14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A MUTUAL INFORMATION APPROACH TO CONTOUR BASED OBJECT TRACKING

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

An object tracking scheme based on mutual information is proposed in this paper. First, coarse tracking is performed by using mutual information maximization. Subsequently the tracking output is refined by the use of a deformable contour scheme based on image gradient and mutual information which allows the tracking system to capture the variation of the tracked object contour. The scheme was tested on hand sequences created for testing gesture recognition algorithms under difficult illumination conditions and found to perform better than a scheme based on the Kullback-Leibler distance and a scheme based on gradient information.

1. INTRODUCTION

Active contours have been extensively used in computer vision and in object tracking in particular. Various modifications of the initial active contour model were proposed. Some of them introduce different cues to the active contour model, others use the inside area defined by the active contour, and not the active contour points themselves.

The contour model proposed in [1] relies on color gradient in order to avoid the accuracy problems of contour deformation related to the intensity gradient. In [2] a modified contour algorithm based on color (hue) and image intensity is presented. The Lucas - Kanade algorithm is used for interframe tracking and an adjustment scheme is introduced in order to enable accurate tracking after many frames.

The active contour model proposed in [3] relies on color and motion information. Both interframe and intraframe energy terms are used. In [4] a contour tracking scheme based on edges is presented. According to this scheme matching is performed by using a set of curves. The object shape is considered to be known a priori.

In [5] video object segmentation and tracking is performed using ”VSnakes”. A differential active contour energy is defined which reflects the difference between successive contour configurations, rather than the energy of the contour. An active contour tracking system that uses motion information in order to remove background clutter is proposed in [6]. First, an approximation of the the tracked object position is found. The final active contour position is obtained by minimizing an energy function and takes into account the predicted object position and the edge map near the predicted object contour.

Integration of edge tracking and point tracking cues in a deformable model is presented in [7] Another approach to tracking active contours that uses Kullback-Leibler distance is proposed in [8]. The entire area defined by the curve is used to perform energy function minimization.

The use of mutual information as a similarity measure in contour-based object tracking is examined in this paper. Mutual information has been previously used in image registration [9] and as a cue selection criterion in multiple cues tracking systems [10]. Moreover, spatiotemporal mutual information has been used in [11] in order to determine the focus of attention in videoconferencing. In [12] mutual information has been used in face tracking as one of the cues used in a probabilistic tracking scheme. However, mutual information has not been extensively used in contour based object tracking.

The tracking algorithm proposed in the context of this work is based on mutual information maximization. First, an object localization step is performed, followed by a contour tracking process that refines the tracking output. The contour energy is based on mutual information and image gradient information. The use of mutual information inhibits contour attraction to unwanted areas. A similar system based on the Kullback-Leibler distance was implemented for comparison purposes. A refinement scheme based on gradient information only, was also implemented for the same reason. Experimental results on real image sequences show the enhanced performance of the proposed scheme.

2. TRACKING SYSTEM DESCRIPTION

In the context of the present work the initialization is manual and can be done interactively. More precisely, initialization is performed by defining the object outline contour through a set of points. The tracking region translation and rotation is found by using mutual information maximization. The region contour is not allowed to deform during this first tracking stage. This results in a coarse estimate of the tracked region in the next frame. In the second, refinement step, the region contour is allowed to deform. This stage combines image gradient information and mutual information and gives the final region estimate as a set of points that define its boundary.

The proposed tracking system is summarized as follows:

- Initialization through feature point set selection.

This work has been partially conducted in conjunction with the SIMILAR European Network of Excellence on Multimodal Interfaces of the IST Programme of the European Union (www.similar.cc)

- Global region translation and rotation calculation using mutual information maximization (coarse tracking).
- Contour refinement based on gradient and mutual information.

3. COARSE MUTUAL INFORMATION OBJECT TRACKING

In order to perform coarse object tracking, a scheme based on mutual information is used. The tracking region motion parameter vector

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ \vartheta \end{bmatrix}, \tag{1}$$

where x and y are the translation parameters on the horizontal and vertical direction respectively and ϑ is an angle describing the tracking region rotation, is estimated based on the maximization of a mutual information based likelihood.

