A MUTUAL INFORMATION APPROACH TO ARTICULATED OBJECT TRACKING

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

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

ABSTRACT

A mutual information based articulated object tracking scheme is proposed in this paper. Articulation constraints are introduced using a kinematic model. Further constraints are introduced based on the human joint anatomy and flexibility. The tracking scheme is enhanced by using the tracked object texture map image. The tracking history is incorporated in the tracking scheme by using a temporal model or a Kalman filtering scheme. The Kalman filtering scheme greatly enhances the tracking scheme provided the suitable initial conditions are set. The resulting system was tested on arm and finger tracking cases using real image sequences

1. INTRODUCTION

Articulated object tracking, in particular human body part tracking, is important in many scientific fields including medicine, athletics and robotics. The flexibility of the human body and the difficulty to predict its movement when compared to the robot makes human body part tracking a difficult task to perform. Part of the existing techniques make extensive use of human body models [1, 2, 3, 4]. The position and motion of a human arm is estimated by using a Kalman filtering scheme in [5]. Articulated object tracking of highly complicated articulated structures is accomplished in [6] by expressing the transform matrix to an exponential. Statistical training was also used in articulated object tracking systems [7]. Particle filtering was also extensively used in articulated object tracking, especially hand tracking [8, 9, 10].

The constraints imposed by articulation are introduced by using kinematic modelling. Moreover, kinematic modelling is used to predict self occlusions [1], while the problem of singularities in articulated object tracking was examined in [4]. More recently a novel kinematic model of human body motion was introduced in [11]. The constraints imposed are distinguished in "hard" and "soft". The former have to do with velocity and acceleration limits while the latter are probabilistic and associated to previous instances of human motion. In order to perform articulated object tracking different matching criteria based on color [12], luminance [1, 13] or edge information [7] were used.

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).

The use of mutual information as a similarity measure in articulated object tracking is examined in this paper. Mutual information has been previously used in image registration [14]. It has also been used as a cue selection criterion in multiple cue tracking system [15]. Moreover a spatiotemporal mutual information has been used in [16] in order to produce vivid video images sequences in videoconferencing. Nevertheless it has not been used in articulated object tracking.

Moreover, the tracked object texture map image [17] is used instead of tracked object image in order to calculate the mutual information based similarity measure. The use of the tracked object texture map image combined with a suitable confidence map provides a stabilized view of the tracked object [18] that can be used in existing 2-D techniques. The texture map technique was previously used in head tracking, but it can be easily extended to other objects by using a cylinder as an intermediate surface [17]. Experimental results prove that this technique can greatly improve the tracker performance.

The tracking procedure is assisted by a kinematic model closely related to that presented in [1] and a temporal model [7]. A Kalman filtering scheme is also applied. Constraints based on the kinematic behavior of each joint are introduced and are categorized into search range constraints and initial state constraints. These constraint can be seen as "hard" constraints in the context of [11]. Self occlusion is predicted by using a mutual information based scheme. The use of such a scheme not only allows the prediction of self occlusion but in addition it distinguishes the occluding and the occluded object parts. Experimental results on real image sequences show the enhanced performance of the proposed scheme.

2. KINEMATIC MODEL

The kinematic model used is inspired by the model presented in [1]. Additional constraints are imposed in order to make articulated object tracking feasible. These constraints are based on the moving capabilities of the human body. It is well known that some joints allow a large variety of movements in different directions, while others do not allow movement in certain directions. Moreover, the speed attained by different body parts connected to the same joint is different. An example showing the need of imposing constraints on the motion of certain body parts is the human finger, which consists of three parts. The middle part moves to greater extend and more rapidly than the other two parts. Moreover, the human finger is not allowed to move to all directions. Therefore,

the human finger cannot be considered as equivalent to a three part robotic mechanism and subsequently more constraints need to be imposed. These constraints can be divided in two categories:

- Tracker search range constraints.
- Initial state constraints.

Initial state constraints involve the choice of suitable initial kinematic conditions based on the body part being tracked. The choice of initial velocities and accelerations for each body part is crucial for the system performance. The slowly moving parts should be assigned smaller initial velocities and accelerations than the faster moving parts.

