EVALUATION OF TRACKING RELIABILITY METRICS BASED ON INFORMATION THEORY AND NORMALIZED CORRELATION

E. Loutas *N. Nikolaidis* *I. Pitas*

Department of Informatics
University of Thessaloniki
Box 451, Thessaloniki 54124
GREECE
E-mail: {eloutas,nikolaid,pitas}@zeus.csd.auth.gr

ABSTRACT

The efficiency of three tracking reliability metrics based on information theory and normalized correlation is examined in this paper. The two information theory tools used for the metrics construction are the mutual information and the Kullback-Leibler distance. The metrics are applicable to any feature-based tracking scheme. In the context of this work they are applied for comparison purposes on an object tracking scheme using multiple feature point correspondences. Experimental results have shown that the information theory based metrics perform better than the normalized correlation one.

1. INTRODUCTION

Measuring 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 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 [2]. A more recent work on those metrics incorporates them in a tracking scheme in order to perform better tracking [3]. A performance measure based on SSD (Sum of squared differences) was introduced in [4], nevertheless it is used as a single feature point tracking performance measure, and is not directly applicable in object tracking. The use of SSD in object tracking can be found in the early works of [5, 6].

Measuring tracker performance is important in cases of rapid performance degradation such as partial or total occlusion. Different algorithms for handling occlusions were presented. However, quantitative measures of the tracker performance not based on ground truth data are not generally proposed.

Evaluation of the efficiency and experimental comparison of three tracker reliability metrics, applicable to feature-based tracking schemes, is performed in this paper. The first metric is based on mutual information and is presented in detail in [7]. In this paper, the implementation of the metric is based on 15×15 windows rather than on single-pixel feature points in order to be more close to existing tracking algorithms. The second metric is novel and is based on the Kullback-Leibler distance [8]. Kullback-Leibler distance was previously used in face detection [9], in sound processing [10] and in video indexing and retrieval, while a fast approximation of Kullback-Leibler distance between two dependence tree models is presented in [11]. Although the Kullback-Leibler distance is similar to the mutual information its asymmetry [10] makes the comparison of the two metrics important. The third metric is based on normalized correlation [12], properly adjusted to particularities of object tracking.

The tracking evaluation metrics are applied for comparison purposes on the same feature based tracking scheme presented in [13]. The scheme performs object tracking by minimizing the sum of squared differences of a large set of feature points generated in the tracking region. The algorithm presented in [6] 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 contributions of present work are the introduction of the new reliability metric based on the Kullback-Leibler distance and the comparison of the mutual information based metric [7], the Kullback-Leibler based metric and the normalized correlation based metric in a unified way.

2. MUTUAL INFORMATION METRIC

Let $\mathbf{x}^r_{i,j}$ and $\mathbf{x}^c_{i,j}$ represent the coordinate vectors of the i^{th} pixel belonging to the 15×15 feature point window j in the reference and current frame, respectively. During the tracking process, a pixel set

$$S_1 = [\mathbf{x}^r_{1,1}, \ldots, \mathbf{x}^r_{W,1} \ldots, \mathbf{x}^r_{W,M_1}]^T \quad (1)$$

is tracked to a pixel set

$$S_2 = [\mathbf{x}^c_{1,1}, \ldots, \mathbf{x}^r_{W,1} \ldots, \mathbf{x}^c_{W,M_2}]^T. \quad (2)$$

where $W = 15^2$ and M_1, M_2 are the number of feature points at the reference and current frame. Let N_T be the maximum number of feature points, which is considered constant during the tracking procedure. Obviously, $M_1 \leq N_T$, $M_2 \leq N_T$. Let also U, V be two random variables expressing grayscale values in the reference and target image with $p(u), p(v)$ their marginal probability mass functions and $u_i = J_r(\mathbf{x}^r_{k,l})$, $v_j = J_c(\mathbf{x}^c_{k,l})$ their possible outcomes, where J_r and J_c are the reference and current image

The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme.

respectively and $\mathbf{x}^r_{k,l} \in S_1, \mathbf{x}^c_{k,l} \in S_2$. The mutual information of two random variables U, V with a joint probability mass function $p(u, v)$ is defined as [14]:

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

where N_{max} is the maximum number of the available grayscale levels. The probability mass functions $p(u)$,$p(v)$ and $p(u, v)$ are empirically determined by obtaining the histograms of the grayscale values of the sets S_1 and S_2. In order to take into account the lost feature points during the tracking process a cost function E_m is defined:

$$E_m(U,V,N_T,M_2) = c_1(\frac{I(U,V)}{I_{max}(U,V)} - \lambda_1 \frac{N_T - M_2}{N_T} + c_2) \qquad (4)$$

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 [15]:

