Accelerometer-Based Hand Gesture Recognition using Feature Weighted Naïve Bayesian Classifiers and Dynamic Time Warping

David Mace
Archbishop Mitty High School
davidmace13@
mittymonarch.com

Wei Gao
Boston University
ECE Department
davidgao@bu.edu

Ayse K. Coskun
Boston University
ECE Department
acoskun@bu.edu

ABSTRACT
Accelerometer-based gesture recognition is a major area of interest in human-computer interaction. In this paper, we compare two approaches: naïve Bayesian classification with feature separability weighting [1] and dynamic time warping [2]. Algorithms based on these two approaches are introduced and the results are compared. We evaluate both algorithms with four gesture types and five samples from five different people. The gesture identification accuracy for Bayesian classification and dynamic time warping are 97% and 95%, respectively.

Author Keywords
Gesture recognition; accelerometer; Bayesian classifier; feature separability weighting; dynamic time warping

ACM Classification Keywords
I.5.2[Pattern Recognition]:Design Methodology – Classifier design and evaluation

INTRODUCTION
Gesture recognition is a growing area of interest because it provides a natural, 3D interface for humans to communicate with computers. In this paper, we present two methods to recognize hand gestures using a 3-axis accelerometer. Using an accelerometer has lower complexity and cost compared to camera-based gesture recognition. In addition, accelerometers worn on the hands provide better flexibility as the user does not need to face a particular direction as in the case with the camera.

The most prevalent algorithm for accelerometer-based gesture recognition is the Hidden Markov Model (HMM) [3]. Using a sufficient set of training samples for each gesture type, an HMM provides very high accuracy in recognition and robustness. This algorithm, however, requires a huge amount of training samples. In this paper, we discuss two approaches that require a much fewer number of training samples but still provide high accuracy.

IUI'13 Companion, March 19–22, 2013, Santa Monica, CA, USA.
ACM 978-1-4503-1966-9/13/03.

EQUIPMENT
Our implementation uses a TI eZ430-Chronos Watch, which is cheap and simple to use, as the accelerometer data provider. The watch contains a VTI-CMA3000 3-axis accelerometer, with a measurement range of 2g, 8-bit resolution, and 100Hz sampling rate.

We use an ASUS TF300T Android tablet to run our algorithms (which are all implemented with Java); however, our implementation can be used with any Android device and can be ported to other mobile platforms. The tablet receives accelerometer data from the watch through an RF-receiver with USB interface, which is recognized as a serial port inside of Android.

METHODS
The proposed gesture recognition methods can be split into three main phases: preprocessing, rotational normalization, and weighted feature classification or dynamic time warping as the last step.

Preprocessing:
The raw data set received from the accelerometer is noisy and contains still frames so the data must be adjusted before they can be properly classified.

The jolty nature of hand motions and the discrete sampling of the gestures contribute white noise to the data. The actual gesture motion has a very low frequency so a first order low-pass algorithm is used to extract the motion. Still frames at the beginning and end of the data that are not part of the gesture are also removed. Still frames are detected by checking the average acceleration in each 0.5 second window and checking whether acceleration is below a constant threshold.

Rotation Normalization:
One issue in gesture recognition that has not been explored in depth is correctly recognizing the same gestures drawn in different planes of motion. Normalizing the gestures' rotation provides greater accuracy and allows for more realistic, orientation-independent motion.

In this phase, first, the best fit plane of the acceleration vectors is found. The rationale behind this is that if the motion lies in a single plane, then the acceleration vectors of a closed shape (e.g., a circle) should on average lie in that main plane. As there could be many motion vectors in the motion that do not lie in the main plane even after using a low-pass filter, all acceleration segments between points of inflection are added up to form one vector. In this way, we can identify the general direction of the user's motion, rather than identifying each motion segment.

After the best fitting main plane is found, each vector is normalized relative to this plane. To account for rotation around the z axis, the best fitting line inside the main plane is found and all points are re-calculated relative to this line.

