INFORMATION THEORY-BASED ANALYSIS OF PARTIAL AND TOTAL OCCLUSION IN OBJECT TRACKING

E. Loutas *C. Nikou* *I. Pitas*

Department of Informatics
University of Thessaloniki
Box 451, Thessaloniki 540 06
GREECE
E-mail: {eloutas,nikou,pitas}@zeus.csd.auth.gr

ABSTRACT

Metrics based on mutual information and not resorting to ground truth data are proposed in this paper in order to measure tracking reliability under occlusion. The variations of the proposed metrics can be used as a quantitative estimate of changes in the tracking region, caused by occlusion, sudden movement or the deformation of the tracked object. The proposed metric 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

The measure of the performance of a tracking algorithm is considered an open issue in the tracking research community. Most of the tracking techniques use subjective evaluation methods, while some of them use quantitative evaluations based on ground truth data [1]. The implementation of reliability measures not resorting in 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 [2]. Measuring tracker performance is important in cases of rapid performance degradation such as partial or total occlusion. Different algorithms handling occlusion were presented [3, 4, 5, 6, 7, 8, 9]. Nevertheless, quantitative measures of the tracker performance are not generally proposed. Mutual information was first introduced in computer vision in [10]. In [11, 12] it was applied to combine the outputs of multiple tracking algorithms in order to improve the overall tracker performance.

The use of mutual information in object tracking as a tool for extracting information concerning the condition of the object being tracked is assessed in this paper. The tracking process is modeled as a communication between a transmitter and a receiver through a channel. Two metrics based on mutual information are introduced: The first metric introduced is used as a quantitative measure of the tracking process. Moreover, its variations can increase user understanding over the behavior of the tracking region and are closely related with changes in the that 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 the tracker performance and assist an event detection scheme. The second metric introduced is used as a quantitative measure of the tracking algorithm initialization efficiency and is closely related to the metric used as a measure of the tracking process.

Object tracking is performed by minimizing the sum of squared differences of a large set of feature points generated in the tracking region presented in [13]. The algorithm presented in [14] is used for feature point tracking. Kalman filtering motion prediction is employed to estimate the tracked region position during occlusion. Robustness to partial occlusion is achieved by estimating the motion of the lost feature points, using the estimated motion of the bounding box [13].

The main contribution of the presented work is the introduction of mutual information based metrics as measures of tracking reliability. Moreover, the use of the metric is extended to the analysis of partial and total occlusion in object tracking.

This study has been partially supported by the Commission of the European Communities, in the framework of the project IST-1999 20993 CARROUSO (Creating, Assessing and Rendering of High Quality Audio-Visual Environments in MPEG-4 context).

2. MUTUAL INFORMATION METRICS

The 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 tracking process is characterized by loss of information caused by feature point rejection and wrong feature point correspondences [13, 14]. Mutual information is a well established measure of the amount of information transmitted through the communication channel [15, 16]. Therefore, it can be used as a quantitative measure of the tracking process.

2.1. Tracker performance 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

$$S_1 = [\mathbf{x}_1^r, \ldots, \mathbf{x}_N^r]^T \tag{1}$$

is tracked to a feature point set

$$S_2 = [\mathbf{x}_1^c, \ldots, \mathbf{x}_M^c]^T, \tag{2}$$

with $M \leq N_{max}$. Let U, V be two random variables with $p(u), p(v)$ their marginal probability mass functions 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 two random variables U, V with a joint probability mass function $p(u, v)$ is defined as:

$$I(U,V) = \sum_{i=1}^{N_{max}} \sum_{j=1}^{N_{max}} P(u_i, v_j) \log_2 \frac{p(u_i, v_j)}{p(u_i)p(v_j)}, \quad (3)$$

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:

$$E_m(U,V,N,M) = c_1\left(\frac{I(U,V)}{I_{max}(U,V)} - \lambda_1 \frac{N-M}{N} + c_2\right) \quad (4)$$

