an object verification process based on mutual information was also proposed and applied after object disocclusion. The use of the information theory based metrics combined with an occlusion handling scene provide an object tracking algorithm performing better than [17] in partial and total occlusion situations.

Experimental results have shown that the algorithm correctly detects and processes partial and total occlusion situations. The interpretation of variations of the proposed metrics may lead to a thorough understanding of the object tracking process in many computer vision applications.

The information theory based metrics behave better in partial occlusion situations than the normalized correlation based metric. A clear distinction in performance between the two information theory based metrics cannot be easily extracted. Nevertheless, the mutual information having the advantage of being symmetrical and including spatial information seems to be the preferred choice.

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*, 20(10):1025–1039, 1998.

[2] N. Peterfreund. Robust tracking of position and velocity with kalman snakes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(6):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*, 22(5):544–549, 2000.

[4] U. Uenohara and T. Kanade. Geometric invariants for verification in 3-d object tracking. In *International Conference on Intelligent Robots and Systems '96 IROS 96 Proceedings of the 1996 IEEE/RS*, volume 2, pages 785–790, 1996.

[5] S. Manku, P. Jain, A. Aggarwal, L. Kumar, and S. Banerjee. Object tracking using affine structure for point correspondences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997.

[6] D. P. Huttenlocher, J. J. Noh, and W.J. Rucklidge. Tracking non rigid objects in complex scenes. In *Proceedings of the International Conference on Computer Vision*, pages 93–101, 1993.

[7] H. T. Nguyen, M. Worring, and R. van den Boomgaard. Occlusion robust adaptive template tracking. In *Proceedings of the International Conference on Computer Vision*, volume I, pages 678–683, 2001.

[8] S. L. Dockstader and A. M. Tekalp. Multiple camera tracking of interacting and occluded human motion. *Proceedings of the IEEE*, 89(10):1441–1455, 2001.

[9] C. Rasmussen and G. D. Hager. Probabilistic data association methods for tracking complex visual objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):560–576, 2001.

[10] T. Schoepflin, V. Chalana, D. R. Haynor, and Y. Kim. Video object tracking with a sequential hierarchy of template deformations. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(11):1171–1182, 2001.

[11] C. Erdem, A. M. Tekalp, and B. Sankur. Metrics for performance evaluation of video object segmentation and tracking without ground truth. In *Proc. of 2001 Int. Conf. on Image Processing*, volume II, pages 69–72, 2001.

[12] C. E. Erdem, A. M. Tekalp, and B. Sankur. Video object tracking with feedback of performance measures. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(4), 2003.

[13] D. Fleet J. Barron and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[14] P. Viola and W. M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[15] H. Kruppa and B. Schiele. Context-driven model switching for visual tracking. In *9th International Symposium on Intelligent Robotic Systems, Toulouse, France.*, 2001.

[16] H. Kruppa and B. Schiele. Using mutual information to combine object models. In *8th International Symposium on Intelligent Robotic Systems 2000, Reading, UK.*, 2000.

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

[18] S. Birchfield. *Depth and Motion Discontinuities.* Ph.D. Thesis, Stanford University, 1999.

[19] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1998.

[20] J. Wang and E. Adelson. Representing moving images with layers. *IEEE Transactions on Image Processing*, 3(5):625–638, 1994.

[21] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 593–600, 2000.

[22] S. Haykin. *Communication Systems-3rd ed.* J. Wiley, 1994.

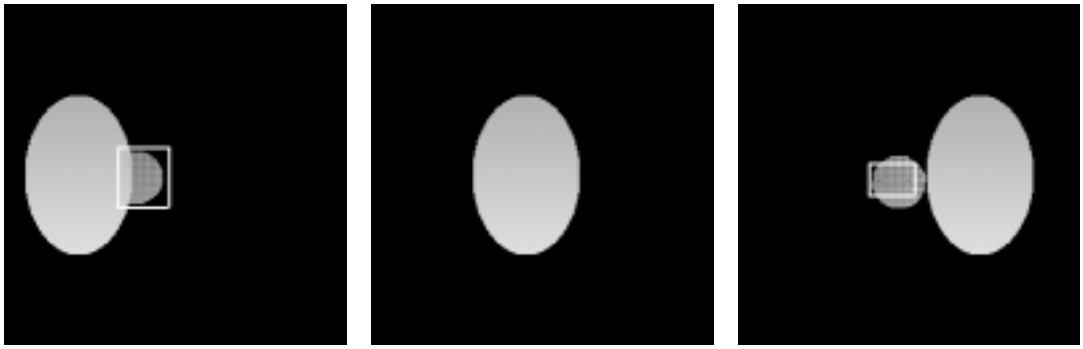
[23] F. M. Reza. *An introduction to information theory.* Dover, 1994.