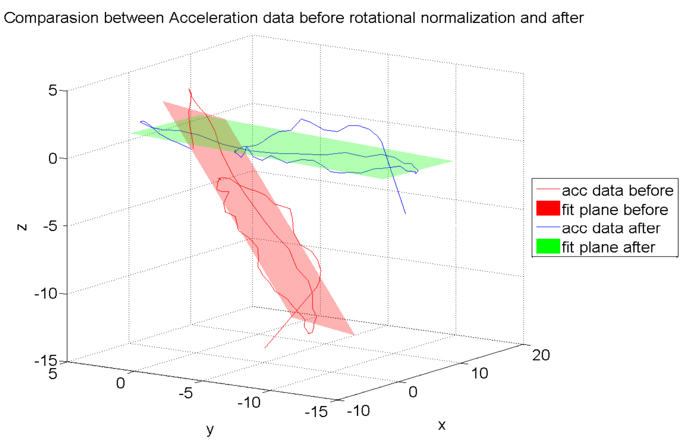

Figure 1. The red curve represents a circle gesture performed in the yz plane and the blue curve represents the result after rotational normalization is applied.

Feature Weighted Naïve Bayesian:

Naïve Bayesian Classification [1] is a promising technique in gesture recognition because it can make accurate predictions by using statistical measures to calculate *membership probabilities*. In our implementation of this algorithm, twenty statistical features are extracted from the acceleration data. These include common statistical measures such as interquartile range, average energy, maximum of absolute value, and standard deviation.

Before a user operates the system, the user registers a set of training gestures. A weight is calculated for each feature type based on the similarity of feature measures of the trained gestures. A value of close to 1 represents very precise measures and a value close to 0 represents imprecise measures. When the user is running the gesture recognition system, feature measures are extracted from the user's registered gesture. The proximity of each feature measure to the average trained feature measure of each gesture type is calculated by a normal distribution. Then this proximity value is multiplied by the feature weight that was calculated in the training phase. All of these multiplied values are added together and the system predicts the gesture type with the greatest value as the user gesture.

Dynamic Time Warping (DTW):

DTW is a widely used algorithm in gesture recognition that calculates the similarity between two time-series data sets. This algorithm is based on the idea that to find the time-independent similarity between a gesture and a template, the i^{th} point of the gesture can be aligned (warped) to the j^{th} point of template [2].

In this algorithm, first a matrix is calculated. Each element a(i,j) in the matrix represents the geometrical distance between the sample data at time t(i) and template data (collected in training phase) at time t(j). Any gesture that is "close" to the template data is likely to be of the same gesture type. Second, a path in the matrix "a" is found so that among all of the paths from a(0,0) to a(n,m), the sum of all the elements on the path is minimized.

The above two steps give a value representing the similarity between one sample data set and one template (training) data set. Then these steps are completed for all of the sample/template data pairs. The pair that has the smallest "path sum value" indicates the predicted gesture.

RESULTS

We test both techniques using 5 gesture samples of 4 gesture types (circle, figure eight, square, star) from 5 different people. The average accuracy is 97% for the feature separability weighted Bayesian Classifier, and 95% for the dynamic time warping.

Both of the proposed methods have comparable accuracy with Hidden Markov Models and k-mean algorithms [3,4]. However, feature separability weighted naïve Bayesian classifiers and dynamic time warping run faster on large data sets and require a fewer number of training samples.

APPLICATIONS

We built an intelligent alarm clock Android application that uses the Chronos watch to detect if a user is asleep by checking for simple gestures [5]. We are also in the process of building Android applications that leverage the Chronos watch and gesture recognition in password detection and hand motion controlled features in media players.

ACKNOWLEDGEMENT

Part of this research has been conducted during David Mace's internship in Boston University's Research Internships in Science and Engineering (RISE) program.

REFERENCES

[1] P. Paalanen. "Bayesian Classification Using Gaussian Mixture Model and EM Estimation: Implementations and Comparisons." Lappeenranta University of Technology, Information Technology Project Report. http://www2.it.lut.fi/project/gmmbayes/downloads/doc/report04.pdf

[2] E. Keogh, C. A. Ratanamahatana, "Exact Indexing of Dynamic Time Warping", In *Proceedings of International Conference on Very Large Data Bases (VLDB),* pp. 406-417, 2002.

[3] A. D. Wilson, "Parametric Hidden Markov Models for Gesture Recognition" , *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.21, no.9, pp.884-900, Sep 1999.

[4] T. Schlomer, B. Poppinga, N. Henze and S. Boll, "Gesture Recognition with a Wii Controller", In *Proceedings of International Conference on Tangible and Embedded Interaction (TEI),* pp. 11-14, 2008.

[5] http://processors.wiki.ti.com/index.php/Persistent_alarm_clock