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

The term $\frac{N_T - M_2}{N_T}$ is a penalizing quantity depending on the number of the lost feature points during the tracking process. The constants $c_1 = 0.5, c_2 = 1, \lambda_1 = 1$ are set in order to ensure that:

$$0 \leq E_m \leq 1. \qquad (6)$$

In the case of total occlusion: $p(v)$ and $p(u, v)$ cannot be calculated since $S_2 = \emptyset$ and $\frac{I(U,V)}{I_{max}(U,V)}$ is set to 0 therefore:

$$\frac{I(U,V)}{I_{max}(U,V)} = 0 \quad and \quad \frac{N_T - M_2}{N_T} = 1 \qquad (7)$$

since $M_2 = 0$ leading to the minimum value of E_m. The maximum value of E_m occurs when:

$$I(U,V) = I_{max}(U,V) \quad and \quad N_T = M_2 \qquad (8)$$

3. KULLBACK-LEIBLER DISTANCE BASED TRACKING METRIC

The Kullback-Leibler distance is defined as [8]:

$$D(p(u)||p(v)) = \sum_{i=1}^{N_{max}} p(u_i) \log_2 \frac{p(u_i)}{p(v_i)} \qquad (9)$$

and measures the similarity between $p(u_i)$ and $p(v_i)$. It is always non negative and is not symmetric [10], i.e. in general:

$$D(p(u)||p(v)) \neq D(p(v)||p(u)). \qquad (10)$$

The Kullback-Leibler distance can be symmetrized [10]. A symmetrical form of the Kullback-Leibler distance can be provided as

$$D(p(u)||p(v))_{s_1} = \frac{1}{2}(D(p(u)||p(v)) + D(p(v)||p(u))), \quad (11)$$

$$D(p(u)||p(v))_{s_2} = \sqrt{D(p(u)||p(v))D(p(v)||p(u))}, \qquad (12)$$

or

$$D(p(u)||p(v))_{s_3} = \frac{D(p(u)||p(v))D(p(v)||p(u))}{(D(p(u)||p(v)) + D(p(v)||p(u)))}, \quad (13)$$

Therefore a variety of Kullback-Leibler-based metrics can be derived. An upper bound for the metric in equation (9) can be easily found as follows, since:

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

The first term is negative or zero, while the second is positive. Therefore, an upper bound of the Kullback-Leibler distance (9) is:

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

The upper bounds of the two forms of the symmetrized Kullback-Leibler distance provided are:

$$D(p(u)||p(v))_{s_{1_{max}}} = \frac{1}{2}(D(p(u)||p(v))_{max} + D(p(v)||p(u))_{max}) \qquad (16)$$

for equation (11) and

$$D(p(u)||p(v)))_{s_{2_{max}}} = \sqrt{D(p(u)||p(v))_{max} D(p(v)||p(u))_{max}}. \qquad (17)$$

for equation (12)

A similar metric to $E_m(U, V, N_T, M_2)$ based on the Kullback-Leibler distance can be defined as:

$$E_K = c_1(1 - \frac{D(p(u)||p(v))}{D_{max}(p(u)||p(v))} - \lambda_1 \frac{N_T - M_2}{N_T} + c_2) \quad (18)$$

and by construction is expected to behave similarly to E_m. In the context of present work $c_1 = 0.5$, $c_2 = 1$, $\lambda_1 = 1$. Large values of E_K imply a better matching between the reference and the current region.

In order to obtain a metric for the symmetrical form of equation (13) we use the values of $D(p(u)||p(v))$ and $D(p(v)||p(u))$, normalized with respect to their maximum values. Therefore equation (18) will be in this case replaced with:

$$E_K = c_1(1 - \frac{\frac{D(p(u)||p(v))}{D_{max}(p(u)||p(v))} \frac{D(p(v)||p(u))}{D_{max}(p(v)||p(u))}}{(\frac{D(p(u)||p(v))}{D_{max}(p(u)||p(v))} + \frac{D(p(v)||p(u))}{D_{max}(p(v)||p(u))})} - \lambda_1 \frac{N_T - M_2}{N_T} + c_2) \qquad (19)$$

4. NORMALIZED CORRELATION BASED METRIC

Normalized correlation is defined as [12]:

$$Cor = \frac{\sum_{i=1}^{W} \sum_{j=1}^{N_k} J_r(\mathbf{x}^r_{i,j}) J_c(\mathbf{x}^c_{i,j})}{\sqrt{\sum_{i=1}^{W} \sum_{j=1}^{N_k} J_r^2(\mathbf{x}^r_{i,j}) \sum_{i=1}^{W} \sum_{j=1}^{N_k} J_c^2(\mathbf{x}^c_{i,j})}}, \quad (20)$$

where

$$N_k = min(M_1, M_2), \qquad (21)$$

expresses the similarity between J_r and J_c and can be used to construct a metric similar to those already presented in Eq. (4) ,(18). The metric constructed is of the form:

$$Corr = c_1(Cor - \lambda_1 \frac{N_T - M_2}{N_T} + c_2) \tag{22}$$

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

$$0 \leq Corr \leq 1 \tag{23}$$

if the values of the constants c_1,c_2,λ_1 are appropriately chosen. In the context of present work $c_1 = 0.5$, $c_2 = 1$, $\lambda_1 = 1$.

