Recognition of Continuous Mouse Gesture Sequences

I. Problem Statement

Gestures, ranging from simple mouse movement to complex full-body motion, are a proven form of input in human-centric user interfaces [1]. However, a significant barrier to widespread adoption of gesture-centric interfaces is the problem of accurate recognition of gestures, both in isolation and in sequence. Gesture recognition, frequently involves solving two problems: segmentation of a gesture sequence and recognition of individual gestures.

We focus on the problem of temporal segmentation during on-line recognition of a sequence of custom mouse gestures. Our approach, described below, captures the mouse position over time, and thus is specific to on-line gesture recognition. However, our work is a step towards addressing the more general problem of off-line recognition of custom symbols, which is an important problem in law enforcement, where the United States Federal Bureau of Investigation Safe Streets and Gang Unit commonly encounters handwritten communication involving custom gestures [2]. To our knowledge, this problem has not been addressed in the literature.

II. Related Work

Yang et al [3] present work on recognition of individual gestures in continuous gesture sequences, but unlike our work, they train Hidden Markov Models (HMMs) on continuous gestures; we view this as a deficiency due to the *a priori* definition of gesture sequences. The work most closely related to our problem and approach is focused on recognition of Chinese handwriting [4]. 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 [4] due to the similarities between 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 [5]. Additionally, we cannot assume the existence of any markers that separate gestures, as is typically found in Chinese writing [4]. As a result, we employ an approach based on an efficient "search" over the space of segmented gesture sequences, aiming to find the most accurate gesture segmentation.

We will build upon the mouse gesture recognition system developed by Tanguay [6], which performs on-line recognition of individual mouse gestures via HMMs. We will enhance his system by adding a segmentation routine that accurately splits a continuous gesture sequence.

In our approach, we first split the multi-gesture data stream 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 of a match between the window data and each trained HMM. We subsequently iteratively increase the window set score by independently increasing or decreasing the duration of each window by a delta based on the time variance of gestures in the training set. If the window score increased between iterations *k* and *k-1*, we repeat the previous action; otherwise, we try the other action. Our routine terminates when the delta change reaches zero.

IV. Proposed Evaluation

We propose to evaluate our routine by measuring the success rate for segmenting a gesture sequence into the correct gestures. We define the success rate as the percentage of correctly identified individual gestures. As noted above, our problem and segmentation routine is comparable with the work completed by Hong et al [4]. As a result, we will compare our success rates with their published results.

To avoid author bias during evaluation, we will conduct a micro-study involving one user. 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.

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] 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.

[5] T. Starner, J. Makhoul, R. Schwartz, and G. Chou. On-line 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.

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