Estimation of 2D Motion Trajectories from Video Object Planes and Its Application in Hand Gesture Recognition

M.K. Bhuyan, D. Ghosh, and P.K. Bora

Department of Electronics and Communication Engineering,
Indian Institute of Technology, Guwahati 781039, India
{manas_kb, ghosh, prabin}@iitg.ernet.in

Abstract. Hand gesture recognition from visual images finds applications in areas like human computer interaction, machine vision, virtual reality and so on. Vision-based hand gesture recognition involves visual analysis of hand shape, position and/or movement. In this paper, we present a model-based method for tracking hand motion in a complex scene, thereby estimating the hand motion trajectory. In our proposed technique, we first segment the frames into video object planes (VOPs) with the hand as the video object. This is followed by hand tracking using Hausdorff tracker. In the next step, the centroids of all VOPs are calculated using moments as well as motion information. Finally, the hand trajectory is estimated by joining the VOP centroids. In our experiment, the proposed trajectory estimation algorithm gives about 99% accuracy in finding the actual trajectory.

1 Introduction

The use of human hand as a natural interface for human-computer interaction (HCI) serves as the motivation for research in hand gesture recognition. While static hand gestures are modelled in terms of hand configuration, as defined by the flex angles of the fingers and palm orientation, dynamic hand gestures require hand trajectories and orientation in addition to these. So, appropriate interpretation of dynamic gestures on the basis of hand movement in addition to shape and position is necessary for recognition.

Another form of dynamic hand gesture that is in common use is in which the 2D gesture trajectory alone builds up a particular message. Examples of such gestures are shown in Fig. 1. Therefore, an essential step in dynamic hand gesture recognition is estimation of 2D gesture trajectory.

All approaches to hand/finger tracking require to locate hand regions in video sequences. Skin color offers an effective and efficient way to segment out hand regions. However, many of these techniques are plagued by some special difficulties such as large variation in skin tone, unknown lighting conditions and dynamic scenes. A solution to this is the 3D model-based approach, in which the hand configuration is estimated by taking advantage of 3D hand models [1].

However, they lack the simplicity and computational efficiency. Therefore, an alternative approach that is computationally simpler is the appearance-based model. Another important approach for hand tracking is the prediction algorithm combined with Kalman tracking and application of probabilistic reasoning for final inference [2]. But, this approach requires fine initial model and also requires additional computations to take care of rotation and changes in shape of the model. Moreover, it is very much sensitive to background noise. In this paper, we now propose a novel method for gesture trajectory estimation that is computationally much more simpler and also robust to background noise.

Fig. 1. Hand gestures for representing a circle, two, square and a wavy hand

In our proposed technique, we use the concept of object-based video abstraction for segmenting the frames into video object planes (VOPs), as used in MPEG-4, where hand is considered as a video object (VO) [3]. The VOPs are binarized to form binary alpha plane – the hand-pixels are assigned “1” and the background pixels are assigned “0”. Thus, a binary model for the moving hand is derived and is used for tracking in subsequent frames. The Hausdorff tracker [4,5] is used for the purpose. The core of this algorithm is an object tracker that matches a two-dimensional (2D) binary model of the object against subsequent frames using the Hausdorff distance measure. The best match found indicates the translation the object has undergone, and the model is updated with every frame to accommodate for rotation and changes in shape. Next, we propose a centroid based method for estimating hand trajectory, where centroid of VOPs are calculated on the basis of moment as well as from respective motion vectors. The computed centroids are joined in the temporal sequence to get the final estimated trajectory.

2 Hausdorff Object Tracker for Hand Tracking

2.1 Hausdorff Distance

The Hausdorff distance measure can be used to measure the similarity between two shapes. It is defined as the maximum function between two sets of points:

$$H(O, I) = \max\{h(O, I), h(I, O)\} \quad (1)$$

where $h(O, I) = \max_{o \in O} \min_{i \in I} \|o - i\|$

and $h(I, O) = \max_{i \in I} \min_{o \in O} \|i - o\|$

$o_1, \dots, o_m$ and $i_1, \dots, i_n$ denote feature points in the sets O and I , respectively.