Let U^r, U^c be two random variables expressing pixel grayscale values of the tracked region in the reference J_r and current J_c frame and $u_i^r, u_j^c,\ 1 \le i, j \le N_{max}$ their possible outcomes where N_{max} is the maximum number of the available grayscale levels. Let also $p(u^r), p(u^c)$ and $p(u^r, u^c)$ be their marginal and joint probability density functions. The mutual information of the tracked region in the reference and current frame is defined as [13]:

$$I_{coarse}(U^r, U^c) = \sum_{i=1}^{N_{max}} \sum_{j=1}^{N_{max}} p(u_i^r, u_j^c) \log_2 \frac{p(u_i^r, u_j^c)}{p(u_i^r) p(u_j^c)}, \tag{2}$$

The probability density functions $p(u^r)$,$p(u^c)$ and $p(u^r, u^c)$ are determined by obtaining the histograms of the grayscale values of the reference and target regions.

The maximum mutual information for a particular prior $p(u^r)$ is [14]:

$$I_{maxcoarse}(U^r, U^c) = -\sum_{k=1}^{N_{max}} p(u_k^r) \log_2 p(u_k^r) \tag{3}$$

and reaches its maximum value when

$$p(u_k^r) = \frac{1}{N_{max}}, \qquad 0 \le k < N_{max}. \tag{4}$$

We define the probability of resemblance between the reference and the target region based on the mutual information as:

$$p_{MIcoarse} = \frac{I_{coarse}(U^r, U^c)}{I_{maxcoarse}(U^r, U^c)}. \tag{5}$$

Since $I(U^r, U^c)_{coarse} \ge 0$ [15],

$$0 \le p_{MIcoarse} \le 1. \tag{6}$$

A large value of $p_{MIcoarse}$ indicates a strong match between the reference and the current regions, while a small value indicates a weaker match. In order to find the tracking region motion parameter vector, the position in the current frame that maximizes $p_{MIcoarse}$ between the current and the reference frame is found by exhaustively checking candidate positions within a certain search range. In our case, the translation search range was set to 15 pixels and the angle search range was set to 4.6 degrees.

4. CONTOUR REFINEMENT SCHEME

In order to account for tracking region deformations, a contour refinement scheme is applied after the coarse tracking step. The contour deformation is governed by internal energy terms and external terms. Two external forces are applied. The first is based on image gradient, while the second term is based on mutual information.

The classic deformable contour energy equation is expressed as follows:

$$\varepsilon = \int (\alpha E_{cont} + \beta E_{cur} + \gamma E_{image}) ds \tag{7}$$

The terms E_{cont}, E_{cur} represent the contour internal energy. The first term is the continuity term, which prevents the formulation of clusters of contour points. The second term is called the smoothness term. The aim of this term is to avoid contour oscillations. The third energy term E_{image} corresponds to an external force attracting the contour points. Usually the contour points are attracted to image edges.

In the discrete case, the contour is represented by a chain of N image points $\mathbf{a}_1 \ldots \mathbf{a}_N$. The continuity term is expressed as:

$$E_{cont} = (\overline{d} - ||\mathbf{a}_i - \mathbf{a}_{i-1}||)^2 \tag{8}$$

where $\overline{d}$ is the average distance between pairs $(\mathbf{a}_i, \mathbf{a}_{i-1})$ of adjacent points.

The second term is expressed as:

$$E_{cur} = ||\mathbf{a}_{i-1} - 2\mathbf{a}_i + \mathbf{a}_{i+1}||^2. \tag{9}$$

The third term is expressed as:

$$E_{image} = -||\nabla I||. \tag{10}$$

The gradient values of all contour points $\mathbf{a}_1 \ldots \mathbf{a}_N$ are used for the gradient term calculation. In order to obtain the gradient values for the candidate contour points, the image gradient map is calculated.