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 [17]:

$$I_{max}(U,V) = -\sum_{i=1}^{N_{max}} p(u_i) \log_2 p(u_i) \quad (5)$$

The term $\frac{N-M}{N}$ is a penalizing quantity depending on the number of the lost feature points during the tracking process. The constants c_1, c_2, λ_1 are set in order to ensure that:

$$0 \le E_m \le 1. \quad (6)$$

In the case of total occlusion:

$$\frac{I(U,V)}{I_{max}(U,V)} = 0 \quad and \quad \frac{N-M}{N} = 1 \quad (7)$$

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

$$I(U,V) = I_{max}(U,V) \quad and \quad N = M \quad (8)$$

2.2. Tracker initialization performance metric

Since the feature point set S_1 generated on the the initial frame belongs to the set of the possible feature point set configurations, a metric measuring the reliability of S_1 is also defined. This metric characterizes the efficiency of the selected region for tracking.

Each feature point set is characterized by its entropy:

$$H_{S_k} = -\sum_{i=1}^{N_{max}} p_k(u_i) log_2 p_k(u_i), \quad (9)$$

Let N_k be the number of feature points generated in the tracked region. In general, $N_k \le N_s$. H_{S_k} is maximized when

$$p_k(u_i) = \begin{cases} \frac{1}{log_2 N_k} & N_k \le N_{max} \\ \frac{1}{log_2 N_{max}} & N_k > N_{max} \end{cases} \quad (10)$$

The maximal symbol value the communication alphabet can reach 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. Therefore, the proposed metric is formulated as:

$$E_i = \begin{cases} \frac{H_{S_k}}{log_2 N_k} & N_T \le N_k < N_{max} \\ \lambda_H \frac{H_{S_k}}{log_2 N_k} + \lambda_F \frac{N_k}{N_s} & N_k < N_{max},\ N_k < N_T \\ \frac{H_{S_k}}{log_2 N_{max}} & N_T \le N_k, \quad N_{max} \le N_k \\ \lambda_H \frac{H_{S_k}}{log_2 N_{max}} + \lambda_F \frac{N_k}{N_s} & N_{max} \le N_k,\ N_k < N_T \end{cases} \quad (11)$$

Threshold N_T is usually a fraction of the user specified feature point number N_s. The penalizing term is introduced only when $N_k < N_T$, that is, when the number of generated feature points is smaller than a fraction of the user specified feature point number.

3. RESULTS

The proposed metrics were tested using the object tracking algorithm presented in [13]. In order to evaluate the metric efficiency in partial and total occlusion situations, image sequences containing total occlusion and severe partial occlusion were used. Curves showing the variations of metrics E_m and E_i during the tracking process were acquired for different occlusion situations.

In Figure 1, the tracking region (head of the football player) is occluded by the foot of another football player. The value of the cost function E_m for each frame of the image sequence appears in Figure 2. E_m drops at its minimum value during total occlusion.

The face in Figure 3 is partially occluded and at the end of partial occlusion the tracking region remains intact. The beginning of partial occlusion in frame 34 is marked by a sudden drop in E_m. The beginning of disocclusion is also marked by a drop in frame 101. Drops in frames 49, 57, 73, 80, 86, 93 are also important as they mark the movement of the occluding region. Another example of a lab image sequence is presented in Figure 5. The peaks appearing in Figure 6 represent static scenes (little or no motion between the current frame and the previous frame).

The variations of the initialization performance metric E_i (eq 11) with respect to the minimum allowed distance in pixels between feature points in the reference frame are presented in Figure 7 for the football image sequence. The value of E_i generally increases when the minimum allowed distance between features increases. A rapid decrease in the E_i value is noticed when the minimum allowed distance between features increase causes $N_k \ll N_s$.

4. CONCLUSIONS

In this paper, metrics based on mutual information were established as a reliability measure to the object tracking process. The metrics proposed provide also a means to detect abrupt changes. These changes include partial occlusion, full occlusion or movement of the occluding object. Motion detection of the tracked object is also possible in static scenes.

Experimental results show that the variations proposed metrics can be used as a means to acquire a more thorough understanding of the object tracking process in many computer vision applications.

