Recognition of Continuous Mouse Gesture Sequences

I. Problem Statement

Gestures, ranging from simple mouse movement to complex full-body motion [1], are a proven form of input in human-centric user interfaces. However, a significant barrier to adoption of gesture-centric interfaces is the problem of accurate gesture recognition, which includes both recognition of individual gestures and individual gestures comprising a gesture sequence. Gesture recognition, henceforth generalized as symbol recognition, frequently involves correct segmentation of a symbol sequence and accurate recognition of individual symbols.

The gesture problem that we focus on in this paper is temporal segmentation during on-line recognition of dynamic mouse gestures. In our proposed test system, the user draws a continuous sequence of user-define gestures using a mouse. Our approach, described below, captures the mouse position over time, and thus is specific to on-line gesture recognition. However, we view this work as a step towards addressing the more general problem of off-line recognition of custom-made symbols, which, to our knowledge, has not been addressed in the literature. Recognition of custom gestures is an important problem in law enforcement, where the United States Federal Bureau of Investigation Safe Streets and Gang Unit commonly encounters handwritten communication of custom gestures [2].

II. Related Work

Yang et al [3] present work on recognition of individual gestures in continuous gesture sequences, but unlike our work, they train HMMs on continuous gestures; we view this as a deficiency due to the *a priori* definition of gesture sequences. Groner [4] segments symbol sequences primarily by placing an interface constraint upon the user; the user is required to temporarily halt input between symbols. We view this approach as deficient due to the heavy user burden. The work most closely related to our problem and approach is focused on recognition of Chinese handwriting [5]. Hong et al use an iterative segmentation technique that uses whitespace separation to split character sequences into individual characters. Their approach is similar to ours, but is intended for either off-line or on-line recognition problems, and thus does not take advantage of the temporal data that we employ. To our knowledge, there is no directly comparable system in the literature; however, we will compare our results to those presented by Hong et al [5] due to the similarity in our problem and approach.

III. Proposed Implementation

Our problem focuses on segmentation of user-defined gestures. As a result, we cannot leverage grammar or other linguistic features, as was done successfully by Starner et al [6]. Additionally, we cannot assume the existence of any markers to separate gestures, as is typically found in Chinese writing [5]. As a result, we employ an approach based on an efficient "search" over the space of segmented gesture sequences, aiming to find the gesture segmentation that most accurately recognizes individual gestures.

We will build upon the mouse gesture recognition system developed by Tanguay [7], which performs on-line recognition of individual mouse gestures. We will enhance his system by adding a segmentation routine that accurately splits a continuous gesture sequence. Our segmentation technique is focused on splitting the complete temporal data stream (representing 1..N individual gestures) into (shorter) sub-streams (each corresponding to one gesture).

In our approach, the data stream is first split temporally into N sub-streams (each corresponding to an individual gesture) of identical duration equal to the mean time taken to draw an individual gesture (calculated from the training set). The duration of each sub-stream can be conceptualized as a "window" overlaid on the complete data stream; N sub-streams correspond to N windows. After configuring the initial "window set", we calculate a "window set score" which is the mean of all the individual window scores; each window score, in turn, is equal to the maximum of the probabilities between a given window sub-stream and each trained HMM. We subsequently iteratively increase the score of the window set by independently increasing or decreasing the duration of each window with a delta based on the variance in the gesture training set. The decision to modify each window is determined by comparing the window score in iteration *k* to the score in iteration *k-1*. If the score for a given window increased, the action taken in the previous iteration (either growing or shrinking the window) is repeated; conversely, if the score decreased, the alternate action should be taken. In order to ensure the process terminates, each successive step in the iteration will modify the window duration by a smaller percentage than the previous iteration.

IV. Proposed Evaluation

Because our proposed work is to implement the aforementioned segmentation routine, we propose to evaluate the success of this routine by measuring the success rate for correctly segmenting a gesture sequence into the correct gestures. We define the success rate as the number of correctly identified individual gestures from a number of gesture sequences.

To avoid author bias during evaluation, we will conduct a micro-study involving one user. The user will engage in a brief training phase; during this phase, the user will define a half-dozen custom gestures and train HMMs by performing each gesture a number of times. After the training phase, the user will engage in a recognition phase, where the user will arbitrarily create gesture sequences of lengths 1, 2, and 3. Gestures with length 1 serves to validate the system accuracy for individual gestures. Gestures of lengths 2 and 3 serves to quantify the success rate of our segmentation routine.

As noted above, our problem and segmentation routine is comparable with the work completed by Hong et al [5]. As a result, we will compare our success rates with their published results.

V. References

[1] S. Mitra and T. Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.

[2] Lyddane, Donald. United states attorneys' bulletin. Technical report, United States Department of Justice Executive Office for United States Attorneys, May 2006.

[3] J. Yang, Y. Xu, and C. S. Chen. Gesture interface: Modeling and learning. In *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pages 1747–1752, 1994.

[4] G. F. Groner. Real-time recognition of handprinted text. In *Proceedings of the November 7-10, 1966, fall joint computer conference*, page 591âĂŞ601, 1966.

[5] C. Hong, G. Loudon, Y. Wu, and R. Zitserman. Segmentation and recognition of continuous handwriting chinese text. *International journal of pattern recognition and artificial intelligence*, 12(02):223–232, 1998.

[6] T. Starner, J. Makhoul, R. Schwartz, and G. Chou. Online cursive handwriting recognition using speech recognition methods. In *Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on*, pages V–125. IEEE, 1994.

[7] D. O. Tanguay Jr. Hidden markov models for gesture recognition. Master's thesis, Massachusetts Institute of Technology, 1995.