In the context of this work the contour equation is reformulated with the addition of a new energy term based on mutual information:

$$\varepsilon = \int (\alpha E_{cont} + \beta E_{cur} + \gamma E_{image} + \zeta MI) ds \tag{11}$$

where

$$MI = -p_{MIcontour}. \tag{12}$$

This new energy term is evaluated as follows: A 15×15 window $\mathbf{W}_i^r$ is considered around each of the N image points $\mathbf{a}_i^r$ that define the contour in the reference frame. In the same manner, a 15×15 window $\mathbf{W}_i^c$ is considered around each of the N image points $\mathbf{a}_i^c$ that define the contour in the current frame. Let $\mathbf{W}^r$ denote the area in the reference frame formed by the union of all windows $\mathbf{W}_i^r$, and $\mathbf{W}^c$ the area in the current frame formed by the union of all $\mathbf{W}_i^c$. The mutual information $I_{contour}(U^r, U^c)$ between $\mathbf{W}^r$, $\mathbf{W}^c$ is evaluated as in (2) but in this case the probability density functions

$p(u^r), p(u^c)$ and $p(u^r, u^c)$ are evaluated over $\mathbf{W}^r$ and $\mathbf{W}^c$. The maximum value $I_{maxcontour}(U^r, U^c)$ of $I_{contour}(U^r, U^c)$ is evaluated using an expression similar to (3). Thus, $p_{MIcontour}$ is finally evaluated as

$$p_{MIcontour} = \frac{I_{contour}(U^r, U^c)}{I_{maxcontour}(U^r, U^c)}. \tag{13}$$

The formulation of contour equation (11) differs from that of equation (7). The classic contour equation is calculated over the contour that outlines the tracked image region in the current frame only. The addition of the mutual information term into the equation introduces information about the contour of the image region in the reference frame. The negative sign in equation (12) is necessary in order to attract the contour to the direction where the mutual information between the windows around the contour points in the current and the reference frame is maximized. This is necessary in order to ensure maximum similarity between the contours. Therefore the proposed contour equation formulation (eq. (11)) causes contour attraction to high image gradient and high mutual information.

The choice of the coefficients γ and ζ control the behavior of the tracking scheme. A very large γ coefficient can cause unwanted contour attraction to edges, while a large ζ coefficient can cause the contour to be not adequately responsive to changes in the shape of the tracked region.

5. EXPERIMENTAL RESULTS

The proposed tracking scheme was tested on real image sequences used for gesture recognition. The Aalborg video sequence [16] database was used for this purpose. It consists of 16 video sequences recorded in PAL resolution (768×576). The gesture vocabulary consists of 13 gestures. 9 gestures are static and 4 are dynamic. The captured scene includes a messy table environment with normal stuff for paper work. The light setup is arranged so that the table is split up in two parts with the same intensity but different color. In Figure 1 hand tracking results obtained using the proposed method are presented. As can be seen, the human hand is tracked correctly. A simple stick model [17] describing the palm has been used in conjunction with the basic algorithm in this case.

In order to test the effect of the introduction of the mutual information term in the contour refinement step, a similar tracking scheme based on the Kullback-Leibler distance was implemented. Kullback-Leibler distance is defined as:

$$D(U\|V) = \sum_{k=1}^{N_{max}} p(u_k) \log_2 \frac{p(u_k)}{p(v_k)}, \tag{14}$$

The Kullback-Leibler distance is asymmetric, that is:

$$D(U\|V) \neq D(V\|U) \tag{15}$$

whereas mutual information is symmetric. In the alternative scheme, the coarse tracking step is based on mutual information maximization, i.e., it is identical to the one described in Section 3, whereas the contour refinement scheme is based on the following equation:

$$\varepsilon = \int (\alpha E_{cont} + \beta E_{cur} + \gamma E_{image} + \zeta D(U\|V)) ds. \tag{16}$$

Moreover, a tracking scheme where coarse tracking is performed as in Section 3 whereas the refinement is based only on gradient information, i.e., it utilizes (7), was also implemented. Some results are presented in Figures 2 and 3. As can be seen, the Kullback-Leibler distance and the gradient based contour refinement schemes did not perform as well as the mutual information based refinement scheme. The superiority of the proposed method was also verified on other sequences depicting a moving hand and moving fingers that were captured by the authors in a laboratory environment (Figures 4,5,6).

6. CONCLUSIONS

An object tracking scheme using mutual information has been presented in this paper. Coarse tracking is performed by using a mutual information based similarity measure. The tracking process is enhanced by using a contour refinement that also involves mutual information cues. Contour refinement schemes relying on Kullback-Leibler distance and gradient information were also tested as an alternative but did not perform as well. The proposed tracking scheme was successfully tested on real image sequences used for gesture recognition purposes.

