ROBUST SHAPE BASED TWO HAND TRACKER

K. A. Barhate[1], *K. S. Patwardhan*[1], *S. Dutta Roy*[1,*], *S. Chaudhuri*[1], *S. Chaudhury*[2]

[1]Dept of Electrical Engineering
IIT Bombay, Mumbai 400076

[2]Dept of Electrical Engineering
IIT Delhi, New Delhi 110016

ABSTRACT

This paper presents a robust shape-based on-line tracker for simultaneously tracking the motion of both hands, that is robust to cases of background clutter, other moving objects, occlusions of one hand by the other, and a wide range of illumination variations. The tracker is based on an on-line predictive EigenTracking framework. This framework allows efficient tracking of articulate objects, which change in appearance across views. We show results of successful tracking across all possible cases of motion dynamics of both hands during occlusion, and a wide range of illumination conditions.

1. INTRODUCTION

A hand gesture-based interface is very intuitive and natural for man-machine communication. The use of both hands is an obvious choice for such systems. Such a system should therefore be able to handle mutual occlusion of the hands. In this paper, we propose an appearance-based, two hand tracker which can handle all possible cases of motion dynamics during mutual occlusion, across a wide range of illumination conditions.

Many different approaches have been explored in the literature. In [1] the authors use multiple cameras to track the 3-D position, posture and shape of the hand. They use *best view point selection* to handle mutual occlusions. However, having multiple synchronized (albeit uncalibrated) cameras is often not feasible. Mammen *et al.* [2] estimate the occluded observation elements in terms of non-occluded ones and their predicted values. However, authors do not consider all possible cases of mutual occlusions. They also can not identify or associate respective hands across an occlusion. Peterfreund [3] presents a Kalman filter-based active contour model (snake) to track non-rigid objects such as hands. The work uses an optical flow-based detection method to deal with occlusions and image clutter. The method rejects measurements that are inconsistent with previous estimates of image motion. This may not be true for all cases of mutual occlusions. Shamaie and Sutherland [4] approach the occlusion problem in a two hand tracker by modeling the spatial synchronization in bi-manual movements by the position and temporal synchronization using the velocity and acceleration of each hand.

Extensive research work has been carried out in the area of tracking people during occlusion that could be considered analogous to the problem of two hand tracking. However, such systems often make domain-specific assumptions about the features of the objects being tracked, which may not hold for the case of two hands. Sherrah and Gong [5] use a Bayesian network to track multiple interacting body parts like faces and hands, during occlusion. Most of the above systems would fail in case the moving objects change their appearance substantially.

In our approach, we use our Predictive EigenTracker [6] to track the hands. An EigenTracker [7] can track objects that simultaneously undergo image motion and changes in appearance. The paper [6] incorporates a prediction framework in the basic EigenTracker to increase its efficiency. The authors also use an efficient eigenspace update mechanism to learn and track the unknown views of the object, on the fly. We use two such trackers to track the hands. Our algorithm handles all possible cases of occlusion, similar to [4]. Appearance-based tracking also enables us to identify and associate both hands correctly across an occlusion. We incorporate a novel neural network-based colour constancy algorithm to make our tracker robust to variations in the illumination conditions.

2. TWO HAND TRACKER

Figure 1 gives an overview of our two-hand predictive EigenTracker. We use skin colour and motion cues [6] to perform *fully automatic initialization* of the tracker. The EigenTracker approximates the object motion by an affine transformation. This takes into account the effects such as translation, scaling and shear – commonly observed for articulate objects like hands. We use the six affine coefficients as the elements of the state vector, *i.e.*, $\mathbf{X} = [a_0\ a_1\ a_2\ a_3\ a_4\ a_5]^T$. Alternately, one can use coordinates of three object points as the state vector. The state vector transforms the view of

*Author for correspondance, sumantra@ee.iitb.ac.in

TWO HAND PREDICTIVE EIGENTRACKER

A. Delineate moving objects of interest *i.e.*, the two hands
B. REPEAT FOR ALL frames:
1. Obtain image MEASUREMENT optimizing affine parameters $\mathbf{a}$ & reconstruction coefficients $\mathbf{c}$
2. ESTIMATE new affine parameters for both hands using output of step 1
3. FOR EACH hand: IF reconstruction error $\in (T_1, T_2]$ THEN update eigenspace
4. IF reconstruction error for ANY hand very large THEN construct eigenspace afresh

