A Gesture Learning and Recognition System for Multitouch Interaction Design (poster)

Sashikanth Damaraju and Andruid Kerne
Interface Ecology Lab at Texas A&M University
{damaraju, andruid}@cs.tamu.edu

Abstract

To design interfaces for more embodied interaction, we build a system that learns and recognizes multitouch gestures – movements of human fingers on a multitouch surface. We form an example vocabulary of gestures that demonstrates the subtleties of finger movements that can be recognized. A video processing system extracts a feature set representing these movements. Hidden Markov Models (HMMs) are used to learn gestures by example, and consistently recognize them. We present an evaluation that demonstrates the robustness of the methodology. Implications for interaction design are presented.

1. Introduction

As digital systems become more ubiquitous, more integrated into the fabric of human life, we need to discover more human-centered forms of interaction. Multitouch systems capture movements of multiple human fingers on an interactive surface enabling direct hands on interaction. This modality of input enlarges the dimensionality of interaction in comparison with the conventional mouse-pointer interface. Human hands afford more degrees of freedom due to the dexterity inherent in their operation. Recent developments facilitate access to the required multitouch hardware, bringing the potential for supporting new forms of human-centered experience; this potential is as yet hardly realized. Our contribution is a system that understands human expressions of intent by learning and recognizing gestures performed on a multitouch surface.

Researchers have investigated multitouch interaction for decades [1]. Multitouch gestures have generally been limited to coarse movements such as the 'wipe' and 'pile-n-browse' techniques [8], simple primitives such as 'flicks' [5] or fixed combinations of 'chords' [6]. We present a technique that learns and recognizes sophisticated hand gestures on a multitouch surface, providing end-user customizability. Our long term objective is to give human participants an interactive experience in which embodied gestures performed by the human hand are mapped to actions in ways that are natural, meaningful and intuitive.

2. Learning and Recognition Pipeline

Computationally, we define a *gesture* as a human action that begins with placing one or more fingers on the interactive surface, and ends when those fingers are lifted off the surface. A sequence of frames from a camera is captured to record each gesture. For each frame of the sequence, a feature set is derived by an

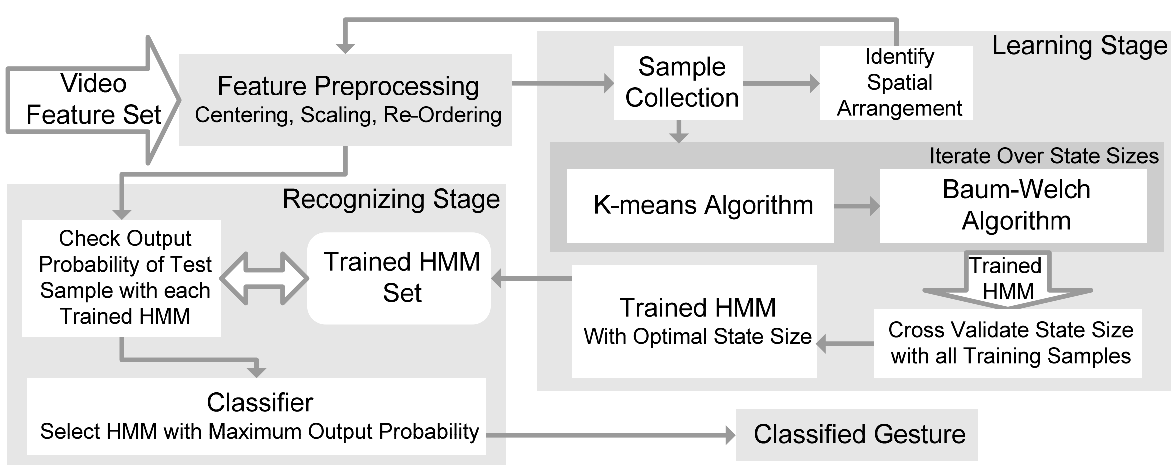

Figure 1. Multitouch Gesture Learning and Recognition Pipeline. Identification of spatial arrangement in the learning phase feeds back to feature pre-processing.

