# Recognition of Continuous Mouse Gesture Sequences

## I. Problem Statement

Gestures are a proven form of input in many human-centric systems. Gestures can be captured from a wide array of sources; from something as simple as the movement of a mouse to something as complex as full-body motion [1]. However, any system driven by gestures is intuitive and more effective to work with when compared to the traditional command-oriented approach [2]. In scenarios involving repeatedly doing a set of tasks through a predetermined sequence of menus, clicks and dialog windows, a gesture-based paradigm of interaction with the system would provide a refreshing alternative to the user. In addition, if the system is robust in identifying the gestures with a high degree of accuracy, a gesture-based interface will be a more accessible way for people with conditions like Carpal-Tunnel to control the system.

In our proposed system, the user draws a continuous sequence of gestures using a mouse. The system recognizes each gesture in the sequence by matching against a trained gesture database. Once a gesture has been identified, an action associated with it is performed. However, if an unknown gesture is encountered, the user can add it to the training database alongwith an associated action.This 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. This is one of the primary benefits of our system, in addition to the previously mentioned potential ways to utilize it.

## II. Related Work

Mitra and Acharya, in their survey on gesture recognition techniques [1], 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, in which they train their HMM model on continuous gestures of varying lengths. Once the system is trained, it is able to identify one or more gestures, based on the models constructed. However, this method has a long training duration and is not robust enough to recognize untrained symbols. Our aim is to exclusively train on individual gestures, which will then be used to identify each gesture in a gesture sequence.

We propose to base our work on the doctoral thesis of Tanguay [4]. However, in addition to recognizing individual gestures, we plan to extend his technique to identify multiple gestures.

## 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, which our system will try to recognize. 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.

This approach seems to be a good starting point and provides a base on exploring other ideas to tackle the problem of segmentation. In addition to our approach, we also plan to explore strategies used to identify segments in handwriting recognition [5] and try to adopt it in our framework.

## 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 an d 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 then compare our results for individual gesture recognition with [4] and those for multiple gesture recognition with [3].

In addition, we will ask for qualitative feedback about the usability and usefulness of the system compared to the normal methods that the user employs to complete his task.

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

[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.

[5] Y. Zheng, H. Li, and D. Doermann. The segmentation and identification of handwriting in noisy document images. *Document Analysis Systems V*, pages 95–105, 2002.