REFERENCES

[1] T. Gevers, "Robust segmentation and tracking of colored objects in video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 776–781, 2004.

[2] N. Eveno, A. Caplier, and P-Y Coulon, "Accurate and quasi-automatic lip tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 5, pp. 706–715, 2004.

[3] Y. Fu, A. T. Erdem, and A. M. Tekalp, "Tracking visisble boundary of objects using occlusion adaptive motion snake," *IEEE Transactions on Image Processing*, vol. 9, no. 12, pp. 2051–2060, 2000.

[4] D. Freedman, "Effective tracking through tree-search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 604–615, 2003.

[5] S. Sun, D. R. Haynor, and Y. Kim, "Semiautomatic video object segmentation using vsnakes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 1, pp. 75–82, 2003.

[6] H. T. Nguyen, M. Worring, R. van den Boomgaard, and A. W. M. Smeulders, "Tracking nonparameterized object conoturs in video," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 1081–1091, 2002.

[7] S. K. Goldenstein, C. Vogler, and D. Metaxas, "Statistical cue integration in dag deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 801–813, 2003.

[8] D. Freedman and T. Zhang, "Active contours for tracking distributions," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 518–526, 2004.

[9] 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.

[10] H. Kruppa and B. Schiele, "Using mutual information

to combine object models," in *8th International Symposium on Intelligent Robotic Systems 2000, Reading, UK.*, 2000.

[11] M. Onishi, T. Kagebayashi, and K. Fukunaga, "Production of video images by computer controlled cameras and its application to tv conference systems," in *Proc. of 2001 Int. Conf. on Computer Vision and Pattern Recognition*, 2001, vol. II, pp. 131–137.

[12] E. loutas, I. Pitas, and C. Nikou, "Probabilistic multiple face detection and tracking using entropy measures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 128–135, 2004.

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

[14] 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.

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

[16] M. Nielsen, M. Stoerring, T. B. Moeslund, and E. Granum, "A procedure for developing intuitive and ergonomic gesture interfaces for hci," in *In Gesture-Based Communication in Human-Computer Interaction: 5th International Gesture Workshop GW 2003, Genova, Italy. Selected Revised Papers, LNCS 2915 / 2004, pages 409-420*, April 2003.

[17] J. Rehg and T. Kanade, "Visual tracking of high dof articulated structures: an application to human hand tracking," in *Proceedings of the European Conference on Computer Vision*, 1994, pp. 35–46.

Figure 1: Hand tracking sequence – Mutual information based tracking refinement scheme

Figure 2: Hand tracking sequence – Kullback-Leibler distance based tracking refinement scheme.

Figure 3: Hand tracking sequence – gradient based tracking refinement scheme.

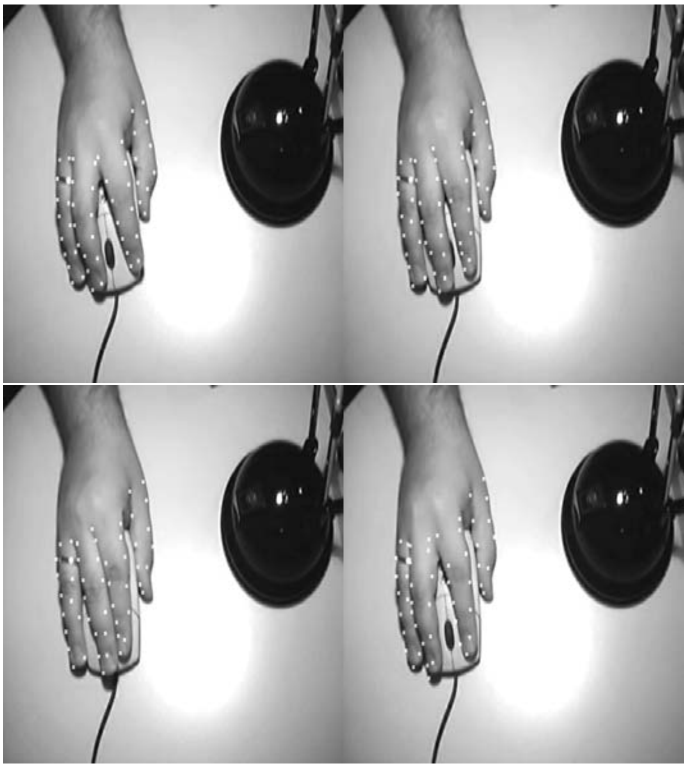

Figure 4: Finger tracking sequence obtained in the lab – Mutual information based tracking refinement scheme