5. EXPERIMENTAL RESULTS

The above metrics were tested using the object tracking algorithm presented in [13]. Image sequences containing total occlusion and severe partial occlusion were used. Curves showing the variations of metrics E_m and E_k and $Corr$ during the tracking process were acquired for different occlusion situations.

Tracking results on the football image sequence, an artificial image sequence and a head and shoulders image sequence containing partial occlusion are presented in Figures 1,2 and 3 respectively. The variations of the metric E_m for the three sequences are presented in Figure 4, while the metric E_k variations are presented in Figure 5. The symmetric version of equation (13) was used. Finally, the variations of the $Corr$ metric are presented in Figure 6. As it can be seen, metric E_k performs similarly to E_m. Further tests on the assymetric version have shown that no significant change in E_k behavior was caused by its asymmetry. The normalized correlation based metric $Corr$ does not behave as well as the information theory based metrics, especially in partial occlusion situations. More specifically, in partial occlusion, $Corr$ exhibits an abnormal behavior as can be seen in Figure 6. It should be also noted that, in the cases of the football and the artificial image sequences, feature point regeneration occurs after object reappearance. This is not the case in the head and shoulders image sequence. This difference is shown in the metric curves. The metrics E_m and E_k increase after the object reappearance in the football and the artificial image sequences, but their values remain low after object reappearance in the head and shoulders sequence.

6. CONCLUSIONS

In this paper, the performance of three different tracker reliability metrics was assessed. Two of the metrics are based on information theory, while the third is based on normalized correlation.

Experimental results have shown that 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. The Kullback-Leibler distance seems to provide more smooth curves. Nevertheless, the mutual information has the advantage of being unique and symmetrical.

7. REFERENCES

[1] 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*, 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] 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*, vol. 13, no. 4, 2003.

[4] K. Nickels and S. Hutchinson, "Estimating uncertainty in ssd-based feature tracking," *Image and Vision Computing*, vol. 20, pp. 47–58, 2002.

[5] P. Anandan, *Measuring Visual Motion From Image Sequences*, Phd dissertation University of Massachusetts at Amherst, 1987.

[6] 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.

[7] E. Loutas, C. Nikou, and I. Pitas, "Information theory-based analysis of partial and total occlusion in object tracking," in *Proc. of 2002 Int. Conf. on Image Processing*, 2002, vol. II, pp. 309–312.

[8] 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.

[9] C. Liu and H-Y Shum, "Kullback-leibler boosting," in *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR '03).*, 2003.

[10] R. Veldhuis, "The centroid of symmetrical kullback-leibler distance," *IEEE Signal Processing Letters*, vol. 9, no. 3, 2002.

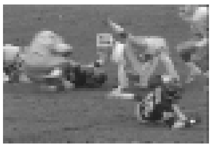
[11] M. N. Do, "Fast approximation of kullback-leibler distance for dependence trees and hidden markov models," *IEEE Signal Processing Letters*, vol. 10, no. 4, 2003.

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

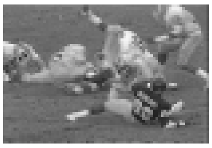
[13] E. Loutas, K. Diamantaras, and I. Pitas, "Occlusion resistant object tracking," in *Proc. of 2001 Int. Conf. on Image Processing*, 2001, vol. II, pp. 65–68.

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

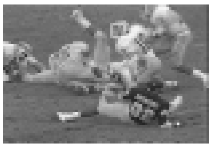
[15] M. Skouson, Q. Guo, and Z. Liang, "A bound on mutual information for image registration," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 843–846, 2001.



(a)

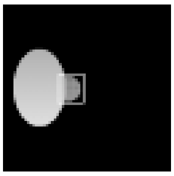


(b)

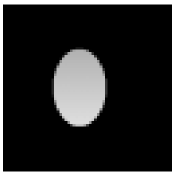


(c)

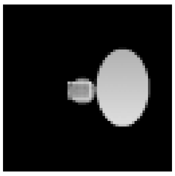
Fig. 1. Football image sequence. Tracking of the head of the football player: (a) before total occlusion, (b) total occlusion, (c) region reappearance.