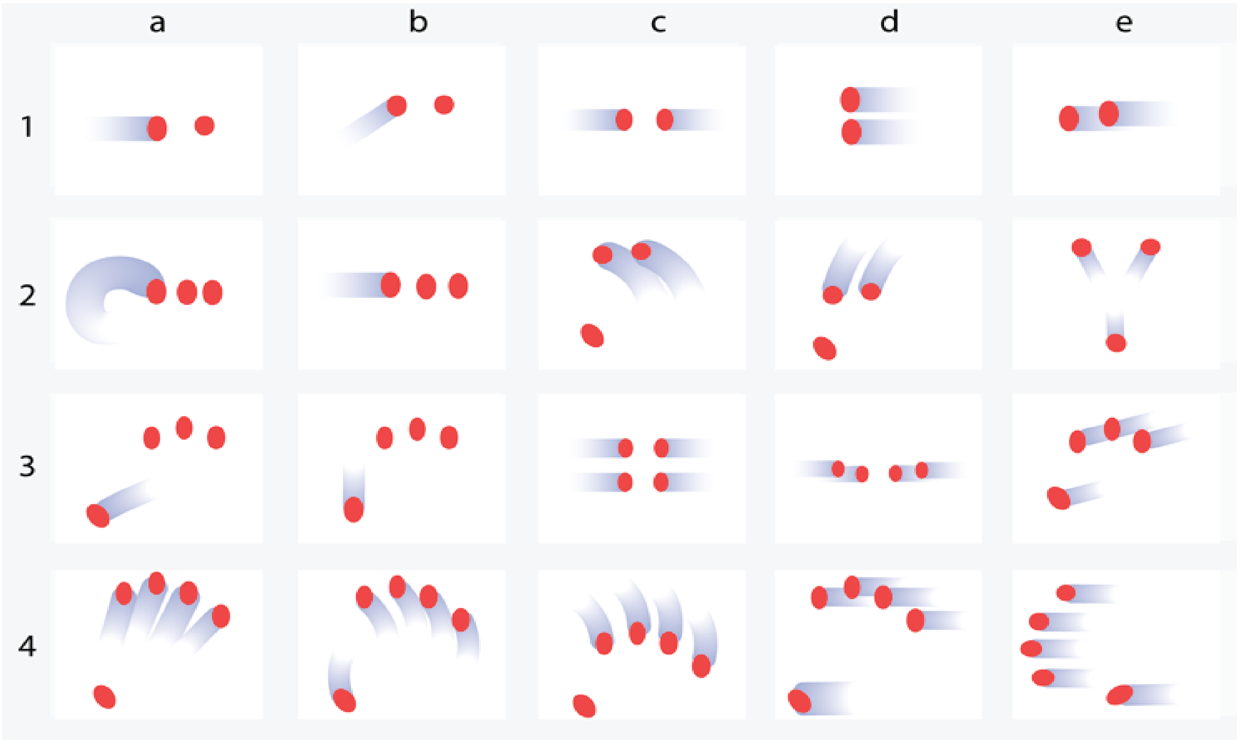

Figure 2. Vocabulary of gestures demonstrating the subtleties that can be recognized. The starting positions of the fingers are shown in red, the blue trails shows the path during the gesture.

open source video processing and feature acquisition toolkit [3], and used to recognize the gesture through a pipeline of processing stages. Figure 1 shows the stages involved in the learning and recognition pipeline.

2.1. Training Set

Our goal is to build a practical system that uses natural movements of the hand to perform complex actions within an application. To start, we developed a vocabulary of 40 different gestures for testing the current learning and recognition system (Figure 2). The gestures were formed to exemplify subtle differences in initial positions of the fingers and directions of movement. They are grouped by the number of fingers used to perform each one. Twenty samples of each gesture were collected from each of 10 users, 9 male and 1 female, all of whom are right-handed. We expect that the ability to recognize a vocabulary of gestures with subtle variations, such as the direction of movement of the thumb in gestures 3a and 3b, will be critical for developing expressive applications.

2.2. Feature Preprocessing

The feature preprocessing stage normalizes the data in the X, Y feature space to allow the gesture to be recognized irrespective of where it was performed on the screen. The samples of the training set are also scaled to a constant size for improved recognition.

The ordering of the fingers in the feature set provided by video processing is determined by the temporal order of placement of the fingers on the surface in the first frame of a gesture. Without preprocessing, this ordering is inconsistent across different samples of the same gesture, which is a problem for the learning and recognition system. We develop a solution to this problem of consistent feature ordering, which identifies the spatial arrangement of fingers as horizontal, vertical, or radial. The spatial arrangement is identified for the first frame of each gesture during the learning stage, and then applied during preprocessing during subsequent frames (see feedback line at the top of Figure 1).