3. TEXTURE MAP CALCULATION

The object to be tracked is mapped onto a 2D texture map, by using a cylinder as an intermediate surface. The 2D texture map is constructed using the following formula:

$$(x, y) = (\frac{r}{c}(\theta - \theta_0), \frac{1}{d}(h - h_0)) \tag{1}$$

Parameters θ and h are explained in figure 1, c and d are scale factors and θ_0, h_0 are used to shift the object to be tracked over the texture map. A confidence map is superimposed on the resulting texture map in order to alleviate the deformation of the tracked object texture map. The deformation is significant in the left and right edges of the the texture map, while it is minimal in the central area of the texture map. Therefore the confidence map should attenuate the influence of the texture map edges. The confidence map used is of the form [19]:

$$\omega = \begin{cases} [1 - \frac{2}{\pi}(\frac{\pi}{2} - \theta)], \theta \epsilon [0, \frac{\pi}{2}] \\ 1 - \frac{2}{\pi}[\frac{\pi}{2} - (\pi - \theta)], \theta \varepsilon (\frac{\pi}{2}, \pi] \end{cases} \tag{2}$$

As it can be seen, the confidence map reaches its maximum value when $\theta = \frac{\pi}{2}$ and its minimum values when $\theta = 0$ and $\theta = \pi$. The initial image and the texture map before and after the confidence map superposition is shown in figure 2.

4. MUTUAL INFORMATION ARTICULATED OBJECT TRACKING

In order to perform articulated object tracking, a scheme based on mutual information is used. The angle of rotation $\overline{\vartheta}$, for each of the articulated object parts, is estimated based on the maximization of a mutual information based likelihood.

Let N_{max} be the maximum number of grayscale levels and $U^i(\overline{\vartheta}^i_{t-1}), V^i(\overline{\vartheta}^i_t)$ be two random variables with $p^i(u, v), p^i(u)$, $p^i(v)$ their joint and marginal probability mass functions and $\overline{\vartheta}^i_t$ is the the rotation angle of the ith body part at time t.

The mutual information of two random variables U^i, V^i with a joint probability mass function $p(u^i, v^i)$ is defined as [20]:

$$I(U^i(\overline{\vartheta}^i_{t-1}), V^i(\overline{\vartheta}^i_t)) = \sum_{k=1}^{N_{max}} \sum_{l=1}^{N_{max}} p(u^i_k, v^i_l) \log_2 \frac{p(u^i_k, v^i_l)}{p(u^i_k)p(v^i_l)}, \tag{3}$$

The maximum mutual information for a particular prior $p^i(u)$ is [21]:

$$I_{max}(U(\overline{\vartheta}^i_{t-1}), V(\overline{\vartheta}^i_t)) = -\sum_{k=1}^{N_{max}} p(u^i_k) \log_2 p(u^i_k) \tag{4}$$

and reaches its maximum value when

$$p(u^i_k) = \frac{1}{\log_2 N_{max}}, \qquad 0 \leq k < N. \tag{5}$$

We define the prior probability based on the mutual information tracking cue as:

$$p_{MI}(\overline{\vartheta}^i_t | \overline{\vartheta}^i_{t-1}) = c_1 \frac{I(U(\overline{\vartheta}^i_{t-1}), V(\overline{\vartheta}^i_t))}{I_{max}(U(\overline{\vartheta}^i_{t-1}), V(\overline{\vartheta}^i_t))}, \tag{6}$$

where c_1 is a constant. Since $I(U, V) \geq 0$ [20],

$$0 \leq p_{MI}(\overline{\vartheta}^i_t | \overline{\vartheta}^i_{t-1}) \leq 1. \tag{7}$$

A large value of $p_{MI}(\overline{\vartheta}^i_t | \overline{\vartheta}^i_{t-1})$ indicates a strong match between the reference and the target regions, while a small value indicates a weaker match.

