Rule Based Trajectory Segmentation Applied to an HMM-Based Isolated Hand Gesture Recognizer

Jounghoon Beh[1], David Han[2], and Hanseok Ko[1,3]

[1] Institute for Advanced Computer Studies, University of Maryland, College Park, USA
{jhbeh,hsko}@umiacs.umd.edu
[2] Office of Naval Research, USA
david.k.han@navy.mil
[3] School of EE, Korea University, Seoul, Korea
hsko@korea.ac.kr

Abstract. In this paper, we propose a simple but effective method of modeling hand drawn gestures based on their angles and curvature of the trajectories. Each gesture trajectory is composed of a unique series of straight and curved segments. In our Hidden Markov Model (HMM) implementation, these gestures are modeled as connected series of states analogous to series of phonemes in speech recognition. The novelty of the work presented here is the automated process we developed in segmenting gesture trajectories based on a simple set of threshold values in curvature and accumulated curvature angle. In order to represent its angular distribution of each separated states, the von Mises distribution is used. Likelihood based state segmentation was implemented in addition to the threshold based method to ensure that gesture sets are segmented consistently. The proposed method can separate each angular state of training data at the initialization step, thus providing a solution to mitigate ambiguity on initializing HMM. For comparative studies, the proposed automated state segmentation based HMM initialization was considered over the conventional method. Effectiveness of the proposed method is shown as it achieved higher recognition rates in experiments over conventional methods.

Keywords: Trajectory segmentation, hand gesture recognition, hidden Markov model, HMM initialization.

1 Introduction

The proper presentation of hand gesture motion is essential for the performance of an automatic hand gesture recognition system. The trajectory of motion is widely used as a distinct feature to represent not only hand gesture recognition [1],[2] but also action recognition [3] and hand-written character recognition [4],[5],[6].

Hand gesture in this work is represented by a trajectory composed of spatio-temporal sequence of hand positions in 2-D. In addition gesture trajectory can be decomposed into several strokes and curves that maintain a consistent angular tendency. Hand signal modeling in most of the related research has been based on classifying each of such angular tendencies, and connecting them to a Markov chain

C. Stephanidis (Ed.): Posters, Part II, HCII 2011, CCIS 174, pp. 146–150, 2011.

to add temporal property such as Hidden Markov Model (HMM) [7]. In HMM each stroke or curve is modeled as a state with statistical parameters such as mean, mixture weight, and variance that are optimized by Baum-Welch re-estimation formulae [8] with training data. However, in practice, some of the parameters such as a number of states comprising a gesture vocabulary are determined manually. The number of states is an important parameter because an excessive number of states can result an over-fitting problem if the number of training samples is insufficient compared to the number of model parameters. When there are insufficient number of states, HMM's ability to discriminate correctly in turn is reduced. A common way of determining the optimal number of states is by trial and error.

Gunter and Bunke [4] optimized the number of state by iterative refinement of the recognizer performance. Starner, Weaver, and Pentland [1] determined the number of state as a constant regardless of gesture pattern. Lee, Kim, and Kim [5] used the skeleton pattern of handwritten character to determine the number of state for on-line handwriting recognition. Elmezain *et al.* [2] empirically assigned the number of state based on the hand motion trajectory.

In this paper, we propose a Figure-based Trajectory Segmentation (FTS) algorithm to automatically determine the number of states based on a simple set of rules in terms of trajectory angles.

2 Proposed Method

2.1 Figure-Based Trajectory Segmentation

Gesture trajectories are acquired by sampling centroid locations of the hand. The proposed method of the FTS segments the trajectory into a series of states by dividing them as either a straight line or a curve. For a straight line segment, curvature of the segment is expected to be negligibly small. For a curved section, its curvature value is finite and is expected to remain below some value. By a simple thresholding, these two types of trajectory segments can be easily separated. If, on the other hand, the curvature is significantly large, it can be inferred that the trajectory exhibited an abrupt heading change such as a vertex. Additional segmentation may be performed on curves by considering their cumulative value of the curvatures by another threshold. Upon exceeding the threshold, the curve section is segmented, and the trajectory is considered to have transitioned into another segment. Our implementation of the FTS is illustrated in Fig. 1. Curvature is computed [3] by

$$k(t) = \frac{\| \mathbf{v}(t) \times \mathbf{a}(t) \|}{\| \mathbf{v}(t) \|^3} \tag{1}$$

where $\mathbf{v}(t)$ and $\mathbf{a}(t)$ denote velocity and acceleration of hand motion at time t respectively. The cumulative curvature angle is defined by

$$k_c(t) = \sum_{i=t-\tau}^{t} k(i) \| \mathbf{v}(i) \| \tag{2}$$

where τ denotes the time at previous segmentation point.

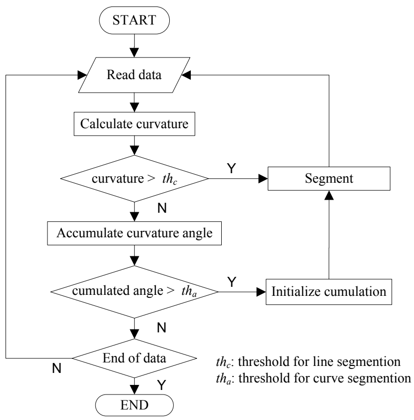

Fig. 1. Flowchart of the proposed FTS

The number of segment is adjustable by tuning the angular change threshold. In this work its value is set to 80°. In Fig. 2 the examples of segmentation for hand gesture trajectory "A" and "D" are depicted.