C. Once occlusion begins:
1. Stop Eigenspace update for both hands
2. Determine which edges of the two hands are observable
3. Derive the unobservable edges from the observable ones
4. Update the translation params. of the affine vector

D. When occlusion is over:
1. If ANY recons. error v. large THEN swap the bounding windows
2. Construct Eigenspace afresh

Fig. 1. Predictive EigenTracking Algorithm for two hands: An Overview

the object onto the eigenspace, given as

$$\mathbf{f}(\mathbf{p}, \mathbf{X}) = \begin{bmatrix} a_0 \\ a_3 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \mathbf{p} \qquad (1)$$

where $\mathbf{p}$ is the position vector of the point. An affine transformation implies a parallelogram bounding box. A parallelogram gives tighter fit to the object being tracked – an important consideration for the on-line eigenspace update mechanism (since we would like to avoid as much of the background as possible). A commonly used model for describing the state dynamics is a second order AR process, given as $\mathbf{X}_t = \mathbf{A}_2\mathbf{X}_{t-2} + \mathbf{A}_1\mathbf{X}_{t-1} + \mathbf{W}_t$, where t denotes time. The values of six affine parameters obtained from the image constitute the measurement $\mathbf{Z}$ (Step B.2 in Figure 1). Measurements are modeled as $\mathbf{Z}_t = \mathbf{B}\mathbf{X}_t + \mathbf{F}_t$. Here, $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{B}$ are the coefficient matrices and $\mathbf{W}_t$, $\mathbf{F}_t$ are zero-mean, white, Gaussian random noise vectors. During gesticulation, hands often undergo considerable changes in appearance. It may not be always possible to learn the multitude of hand poses off-line even for a single person. As in our work on predictive EigenTracking [6], we use efficient eigenspace update mechanism of Chandrasekaran *et al.* [8]. Depending on the reconstruction error of the hands, the system updates the appearance model of the hand (Step B.3 and B.4 in Figure 1). In every frame the tracker checks for overlap of the hands. Occlusion prevents a tracker from making measurements for the two hands. Figure 2 shows a case of mutual occlusion. The next section describes our occlusion handling strategy.

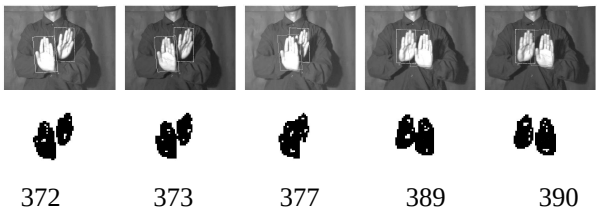

Fig. 2. Detection of occlusion start and end. Images in the upper row are the input images while the images in the lower row show the segmented hands based on skin colour

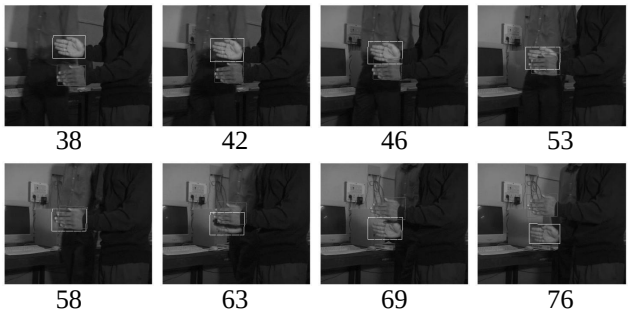

Fig. 3. Occlusion begins at frame 42 and ends at frame 69. Note that the bounding window is a parallelogram. For videos: `http://www.ee.iitb.ac.in/~sumantra/icip04b`

2.1. Occlusion Handling

Two trackers can not be used to track two overlapping objects as such. For the occlusion phase, the system can not update the six-element state vector. For the cases when the hands are not too tilted, we derive the measurements from the bounding extents of the detected skin blob. This increases accuracy of the system during occlusion. If a pair of opposite boundaries is visible, we update the corresponding difference variable (height or width). If only one boundary is visible, we use the corresponding difference variable to estimate position of the other. If both are unobservable, we use the second order AR process to estimate their positions. Once all boundaries are estimated in a frame, we update the translation parameters of the affine vector (we leave the other affine parameters unchanged). Step C in Figure 1