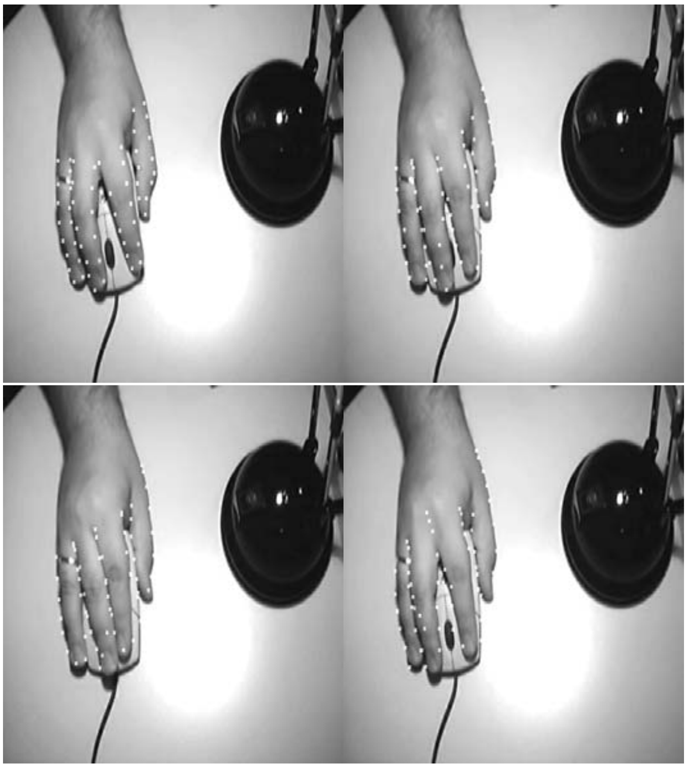

Figure 5: Finger tracking sequence obtained in the lab – Kullback-Leibler based tracking refinement scheme

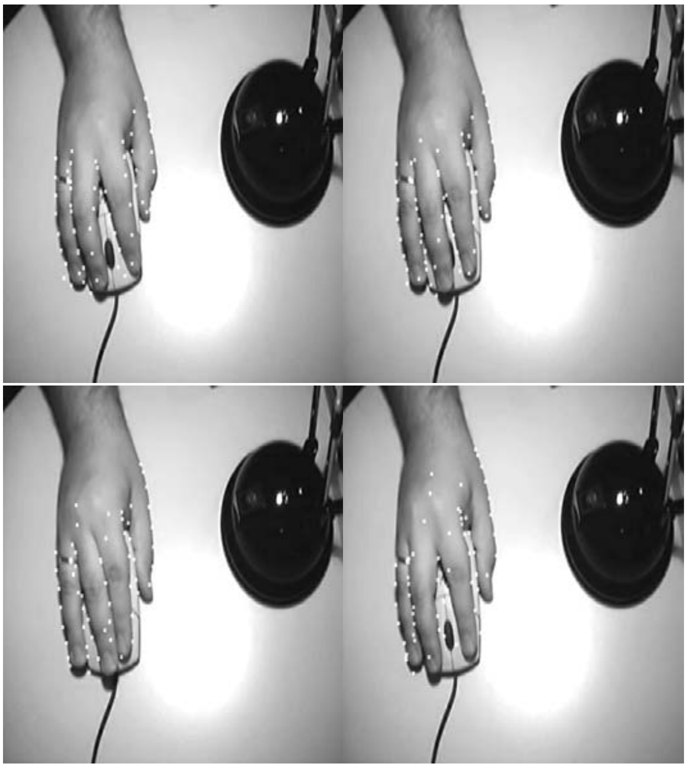

Figure 6: Finger tracking sequence obtained in the lab – Gradient based tracking refinement scheme