However, a computationally less expensive way to approximate the Hausdorff distance measure is the generalized Hausdorff distance in which the distances are sorted in ascending order and the k^{th} value is chosen as

$$h_k(O, I) = k_{o \in O}^{th} \min_{i \in I} \|o - i\| \quad (2)$$

Similarly, $h_l(I, O)$ is defined as the l^{th} value of the ordered distances. The maximum of these two gives the distance $H(O, I)$.

2.2 Hand Tracking and Model Update

The Hausdorff object tracker, which is based on the generalized Hausdorff distance, finds the position where the input hand model best matches the next edge image and returns the motion vector MV_i that represents the best translation from the $(i-1)^{th}$ frame to the i^{th} frame.

As a tracked object moves in a video sequence, it might rotate or change its shape. To allow for this, the model must be updated in every frame. The model is updated by using the motion vector MV_i . First, the model image in frame $(i-1)$, say M_{i-1} , is dilated and shifted by MV_i . Then, the portion of edge image that overlaps this shifted model is selected as the new model image M_i in the i^{th} frame.

2.3 Estimation of Motion Vector MV_i

In our hand tracking algorithm, we propose to use a simplified Hausdorff tracker with reduced computational load. Accordingly, we calculate generalized Hausdorff distance using distance transform algorithm [6]. Computation of distance transformation is accomplished using Chamfer 5–7–11 mask. Further reduction in computation is achieved by searching the shifted hand model in subsequent VOPs only at some possible pixel locations around the model image. The algorithm for determination of motion vector MV_i for i^{th} VOP may accordingly be described as given below.

Algorithm 1 : Estimation of Motion vector

begin

initialize translation vector t_x and t_y .

 Calculate distance transform of edge image.

 Calculate $h_k(O, I)$ for all translations (t_x, t_y) .

 Calculate $h_l(I, O)$ for all translations (t_x, t_y) .

$H(O, I) = \max\{h_k(O, I), h_l(I, O)\}$ for all translations (t_x, t_y) .

 Find $\min\{H_t(O, I)\}$ over all translations (t_x, t_y) .

 Find translation vector $t' = (t'_x, t'_y)$ corresponding to $\min\{H(O, I)\}$.

$MV_i = (t'_x, t'_y)$.

return MV_i

end

3 Proposed Hand Trajectory Estimation Algorithm

3.1 Step 1: Determination of Centroid of Hand Image in a VOP

After determining the binary alpha plane corresponding to each VOP, moments are used to find the center of the hand. The 0^{th} and the 1^{st} moments are defined as:

$$M_{00} = \sum_x \sum_y I(x, y), \quad M_{10} = \sum_x \sum_y xI(x, y), \quad M_{01} = \sum_x \sum_y yI(x, y) \quad (3)$$

Subsequently, the centroid is calculated as

$$x_c = \frac{M_{10}}{M_{00}} \quad \text{and} \quad y_c = \frac{M_{01}}{M_{00}} \quad (4)$$

In the above equations, $I(x, y)$ is the pixel value at the position (x, y) of the image. Since the background pixels are assigned “0”, the centroid of the hand in a frame is also the centroid of the total frame. Therefore, in moment calculation, we may either take the summation over all pixels in the frame or over only the hand pixels.

3.2 Step 2: Calculation of VOP Centroids

After determining the centroid of the first VOP in the gesture sequence, centroid of each subsequent VOP, say the i^{th} VOP, is determined by shifting the centroid of the previous VOP by the motion vector MV_i corresponding to the current frame. However, since slight changes in the shape of the hand model in successive frames cause shifting of actual centroids than that computed from the motion vectors in the previous step, centroids are also computed using equations (3) and (4). Finally, the centroid c_i of the i^{th} VOP is computed by taking the average of the two centroid values computed above.

3.3 Step 3: Trajectory Formation

In this step, the final trajectory is formed by joining all the calculated centroids in sequential manner. However, this trajectory may be noisy due to the following reasons.

- points too close.
- isolated points far away from correct trajectory due to change in hand shape.
- unclosed end points.
- hand trembling.
- unintentional movements.