[24] M. Skouson, Q. Guo, and Z. Liang. A bound on mutual information for image registration. *IEEE Transactions on Medical Imaging*, 20(8):843–846, 2001.

[25] M. N. Do and M. Vetterli. Texture similarity measurement using kullback-leibler distance on wavelet subbands. In *Proc. of 2000 Int. Conf. on Image Processing*, 2000.

[26] A. Papoulis. *Probability, Random Variables, and Stochastic processes.* Mc Graw-Hill, Inc, 1991.

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



(a) (b) (c)

Figure 1: Artificial image sequence: (a) before total occlusion, (b) during total occlusion, (c) region reappearance.

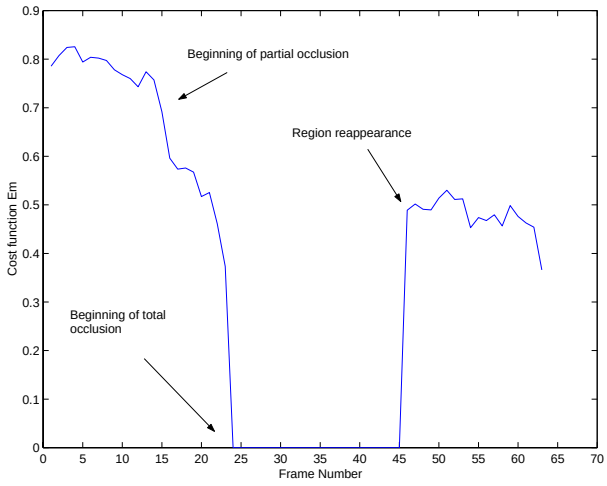
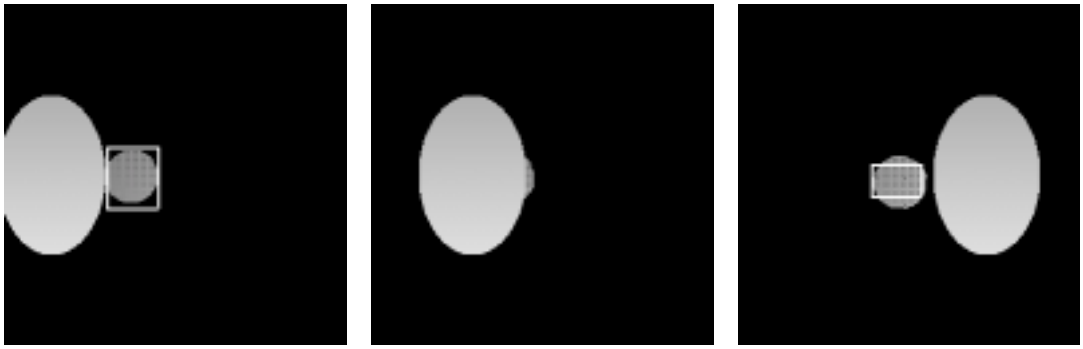


Figure 2: Cost function E_m versus frame number of the artificial image sequence of Fig. 1.



(a) (b) (c)

Figure 3: Artificial image sequence frames characterized by the mutual information cost function as: (a) the start of partial occlusion frame (frame No. 15), (b) the start of total occlusion frame (frame No. 23), (c) the first frame after the total object reappearance (frame No. 45)

(a) (b) (c)

Figure 4: Football image sequence: (a) before total occlusion, (b) during total occlusion, (c) region reappearance.

