Recognition of Continuous Mouse Gesture Sequences

I. Problem Statement

Gestures are a proven form of input in many human-centric systems. As Nielsen describes [1], gestures are particularly effective when the user interface paradigm moves away from a command-oriented approach and toward a task-oriented approach. Command-oriented user interfaces require the user to request an action and subsequently wait for completion, which is clearly sub-optimal when the desired sequence of actions is known in advance.

In our proposed system, the user draws a continuous sequence of gestures. The system recognizes each gesture in the sequence by matching against a trained gesture database. The primary benefit of our proposed system is its suitability for completing task-oriented work. The system provides a capability to queue actions for sequential execution, eliminating the need for a user to alternate between providing input and waiting for an action to be completed, saving scarce time and resources.

II. Related Work

Mitra and Acharya, in their survey on gesture recognition techniques [2], note that Hidden Markov Models (HMMs) are a popular tool used to accurately recognize gestures in dynamic gesture recognition problems. Additionally, they present a categorization of gestures, defining pantomimes as "gestures depicting objects or actions, with or without accompanying speech".

Our work is focused on the problem of using HMMs to recognize individual pantomimes embedded in a "composite pantomime" (pantomime sequence), constructed through simple mouse gestures. Yang et al [3] discuss work similar to ours, with the primary difference being that they train HMMs on sequential gestures; in contrast, our approach is to exclusively train on individual gestures, which will then be used to identify each gesture in a gesture sequence. Our approach is an extension to the work by Tanguay [4]; we propose to recognize more than one gesture given as input.

III. Proposed Implementation

As implied above, we propose to implement a system that uses HMMs to recognize individual gestures within gesture sequences. In our system, the user will manually perform a *single* gesture multiple times to train an HMM. We note that the user will *not* need to conduct training for gesture sequences; eliminating that step reduces the training time and lowers the barrier for system adoption.

During the recognition phase, the user will create multi-gesture sequences. As Mitra et al note, a difficult aspect of this problem is segmentation ambiguity. Our proposed approach to solving the segmentation problem is to apply a technique that aims to automatically segment the complete data stream (representing 1..N individual gestures) into (shorter) sub-streams (each corresponding to one gesture).

Initially, the data stream is 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. 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 mathematical product of 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. 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

We propose to evaluate the system by conducting a micro-study involving one user. Because our proposed work is to develop a sequential gesture recognition algorithm, we propose to evaluate the system primarily by collecting metrics that measure recognition accuracy. During the study, the user will complete an initial training step and a subsequent recognition step, consisting of multiple phases. Recognition phase n will consist of the user providing gesture sequences of length n. We will score the system by computing the proportion of correctly identified individual gestures, where "correct identification" includes both placement of the individual gesture in each gesture sequence and correct identification of the intended gesture. We will conduct three phases of recognition testing (thus evaluating sequential gestures with a maximum length of three), primarily because we anticipate three phases will provide us with enough data to determine the algorithm's effectiveness.

V. References

[1] J. Nielsen. Noncommand user interfaces. *Communications of the ACM*, 36(4):83–99, 1993.

[2] 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.

[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] D. O. Tanguay Jr. *Hidden markov models for gesture recognition*. PhD thesis, Massachusetts Institute of Technology, 1995.