Therefore, in order to take care of this, the final trajectory is smoothened out by considering the mean value of a specified point and its two neighboring points, *i.e.*,

$$(\hat{x}_t, \hat{y}_t) = ((x_{t-1} + x_t + x_{t+1})/3, (y_{t-1} + y_t + y_{t+1})/3) \quad (5)$$

So, dynamic hand gesture (DG) can be interpreted as a set of points in a spatio-temporal space as

$$DG = \{(\hat{x}_1, \hat{y}_1), (\hat{x}_2, \hat{y}_2), \dots, (\hat{x}_t, \hat{y}_t)\}. \quad (6)$$

4 Key Frame Based Trajectory Estimation

The computational burden in estimating hand trajectory can be significantly reduced by selecting the key video frames in the gesture video sequence. The key VOPs are selected on the basis of Hausdorff distance measure, thereby transforming an entire video clip to a small number of representative frames that are sufficient to represent a particular gesture sequence [7,8]. These key frames form the set of gesture frames that best represents the content of the sequence in an abstracted manner. After getting the key frames, hand trajectories are obtained by following the same procedure as that of the previous one, but considering the key VOPs only.

5 Experimental Results

We have tested altogether eight different hand trajectories in view of special applications like robot control and gesture based window menu activation in the Human-Computer interactive (HCI) platform. They are gestures showing one, two, three, five, wavy hand, seven, square and circle as shown in Fig. 2. Figure 3 shows the extracted trajectories for four of these gestures, *viz.*, circle, square, one and two. First column shows the extracted trajectories, second column shows the smoothed trajectory and the third column shows the trajectories obtained from the key frames. Our proposed trajectory estimator gives about 99% of accuracy in finding the actual trajectory. The accuracy criteria is fixed in terms of shape similarity between the extracted gesture trajectory and the corresponding template/prototype trajectory.

Fig. 2. Gestures showing 2D motion trajectory

Fig. 3. Estimated motion trajectories

6 Conclusions and Discussion

It is observed that the Hausdorff object tracker is capable of tracking hand motion in long gesture video sequences in an efficient manner. The VOP based method for segmentation requires no extra computation for rotation and scaling of the object, where the shape change is represented explicitly by a sequence of two dimensional models, one corresponding to each image frame. The major assumption is that the two dimensional shape of the hand changes slowly from one frame to the next. Moreover, trajectory estimated from the corresponding VOPs bears spatial information regarding hand positions in dynamic gestures, which is required in the gesture classification stage.

References

1. Pavlovic, V.I., Sharma, R., Huang, T.S.: Visual interpretation of hand gestures for human-computer interaction: A review, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **19** (7) (1997) 677-695
2. Wu, Y., Huang, T.S.: Self-supervised learning for visual tracking and recognition of human hand, *Proc. American Association for Artificial Intelligence (AAAI)*, (2000) 243-248
3. Bhuyan, M.K., Ghosh, D., Bora, P.K.: Automatic video object plane generation for recognition of hand gestures, *Proc. International Conf. Systemics, Cybernetics and Informatics (ICSCI)*, (2005) 147-152
4. Huttenlocher, D.P., Noh, J.J., Rucklidge, W.J.: Tracking non-rigid objects in complex scene, *4th International Conf. Computer Vision*, (1993) 93-101
5. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the Hausdorff distance, *IEEE Trans. Pattern Analysis and Machine Intelligence*, **15** (9)(1993) 850-863
6. Borgefors, G.: Distance transformations in digital images, *Computer Vision, Graphics and Image Processing*, **34** (1986) 344-371
7. Bhuyan, M.K., Ghosh, D., Bora, P.K.: Finite state representation of hand gestures using key video object plane, *Proc. IEEE Region 10 - Asia-Pacific Conf. (TENCON)*, (2004) 21-24
8. Bhuyan, M.K., Ghosh, D., Bora, P.K.: Key video object plane selection by MPEG-7 visual shape descriptor for summarization and recognition of hand gestures, *Proc. 4th Indian Conf. Computer Vision, Graphics and Image Processing (ICVGIP)*, (2004) 638-643