(a)



(b)



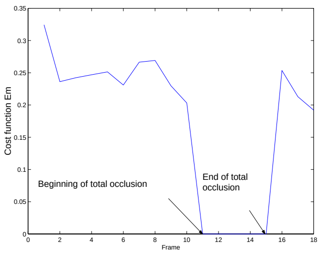
(c)

Fig. 2. Artificial image sequence: (a) before total occlusion, (b) during total occlusion, (c) region reappearance.

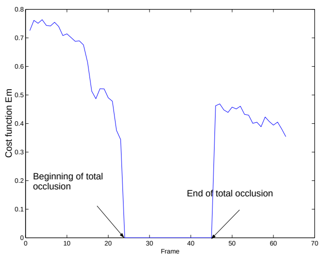
(a)

(b)

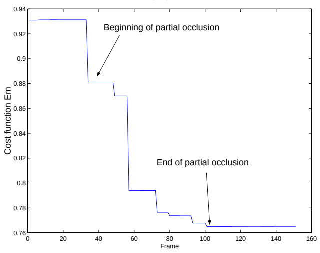
Fig. 3. Head and shoulders image sequence: (a) beginning of partial occlusion (frame 34), (b) disocclusion (frame 101).



(a)

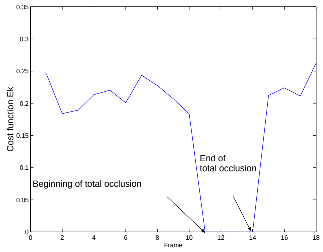


(b)

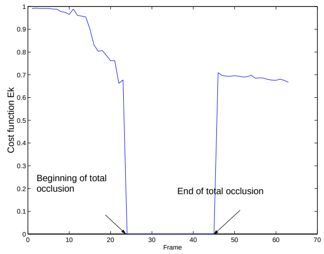


(c)

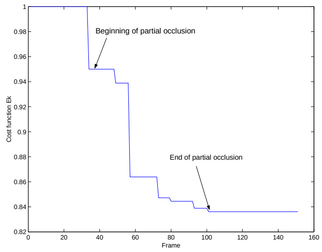
Fig. 4. (a) Values of the cost function E_m for part of the football image sequence (Fig. 1). Beginning of partial occlusion: frame 10. Beginning of total occlusion: frame 11. End of total occlusion: frame 15. (b) Cost function E_m versus frame number of the artificial image sequence of Fig. 2. Beginning of partial occlusion: frame 16. Beginning of total occlusion: frame 24. End of total occlusion: frame 45. (c) Values of the cost function E_m for the head and shoulders image sequence (Fig. 3). Beginning of partial occlusion: frame 34. End of partial occlusion: frame 101.



(a)

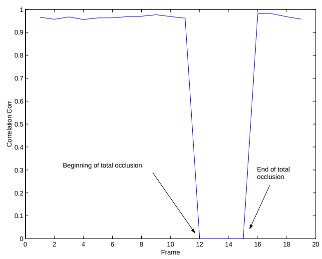


(b)

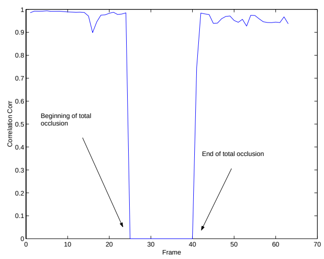


(c)

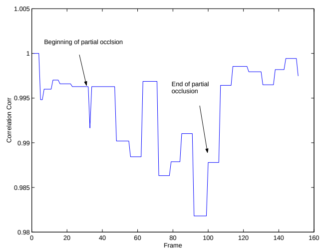
Fig. 5. (a) Values of the cost function E_K versus frame number for part of the football image sequence (Fig. 1). Beginning of partial occlusion: frame 10. Beginning of total occlusion: frame 11. End of total occlusion: frame 14.(b) Cost function E_K for the artificial image sequence (Fig. 2) versus frame number. Beginning of partial occlusion: frame 16. Beginning of total occlusion: frame 24. End of total occlusion: frame 45. (c) Values of the cost function E_K versus frame number for the head and shoulders image sequence (Fig. 3). Beginning of partial occlusion: frame 34. End of partial occlusion: frame 101.



(a)



(b)



(c)

Fig. 6. (a) Normalized Correlation for the football image sequence (Fig. 1) versus frame number. Beginning of partial occlusion: frame 10. Beginning of total occlusion: frame 12. End of total occlusion: frame 15.(b) Normalized Correlation for the artificial image sequence (Fig. 2) versus frame number. Beginning of partial occlusion: frame 16. Beginning of total occlusion: frame 25. End of total occlusion: frame 45. (c) Normalized Correlation for the head and shoulders image sequence (Fig. 3) versus frame number. Beginning of partial occlusion: frame 34. End of partial occlusion: frame 101.