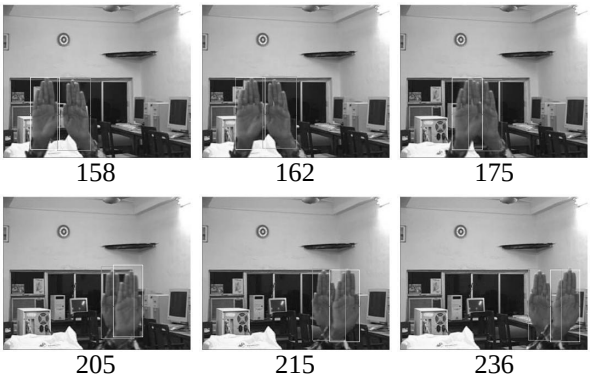

Fig. 4. Both hands moving in the same direction but with different velocities. Occlusion: from frame 162 to frame 236

summarizes our occlusion handling strategy. This scheme requires two measurements prior to occlusion, to be available. In a given video sequence, occlusions can therefore begin from the third frame onwards. This delay also allows the EigenTrackers to generate the appearance models of the hands. Figure 3 shows the result of successful tracking of the hands during occlusion, using our strategy. In Figure 4, we demonstrate the efficacy of our method for another interesting case. Here, both the hands move in the same direction, but with different velocities. Our tracker successfully tracks and associates the hands across the occlusion. If the hands change their direction of motion during occlusion, it is known as a *collision* [4]. *Our occlusion handling strategy works not just for simple occlusions, it works for all cases of collisions as well.* After a collision, for a non-appearance-based method, the hands may get wrongly identified because of the changes in their direction of motion. We use appearance models generated by the EigenTrackers to correctly identify and associate the respective hands after a collision. Figure 5 shows the result of successful tracking across a collision. Here, the hands approach each other from opposite directions, and change their direction of motion during occlusion to return back to their starting positions.

2.2. Automatic Tracker Initialization

Our system is flexible in that it does not require the hands to have a predefined shape at the beginning of the sequence. This makes tracker initialization even more difficult in the presence of multiple moving objects and background clutter. Our tracker performs *fully automatic initialization* under certain conditions. In general, one can use motion cues for object segmentation (Dominant motion analysis [9]), but depending on the application, other cues can be also used to advantage. We combine the motion cues with skin colour cues [10] to initialize our tracker. If multiple objects satisfy

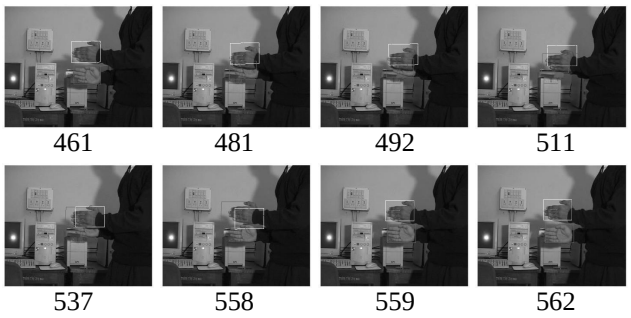

Fig. 5. Hands go back to their original position after occlusion. Occlusion:from frame 481 to frame 559

Fig. 6. Tracker initialization. (a) shows the original image, in (b) the detected skin coloured regions, and (c) shows the tracker initialized

the motion and colour criterion, we assume the two largest blobs to be the hands (Figure 6). Accurate tracker initialization helps the tracker to generate the correct eigenspace representations of both hands.

2.3. Use of Colour Constancy for Robust Tracking

If a gesture sequence is performed in a poorly illuminated environment (as in the upper part of Figure 7), the skin colour detection algorithm may fail to identify skin coloured regions. To make our tracker robust against the variations in illumination conditions, we apply a colour correction algorithm [11] to the input frames before the tracker processes them. This algorithm transforms the image taken under poor illumination conditions to canonical illumination conditions. The canonical conditions are those used in the training phase for skin colour detection [12]. We use a neural network implementing the back-propagation learning rule to perform the transformation. We train it using a skin colour palette under unknown illumination, and a similar palette under known illumination conditions. The lower sequence in Figure 7 shows the results of successful tracking after application of colour transformation algorithm.