(a) Hand motion trajectory for letter "A" (b) Hand motion trajectory for letter "D"

Fig. 2. Example hand motion trajectories "O" mark denote the detected turning point of segmentation

2.2 Determining Number of States

The proposed FTS method was applied to a training database of a set of alphanumeric hand gesture symbols. Due to variability of gesture motion from different individuals, it has been observed that a same gesture symbol may be divided into different number of segments. Among various segmentation lengths of these symbols, we selected the number of states shown to be the most frequent in the training set as the number used for the construction of the HMM models.

2.3 HMM State Parameter Initialization

Since the angular heading of hand motion is considered as an input feature of our recognizer, we employ the von Mises probability density function (PDF) [9] to represent the state distribution of the angular heading, x. The von Mises PDF is given by

$$f(x \mid \mu, \kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)} \tag{3}$$

where $I_0(\cdot)$ is the modified Bessel function of order zero. μ and κ are, respectively, the circular mean and concentration that are analogous to the mean and variance in the normal distribution.

For each state in each gesture HMM, the segmented trajectory data are used to estimate the parameter μ and κ initially. Then, we refine the segment boundary (turning point) by aligning input feature to corresponding state using (3). This procedure is repeated iteratively until there is no change in the segment boundary position. After this refinement the final update of μ and κ are to be the initialized parameters.

3 Performance Evaluation

3.1 Experimental Settings

The following are the set of alpha-numeric symbols we have chosen for our initial study:

1. Numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9
2. Alphabet: A, B, C, D, X, Y, Z
3. Symbols: ↑, ↓, ←, →, @

The hand gesture database consists of samples from 22 people. Three separate hand gesture sets were collected from each individual. For evaluation, 11 people's data are used for training the gesture recognition HMM, and the remaining data were used for testing. Video frame rate in the database was 40 Hz and the image resolution was 160 x 120.

3.2 Results

We compared the performance of the proposed method to conventional approaches [5, 6]. In Table 1, the fixed number method sets the number of state of each HMM to a constant, and the Bakis method sets the number of state of each HMM to the average length of the corresponding sequence of feature multiplied by a constant *f*.

In our implementation of the fixed number method, the best result was obtained with 4 states, and for the Bakis method we obtained the best result with setting *f* to 0.08.

Table 1. Performance comparison

	Fixed number	Bakis	Proposed method
Recognition error rate (%)	9.53	12.99	7.50

From the result the proposed method achieves the lowest error rate. Since training the HMM is based on the maximum likelihood estimation principle, a proper initial estimation is essential for obtaining globally maximized likelihood function for good recognizer performance. The proposed segmentation inherently clusters angular tendencies in a gesture symbol trajectory. It determines the number of states in HMM consistently, therefore it would also result proper initialization the parameters of each state. Another advantage of the proposed method is that it automatically determines the number of states, unlike other method that require manual trial-and-error procedure, thus it simplifies the training part of the algorithm.

4 Conclusions

A rule based trajectory segmentation method to initialize hand gesture HMM is proposed in this paper. The number of state in HMM is automatically determined by the number of segments in the hand gesture by a simple rule based algorithm. Those segments are used to initialize statistical parameters of state of each HMM. From the experimental result, the proposed method reduced the error rate by an average of 31.8% over the conventional methods. Advantages of the proposed segmentation method are that the training phase is simple and yet consistent in making good initial estimates for the HMM.

References

1. Starner, T., Weaver, J., Pentland, A.: Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans. PAMI 20(12), 1371–1375 (1998)
2. Garg, P., Aggarwal, N., Sofat, S.: Vision based hand gesture recognition. World Academy of Science, Engineering and Technology 49 (2009)
3. Rao, G., Yilmaz, A., Shah, M.: View-inveriant representation and recognition of actions. International Journal of Computer Vision 50(2), 203–226 (2002)
4. Gunter, S., Bunke, H.: HMM-based hand written word recognition: on the optimization of the number of states, training iterations and Gaussian components. Pattern Recognition 37, 2069–2079 (2004)
5. Lee, J.J., Kim, J., Kim, J.H.: Data-driven design of HMM topology for on-line handwriting recognition. In: The Proceedings of 7th International Workshop on Frontiers in Handwriting Recognition, pp. 239–248 (2000)
6. Gunter, S., Bunke, H.: Optimizing the number of states, training iterations and Gaussian s in an HMM-based Handwritten Word Recognizer. In: The Proceedings of the 7th International Conference on Document Analysis and Recognition (2003)
7. Juang, B.-H., Rabiner, L.R.: A probabilistic distance measure for hidden Markov models. AT&T Technical Journal 64(2), 391–408 (1985)
8. Baum, L.E.: An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. Inequalities 3(9), 1–8 (1972)
9. Fisher, N.I.: Statistical Analysis of Circular Data, Ch. 3.3.6 The von Mises distiribution, pp. 48–56. Cambridge University Press, Cambridge (1993)