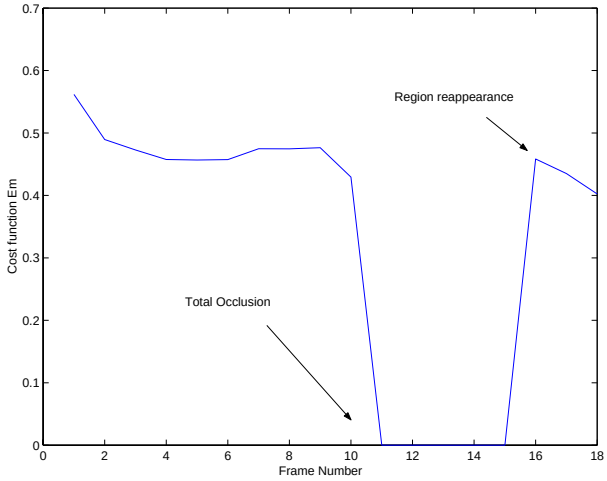
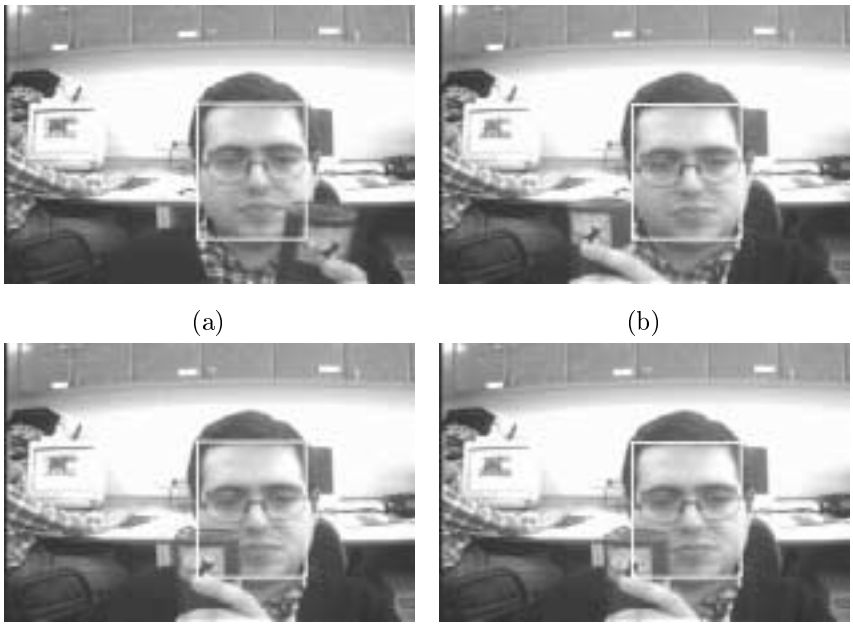


Figure 5: Values of the cost function E_m versus frame number for part of the football image sequence (Fig. 4).

(a) (b) (c)

Figure 6: Lab image sequence: (a) tracked region, (b) partial occlusion, (c) region after occlusion.



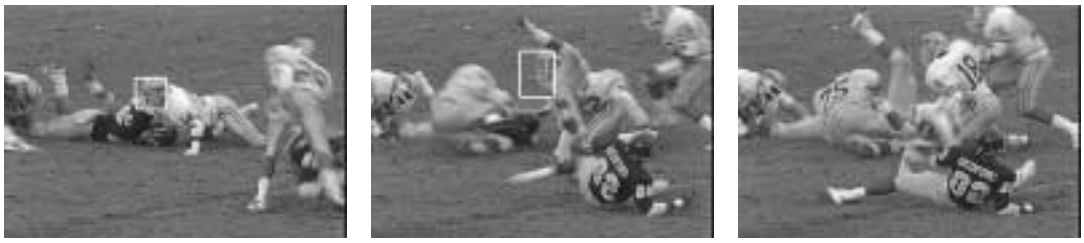
(a) (b)

(c) (d)

Figure 7: Lab image sequence: (a) beginning of partial occlusion (frame 34), (b) disocclusion (frame 101), (c) movement of the occluding region (frame 85), (d) movement of the occluding region (frame 86).