The estimation of the rotation angle $\overline{\vartheta}^i_t$ of the ith articulated object part enables the calculation of the rotation angle $\overline{\vartheta}^{i+1}_t$ of the $i + 1th$ articulated object part. During the articulated object part tracking, the two constraint categories described above are introduced. The mutual information tracking scheme is enhanced by the use of a temporal model. Alternatively, a Kalman filtering scheme can be used, as explained in the following subsections.

4.1. Temporal model

The use of a temporal model as a constraint factor in the tracking process implies little or no knowledge of the tracking process previous history [22]. The previous history on tracking of the ith part of the articulated object is modelled by:

$$p_i(\overline{\vartheta}_t | \overline{\vartheta}_{t-1}) \sim \exp^{(-c(\overline{\vartheta}_t - \overline{\vartheta}_{t-1})^2)} \tag{8}$$

A very small value of constant c will render the temporal model non informative and, thus, not useful for the tracking process.

4.2. Kalman filtering

Kalman filtering can also be used to enhance the tracking process. It is also a way of imposing constraints in the form of initial conditions. In the context of the present work, the constant acceleration model is used. Constraints in the tracking process are imposed by setting suitable velocities and accelerations. The state vector is comprised by the rotation angles of each joint, their velocities $\frac{d\overline{\vartheta}}{dt}$ and accelerations $\frac{d^2\overline{\vartheta}}{dt^2}$,

$$\mathbf{s} = \begin{bmatrix} \overline{\vartheta} \\ \frac{d\overline{\vartheta}}{dt} \\ \frac{d^2\overline{\vartheta}}{dt^2} \end{bmatrix}, \tag{9}$$

while the measurement vector $\mathbf{d}$ is comprised by the rotation angles of each joint:

$$\mathbf{d} = \begin{bmatrix} \overline{\vartheta} \end{bmatrix} \tag{10}$$

4.3. Occlusion prediction

Occlusion and in particular self occlusion is considered as a major problem for many articulated object tracking schemes. In [1] the problem of self occlusion is handled using a set of templates. The occluding and occluded parts are determined by using a set of rules based on the camera position. In the context of present work, self occlusion is handled using a mutual information based scheme.

The use of the matching probability as a reliability measure is insufficient as it does not include by itself the previous history of the body part. It is, thus, necessary to compare the candidate human part region with the initial region. The mutual information of the two regions divided by its maximum serves as a reliability measure, whose values below a threshold signify the presence of occlusion. That is, the reliability measure used is $p_{MI}(\overline{\vartheta}_{t=t_{curr}}|\overline{\vartheta}_{t=0})$. The articulated body part with the higher reliability value *occludes* the part with the lower reliability value. Thus, the occlusion sequence is determined without resorting to the knowledge of the camera position.

5. EXPERIMENTAL RESULTS

The proposed tracking scheme was tested on real image sequences. An one part arm tracking example is presented in Figure 3. The tracking scheme is able to follow the arm even in rotation angles of 90 degrees. A two part arm tracking example is presented in Figure 4. The successful tracking of the upper articulated part is necessary in order to complete the tracking of the lower articulated part. The tracking results were improved by using the tracked object texture map. More specifically, using a texture map allowed the system to successful track the articulated objects in frames where a system without texture mapping failed.

The proposed tracking scheme was also tested on finger tracking. A two finger tracking example is presented in Figure 5. The variations of the reliability metric for each of the two fingers are depicted in Figure 6. Self occlusion reveals itself as a notable difference in the reliability metrics of the two fingers. The occluding finger is the one with the higher reliability metric.

Experiments proved that the use of the temporal model enhances the tracker performance. However, the best results were obtained when the Kalman filtering scheme, combined with successful initial conditions, was applied.

6. CONCLUSIONS

An articulated object tracking scheme using mutual information is presented in this paper. Tracking is performed by using a mutual information based similarity measure. The measure is calculated on the tracked object image, or alternatively on the tracked object texture map, accompanied with a confidence map. The use of the object texture map is found to improve the tracker performance. Articulation constraints are introduced using a kinematic model. Moreover constraints on the tracker search range and initial conditions based on the anatomy and kinematic capabilities of each joint are introduced. The tracking process is enhanced by using a constant acceleration Kalman filter model. Alternatively, a temporal model can be used when suitable initial conditions cannot be chosen for the Kalman filtering scheme. The proposed tracking scheme was successfully tested on real image sequences.