2.3. Learning Stage

To learn gestures, we have randomly selected 20 samples of each gesture from across all users. The remaining samples are used to determine the accuracy of classification of the recognition system. During the learning phase, after the observations have been collected and preprocessed, they are passed to a K-means clustering algorithm. The K-means clustering

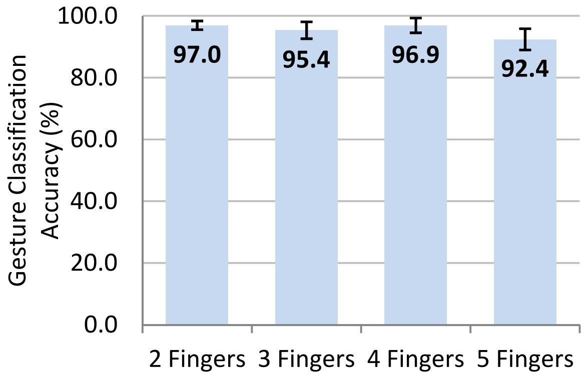

Figure 3. Classification accuracy results for 20 runs of learning gestures with a random training set of 20 samples, and testing with the remaining samples. The gesture vocabulary of 40 gestures is grouped by the number of fingers in each gesture.

efficiently forms an initial estimate of the HMM parameters. The estimate is, in turn, passed to the Baum-Welch algorithm, which tunes these parameters to return a high probability for only the given training sample sequences.

2.4. Recognition Stage

For each sample to be recognized, we pass the sample into each trained HMM. The probabilities output from each HMM are collected and compared. We label the sample as the multitouch gesture whose model returns the highest probability.

3. Results

A random selection of 20 training samples for each gesture from the entire set of all 10 users is used to train HMMs, and the remaining samples used to determine the overall classification rate of the system. The classification rates shown in Figure 3 are for the gestures using two, three, four and five fingers are 97.0%, 95.4%, 96.9% and 92.4%, for an average gesture recognition rate of 91.5%, and the result is statistically significant ($p < .0001$). These results show that our technique classifies gestures within a reasonable accuracy for practical use. As the HMM can be trained using samples from more than one user, it can also be stored within an application or shared across users to avoid re-training.

4. Conclusion & Future Work

We have presented a system that can successfully learn and recognize gestures performed by the human hand on multitouch surfaces. A vocabulary of gestures was built to test the system's ability to distinguish the subtle differences between the gestures. Issues to increase consistency between samples of a gesture were raised, and solutions were developed to solve them. The system can recognize samples of gestures successfully.

This gesture recognition system will serve as a foundation for advanced interactive visual systems. Gestures will be used to design interaction techniques for performing complex actions. Future work will develop interaction techniques mapping gestures to actions of searching, browsing, collecting and organizing within combinFormation, a mixed-initiative information composition platform [2]. The system agents process documents from the internet and form image-text surrogates, placing them on a composition space. combinFormation provides the user with the ability perform a large number of operations, such as expressing interest to direct the agent, navigating to the source document, actions for image manipulation and text editing. Inspired by prior work on choreographing human movements for human-computer interaction [6], we will design interaction techniques to enable participants in performing these complex tasks with well choreographed and embodied gestures. We hypothesize that interfaces with intuitively designed gesture-actions mappings will result in a more expressive, engaging and fluid human computer interaction.

5. References

[1] Buxton, W. Multitouch Systems That I Have Known And Loved. 2008 http://www.billbuxton.com/multitouchOverview.html

[2] Kerne, A., Koh, E., et al., A Mixed-Initiative System for Representing Collections as Compositions of Image and Text Surrogates, *Proc JCDL 2006*.

[3] Nuigroup, Touchlib: A Multi-Touch development kit http://nuigroup.com/touchlib/

[4] Rabiner, L. R., A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, Feb 1989, 77 (2), 257–286.

[5] Tse, E., Shen, C., Greenberg, S., Forlines, C., Enabling interaction with single user applications through speech & gestures on a multi-user tabletop. *Proc ACM AVI 2006.*

[6] Webb, A., Kerne, A., Koh, E., Joshi, P., Park, Y., Graeber, R., Choreographic Buttons: Promoting Social Interaction through Human Movement and Clear Affordances, *Proc ACM Multimedia 2006.*

[7] Westerman, W, Hand Tracking, Finger Identification, and Chordic Manipulation on a Multi-Touch Surface, *Phd Diss* http://www.ece.udel.edu/~westerma/main.pdf

[8] Wu, M., Shen, C., Ryall, K., Forlines, C., Balakrishnan, R., Gesture Registration, Relaxation, & Reuse for Multi-Point Direct-Touch Surfaces, *Proc IEEE TableTop 2006*, 185-192.