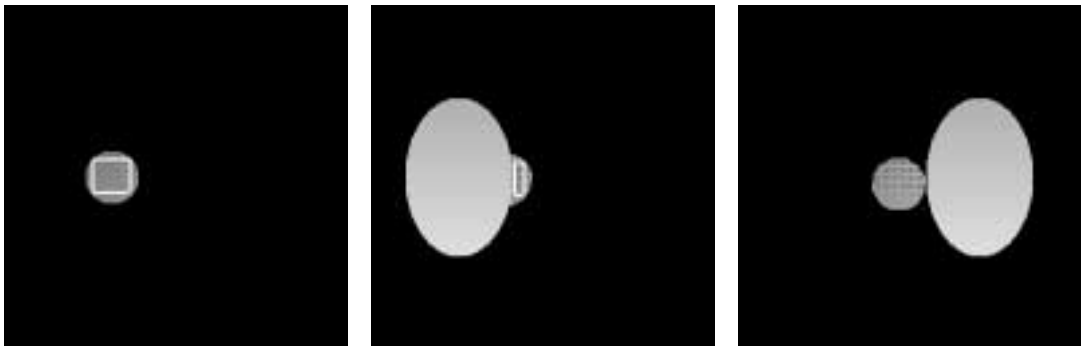
(a) (b)

Figure 8: Lab image sequence: (a) Feature point set before partial occlusion, (b) Feature point set after partial occlusion



(a) (b) (c)

Figure 9: Football image sequence: Tracking without occlusion handling and object reappearance prediction and verification. (a) initial frame, (b) before total occlusion, (c) after total occlusion.



(a) (b) (c)

Figure 10: Artificial image sequence: Tracking without occlusion handling and object reappearance prediction and verification. (a) initial frame, (b) before total occlusion, (c) after total occlusion. Notice tracking degradation in (b) and in (c).

(a) (b) (c)

Figure 11: Lab image sequence: Tracking without occlusion handling and object reappearance prediction and verification. (a) initial frame, (b) during partial occlusion, (c) after partial occlusion. Notice tracking degradation in (b) and in (c).

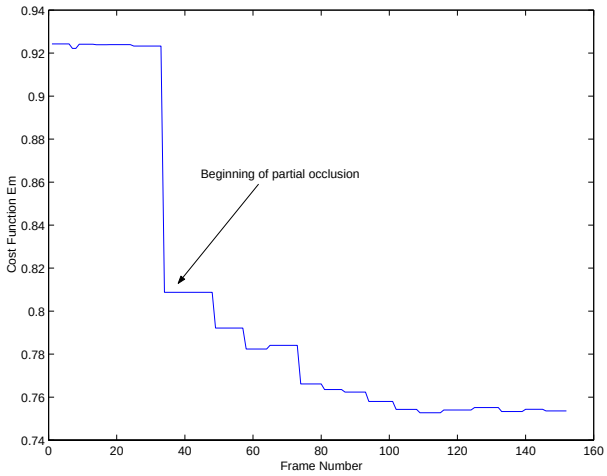


Figure 12: Cost function E_m for the lab image sequence (Fig. 7) versus frame number.

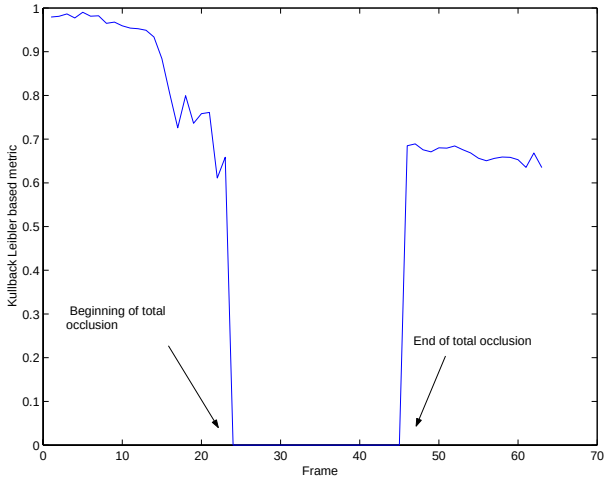


Figure 13: Cost function E_K for the artificial image sequence (Fig. 3) versus frame number.

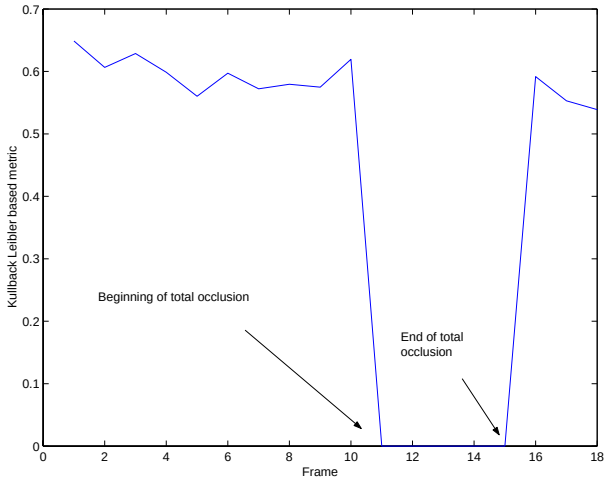


Figure 14: Values of the cost function E_K versus frame number for part of the football image sequence (Fig. 4).