5. REFERENCES

[1] T. Schoepflin, V. Chalana, D. R. Haynor, and Y. Kim, "Video object tracking with a sequential hierarchy of template de-

formations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 11, pp. 1171–1182, 2001.

[2] C. Erdem, A. M. Tekalp, and B. Sankur, "Metrics for performance evaluation of video object segmentation and tracking without ground trouth," in *Proc. of 2001 Int. Conf. on Image Processing*, 2001, vol. II, pp. 69–72.

[3] N. 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.

[4] 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.

[5] B. Galvin, B. McCane, and K. Novins, "Virtual snakes for occlusion analysis," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[6] D. Koller, J. Weber, and J. Malik, "Robust multiple car tracking with occlusion reasoning," in *Proceedings of the European Conference on Computer Vision*, 1994, pp. 189–198.

[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*, 2001, vol. I, pp. 678–683.

[8] S. L. Dockstader and A. M. Tekalp, "Multiple camera tracking of interacting and occluded human motion," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 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*, vol. 23, no. 6, pp. 560–576, 2001.

[10] P. Viola and W. M. Wells, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

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

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

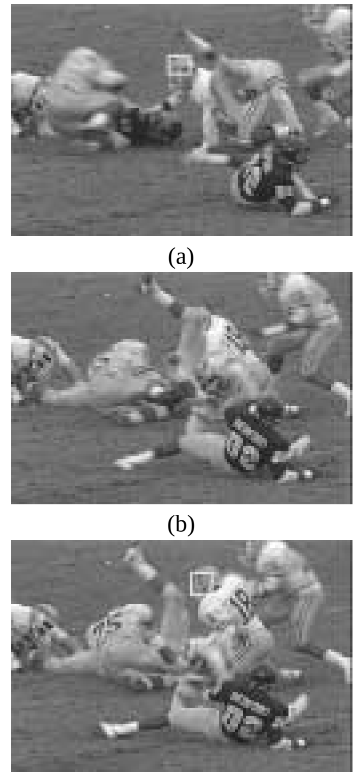
[13] E. Loutas, K. Diamantaras, and I. Pitas, "Occlusion resistant object tracking," in *Proc. of 2001 Int. Conf. on Image Processing*, 2001, vol. II, pp. 65–68.

[14] C. Tomasi and T. Kanade, *Shape and Motion from Image Streams: a Factorization Method - Part 3 Detection and Tracking of Point Features*, 1991.

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

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

[17] M. Skouson, Q. Guo, and Z. Liang, "A bound on mutual information for image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 843–846, 2001.



(a)

(b)

(c)

Fig. 1. Football image sequence. Tracking of the head of the football player: (a) before total occlusion, (b) total occlusion, (c) region reappearance.

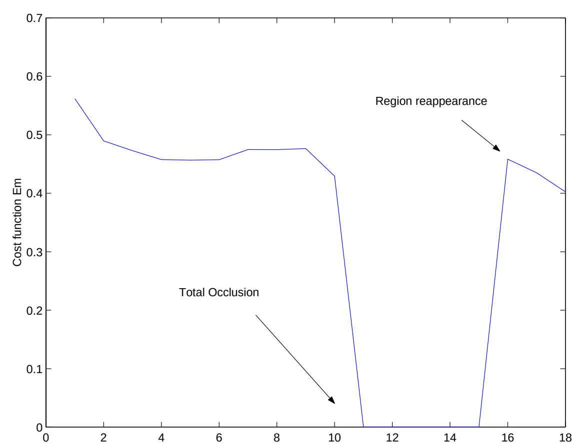


Fig. 2. Values of the cost function E_m for part of the football image sequence (Fig. 1). X-axis: frame number. Y-axis: value of E_m.