3. CONCLUSIONS

This paper presents a two hand, shape-based, on-line tracker. The system is robust to cases of background clutter, other moving objects, and all possible mutual hand occlusions. The EigenTracking framework allows us to identify and cor-

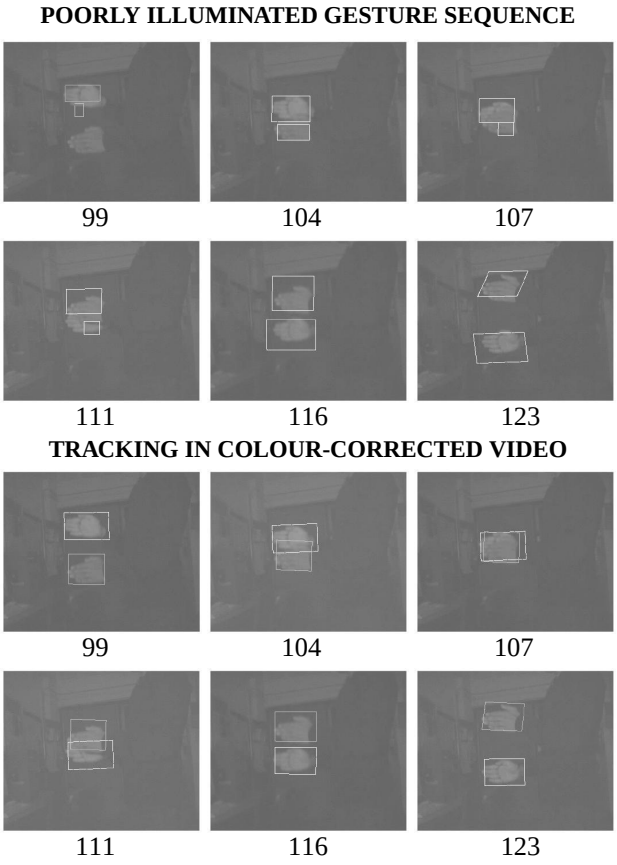

Fig. 7. Use of colour correction for enhanced tracking. Contrast has actually been enhanced in the first set of frames for clarity. Occlusion: from frame 104 to frame 116

rectly associate the hands across an occlusion. For certain cases of hand motion, we propose a framework to take measurements even during occlusion, thus enhancing tracking accuracy. The use of colour constancy makes our tracker robust to poor illumination conditions, as well. We show results of successful tracking for a large number of sequences.

Acknowledgment

Funding support of the Naval Research Board, Government of India, is gratefully acknowledged.

4. REFERENCES

[1] A. Utsumi and J. Ohya, "Multiple Hand Gesture Tracking using Mutliple Cameras," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1999, pp. 473 – 478.

[2] J. Mammen, S. Chaudhuri, and T. Agrawal, "Tracking of both hands by estimation of erroneous observations," in *Proc. British Machine Vision Conference (BMVC)*, 2001.

[3] N. Peterfreund, "Robust Tracking of Position and Velocity with Kalman Snakes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 564 – 569, June 1999.

[4] A. Shamaie and A. Sutherland, "A dynamic model for real-time tracking of hands in bimanual movements," in *5th International Gesture Workshop, Geneva*, April 2003.

[5] J. Sherrah and S. Gong, "Resolving Visual Uncertanity and Occlusion through Probabilistic Reasoning," in *Proc. British Machine Vision Conference (BMVC)*, 2000.

[6] N. Gupta, P. Mittal, S. Dutta Roy, S. Chaudhury, and S. Banerjee, "A Predictive Scheme for Appearance-based Hand Tracking," in *Proc. National Conference on Communications (NCC)*, 2002, pp. 513 – 522.

[7] M. J. Black and A. D. Jepson, "EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation," *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63 – 84, 1998.

[8] S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang, "An Eigenspace Update Algorithm for Image Analysis," *Graphical Models and Image Processing*, vol. 59, no. 5, pp. 321 – 332, September 1997.

[9] M. Irani, B. Rousso, and S. Peleg, "Computing Occluding and Transparent Motions," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5 – 16, January 1994.

[10] J. Mammen, S. Chaudhuri, and T. Agrawal, "Tracking of both hands by estimation of erroneous observations," in *Proc. British Machine Vision Conference (BMVC)*, 2001.

[11] A. Nayak and S. Chaudhuri, "Self-induced Color Correction for Skin Tracking Under Varying Illumination," in *Proc. International Conference on Image Proccesing*, September 2003.

[12] R. Kjeldsen and J. Kender, "Finding Skin in Color Images," in *Proc. Intl. Conf. on Automatic Face and Gesture Recognition*, 1996, pp. 312 – 317.