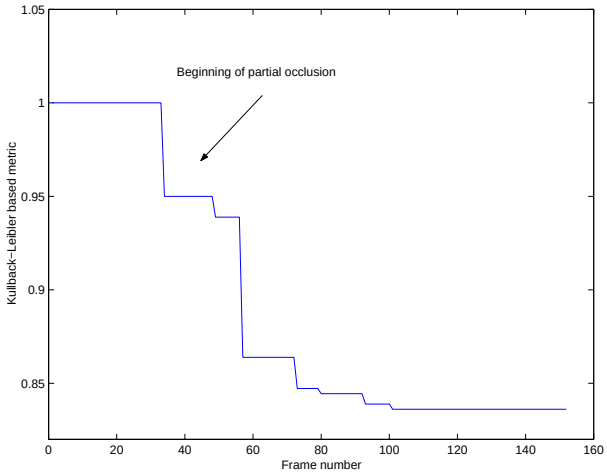


Figure 15: Values of the cost function E_K versus frame number for the lab image sequence (Fig. 7).

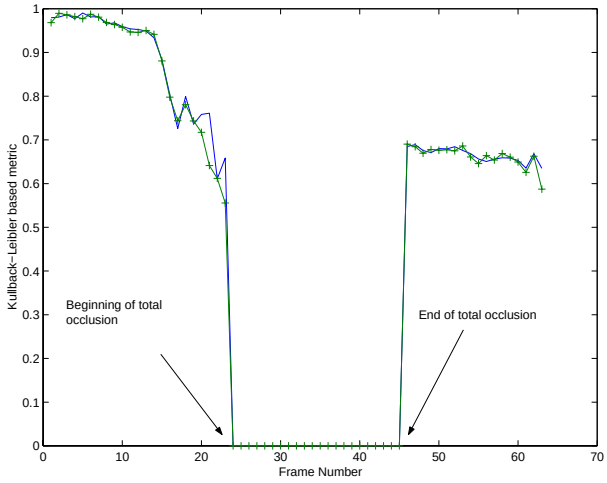


Figure 16: Values of the two forms of the cost function E_K based on $D(p(u)||p(v))$ and $D(p(v)||p(u))$ for the artificial image sequence (Fig. 3).

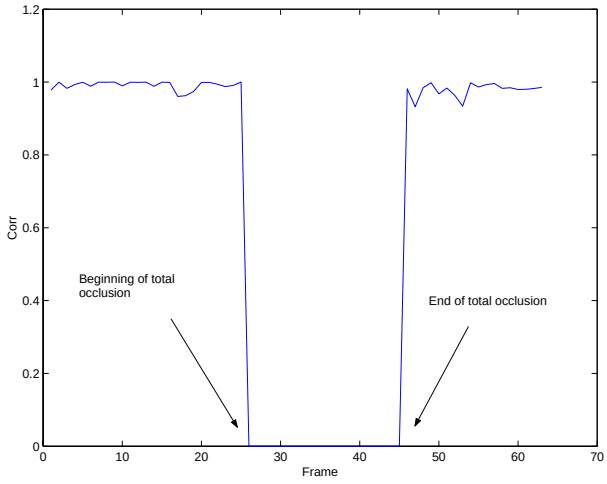


Figure 17: Normalized Correlation for the artificial image sequence (Fig. 3) versus frame number.

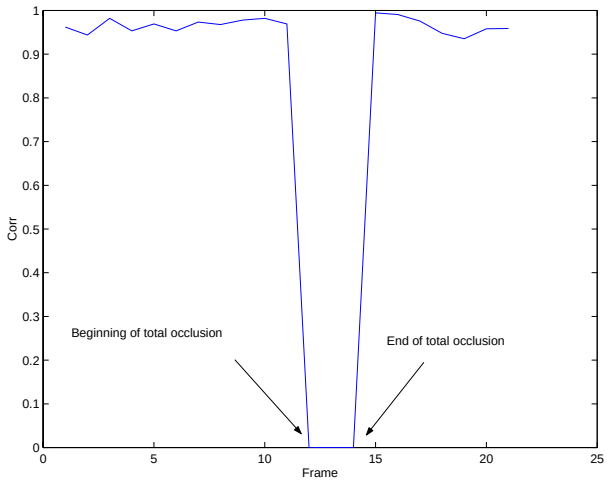


Figure 18: Normalized Correlation for the football image sequence (Fig. 4) versus frame number.