7. REFERENCES

[1] J. Rehg and T. Kanade, "Model-based tracking of self occluding articulated objects," in *Proceedings of the European Conference on Computer Vision*, 1995, pp. 612–617.

[2] D. Gavrila, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[3] H. Sidenbladh, F. De la Torre, and M. Black, "A framework for modeling the appearance of 3d articulated figures," in *IEEE International Conference on Automatic Face and Gesture Recognition (FG), Grenoble, France.*, 2000, pp. 368–375.

[4] D. Morris and J. Rehg, "Singularity analysis for articulated object tracking," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 1998.

[5] L. Goncalves, E. Di Bernardo, E. Ursella, and P. Perona, "Monocular tracking of the human arm in 3d," in *Proceedings of the International Conference on Computer Vision*, 1995.

[6] 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, pp. 8–15.

[7] H. Sidenbladh and M. Black, "Learning image statistics for bayesian tracking," in *IEEE International Conference on Computer Vision (ICCV), Vancouver, Canada.*, 2001, vol. 2, pp. 709–716.

[8] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2000.

[9] A. Blake J. Rittscher, "Classification of human body motion," 1999, pp. 634–639.

[10] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," in *European Conference on Computer Vision*, 2000, vol. 2, pp. 3–19.

[11] S. Dockstader, M. Berg, and A. Tekalp, "Performance analysis of a kinematic human motion model," in *IEEE International Conference on Multimedia and Expo (ICME 2002), Laussane , Switcherland.*, 2002.

[12] E. Polat, M. Yeasin, and R. Sharma, "Detecting and tracking body parts of multiple people," in *Proc. of 2001 Int. Conf. on Image Processing (ICIP 2001)*, 2001, pp. 405–408.

[13] S. 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.

[14] 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.

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

[16] 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.

[17] A. Watt and F. Policarpo, *The Computer Image*, Addison-Wesley, 1998.

[18] M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *IEEE Transactions on Medical Imaging*, vol. 22, no. 4, pp. 322–336, 2000.

[19] Jing Xiao, Takeo Kanade, and Jeffrey Cohn, "Robust full-motion recovery of head by dynamic templates and re-registration techniques," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002.

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

[21] 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.

[22] J. Ruanaidh and W. Fitzgerald, *Numerical bayesian methods applied to signal processing*, Springer-Verlag, 1996.

[23] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

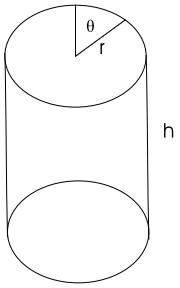

Figure 1: Texture map cylinder parameters

Figure 2: Initial image, texture map without confidence map, texture map with confidence map. The images were obtained from the MIT Vision and Modeling Group face database [23].

Figure 3: One part arm tracking

Figure 4: Two part arm tracking

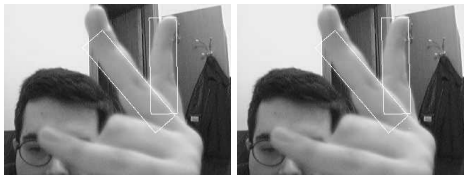

Figure 5: Two finger tracking

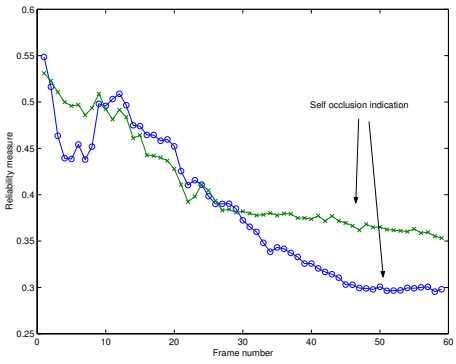

Figure 6: Reliability metric for the tracking procedure of each of the two fingers (Figure 5). The time interval where self occlusion occurs is marked. The finger corresponding to the higher-valued metric occludes the finger related to the lower-valued metric