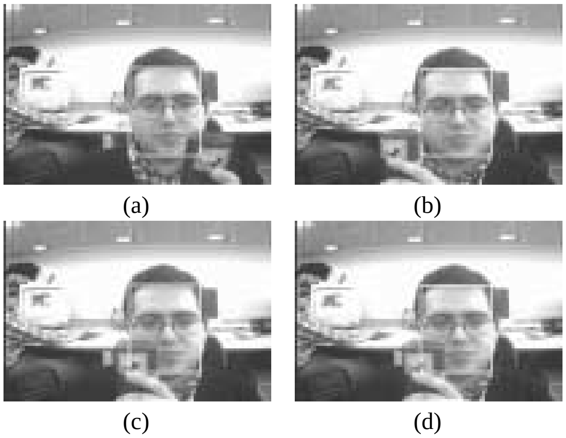


Fig. 3. Lab image sequence I: (a) beginning of partial occlusion (frame 34), (b) disocclusion (frame 101), (c) movement of the occluding region (frame 85), (d) movement of the occluding region (frame 86).

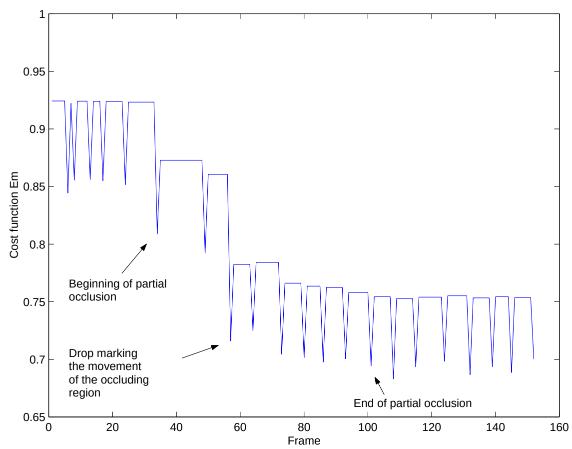


Fig. 4. Values of the cost function E_m for the lab image sequence I (Fig. 3). X-axis: frame number. Y-axis: value of E_m.

Fig. 5. Lab image sequence II: (a) initial frame, (b) partial occlusion.

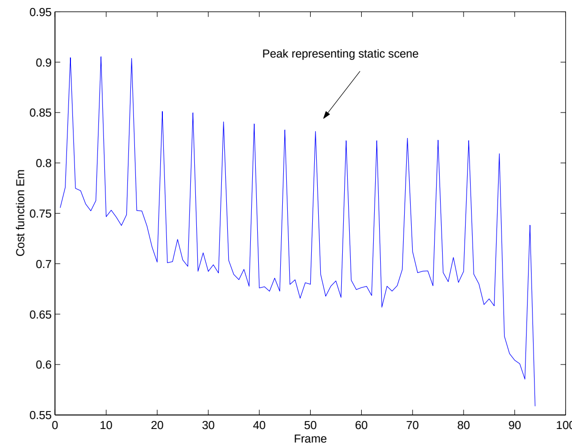


Fig. 6. Values of the cost function E_m for the lab image sequence II 5. X-axis: frame number. Y-axis: value of E_m.

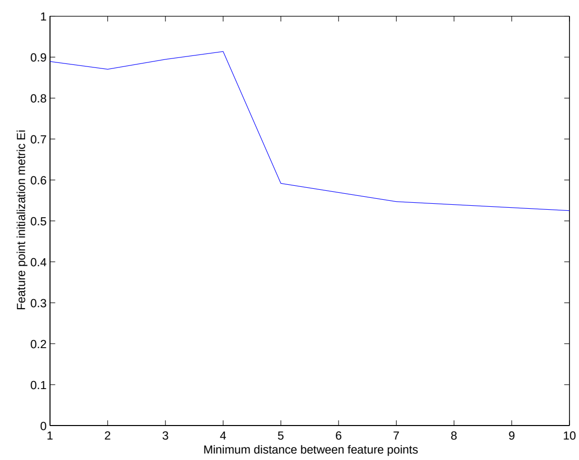


Fig. 7. Values of the cost function E_i for the initialization process in the football image sequence (Fig. 1). X-axis: Minimum distance between feature points in pixels. Y-axis: value of E_i.