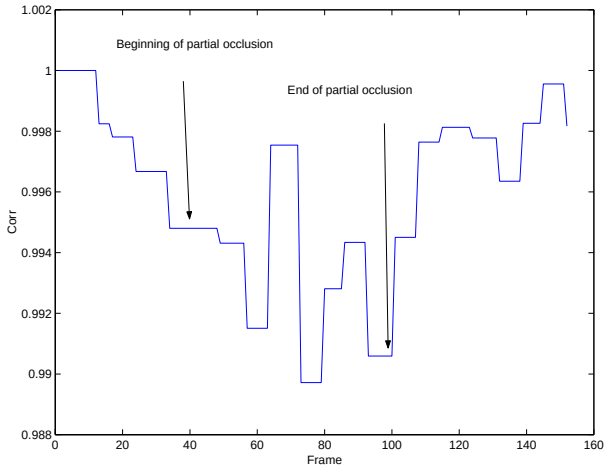


Figure 19: Normalized Correlation for the lab image sequence (Fig. 7) versus frame number.

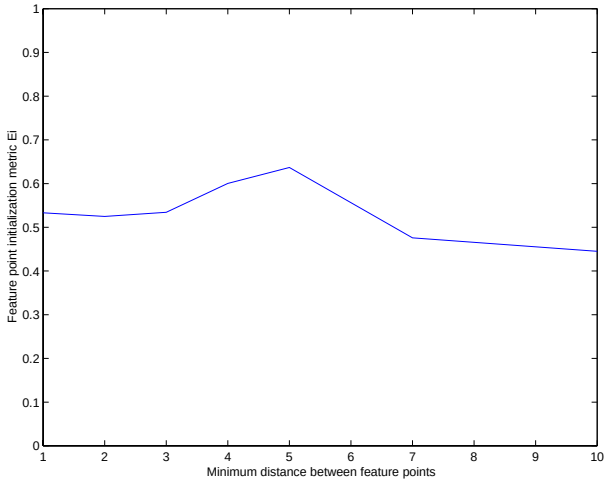


Figure 20: Cost function E_i for the algorithm initialization in the artificial image sequence. Notice that the texture grain size is 5 pixels. (Fig. 1, 2).

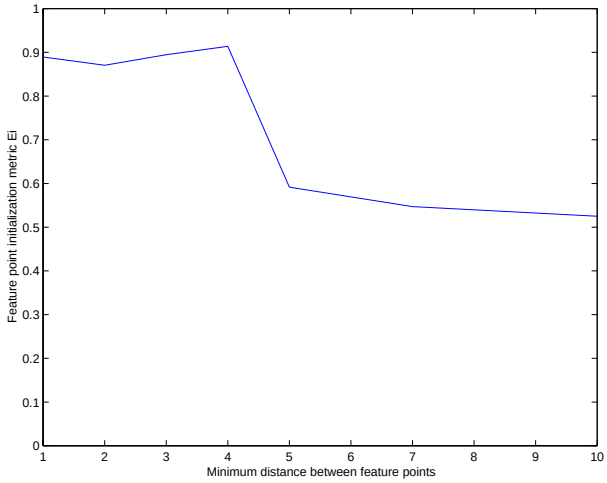


Figure 21: Cost function E_i for the initialization process in the football image sequence (Fig. 4).

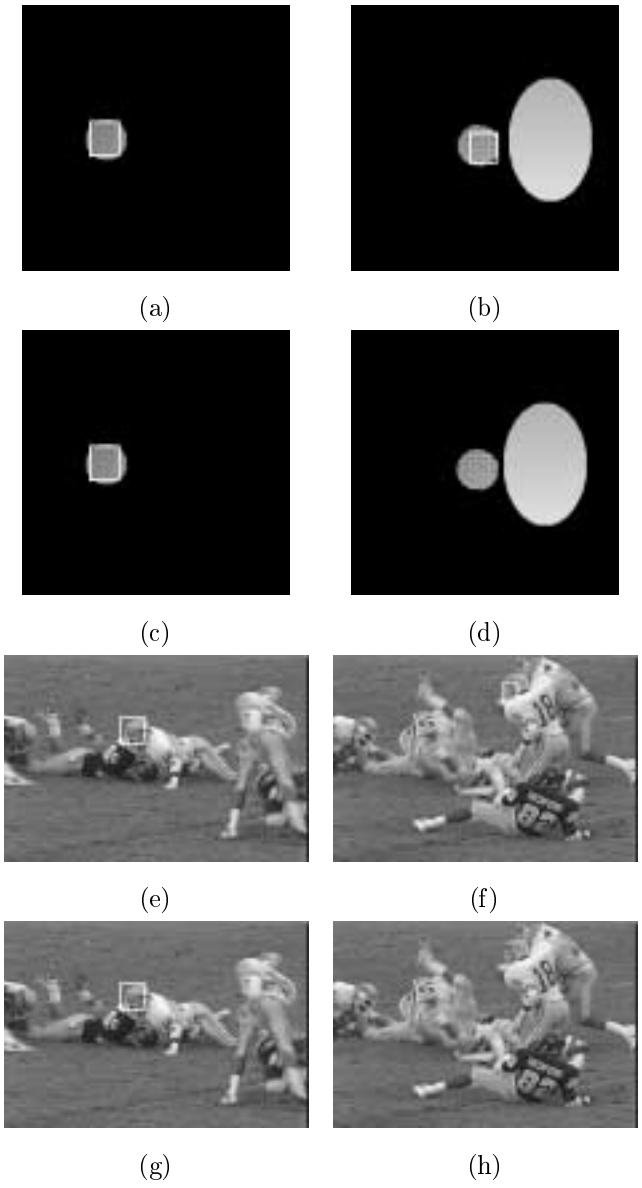
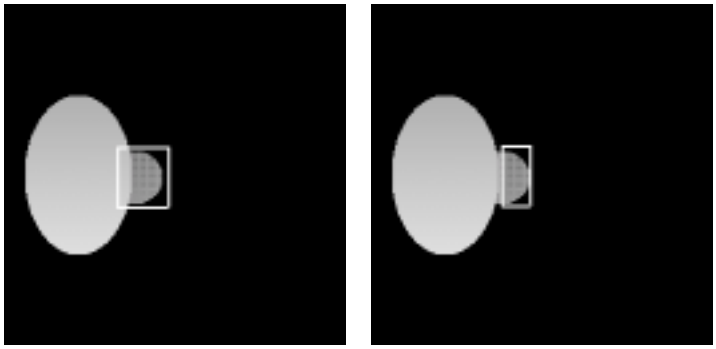
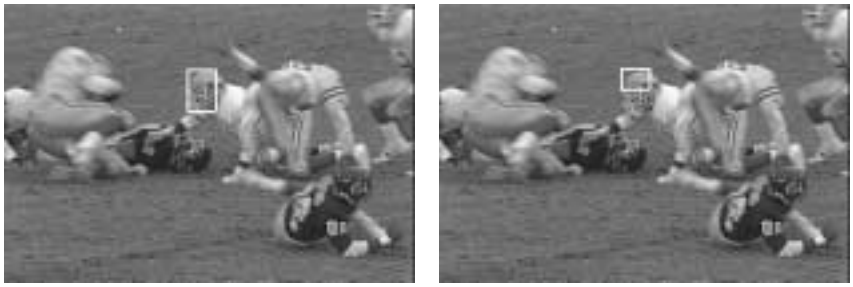


Figure 22: Tracker outputs for the artificial image sequence: (a) and (b) with minimum distance between feature points equal to 5 pixels, (c) and (d) with minimum distance between feature points equal to 10 pixels. Football image sequence I: Tracker output obtained: (e) and (f) with minimum distance between feature points equal to 4 pixels, (g) and (h) with minimum distance between feature points equal to 5 pixels Notice the performance degradation at (c),(d) and (g),(h).



(a) (b)

Figure 23: Tracker outputs obtained: (a) with and (b) without applying the partial occlusion handling scheme in a frame of the artificial image sequence.



(a) (b)

Figure 24: Tracker outputs obtained: (a) with and (b) without partial occlusion handling scheme in the football image sequence.)