Gesture Recognition with Low Quality Signal over Low Energy Bluetooth 4.0

Yang Zhao and J. Brandon Laflen
Sensor and Signal Analytics Lab
GE Global Research Center
Niskayuna, New York 12309
Email: yang.zhao@ge.com, brandon.laflen@ge.com.

***Abstract*—Accelerometer-based gesture recognition enables natural user interaction of handheld tools with PC and other devices. Dynamic time warping (DTW) can achieve high recognition accuracy and does not require extensive training. However, for applications with noisy data and limited data bandwidth, the performance of DTW algorithm decreases significantly. We propose a new gesture recognition method that is robust to low quality signals. We introduce a simple template selection step to obtain robust templates for DTW. Our experimental results show that our algorithm achieves an average 92% classification accuracy even under noisy and low bandwidth situations.**

I. Introduction

Accelerometer-based hand gesture recognition is attractive for many applications in entertainment electronics and mobile computing. Various recognition techniques have been proposed and studied. Hidden Markov model (HMM)-based method is one of the most popular methods in recognition. However, unlike human speech, hand gestures do not have a standard "vocabulary," and it is often desired for users to create their own gestures. Thus, HMM-based methods are not preferred for personalized gesture recognition systems. As an alternative, dynamic time warping (DTW) aligns two signals with different time dynamics and was widely used in speech recognition [1]. DTW directly compares a time series with a template, which is obtained beforehand. It does not require a standard vocabulary library, and also does not require much training. Thus, DTW is preferred in many scenarios. In addition, it achieves high classification accuracy in hand gesture recognition [2].

When we apply DTW to our application with low signal-to-noise ratio (SNR) and low sampling rate signals, we find unsupervised DTW cannot achieve robust performance. We have chosen Bluetooth low energy (BLE; e.g., Bluetooth 4.0 Smart), which provides low cost and low power consumption, for communications in our application. The signals have low SNR and low bandwidth, and the classification accuracy of unsupervised DTW can be below 70%.

To solve this problem, we use more training data to find a robust template instead of directly using an offline gesture sample as the template. That is, we add a simple template selection step into DTW. Before online recognition, we collect several gesture samples instead of one single sample as training data, from which we calculate the DTW distance between pairs of samples to construct a distance matrix. Then we find the template as the one gesture sample that has the minimum total distance between itself and other samples. By using this selected template, the performance of DTW is greatly improved. For recognition of eight gestures as used in a Nokia study [3], the classification accuracy of the unsupervised DTW algorithm can be lower than 70%, while our new method has an average classification rate of 92%.

II. Methods

We present our dynamic time warping (DTW)-based gesture recognition algorithm in this section. We describe how we use template selection to make DTW more robust to template outliers.

Let $\mathbf{v}$ represent a three-axis accelerometer time series: $\mathbf{v} = [\mathbf{x}, \mathbf{y}, \mathbf{z}]$, where $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are x, y, z-axes accelerometer data. The DTW distance between two time series $\mathbf{v}_s$ and $\mathbf{v}_t$ is:

$$d = \text{DTW}(\mathbf{v}_t, \mathbf{v}_s) \tag{1}$$

We propose the following procedure for robust template selection. Assume we have N types of gestures to be classified $g_1, g_2, \cdots, g_N$, and for each gesture type, we have M offline training data for template selection. For the mth training sample $\mathbf{v}_{t_m}$, we use (1) to calculate the DTW distance between itself and other samples. Then for all M samples, a distance matrix $D = [[d_{m,m'}]_M]_M$ can be constructed using the training dataset for that gesture. For each sample, the total DTW distance between itself and other samples can be calculated as $d_m = \sum_{m'=1}^{M} d_{m,m'}$. That is, d_m can be used to measure the total difference of a training sample $\mathbf{v}_{t_m}$ from other samples in the training dataset. Then we choose the sample with the minimum d_m as the template for that gesture type. For each gesture type, we perform the same procedure to build a template library for all N gestures: $\mathbf{v}_t^1, \mathbf{v}_t^2, \cdots, \mathbf{v}_t^N$, where $\mathbf{v}_t^i$ is the template for the ith gesture type. Note that from the training data used in template selection, we can also obtain statistics for each of the x-, y-, and z-axis accelerometer measurements.

After offline template selection, we perform online recognition. For a particular online accelerometer reading $\mathbf{v}_s$, we can use (1) to calculate the distance $d_{i,s}$ between itself and the template $\mathbf{v}_t^i$ of the ith gesture type. For N types of gestures, we have N distances $d_{i,s}$, $i = 1, 2, \cdots, N$. The gesture type with the minimum distance calculated between its template and the sample is our estimated gesture. That is, our final gesture estimate $\hat{g}_s$ is:

$$\hat{g}_s = \arg\min_i d_{i,s} \tag{2}$$

This could be improved by using training statistics, e.g., with Mahalanobis distance.

978-1-4799-1291-9/14/$31.00 ©2014 IEEE

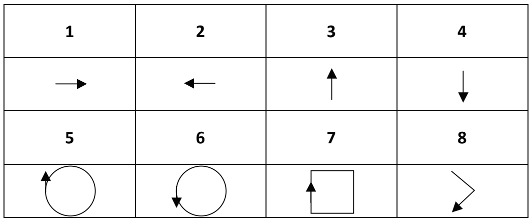

Fig. 1: Gestures and corresponding indexes.

III. Experiments and Results

A. Prototype and Experiments

Our application uses an accelerometer with decreased SNR and sampling rates compared with [2]. The study in [2] uses Bluetooth and has an accelerometer from Analog Devices, ADXL330 operating at 100 Hz. We use TI's CC2540 development kit [4] as our handheld device (BLE slave), and our accelerometer sampling rate is about 40 Hz. As a result, our sampling rate is less than half that of [2]. Further, our accelerometer has a lower SNR. The three-axis accelerometer used in [2] has a range of $-3g$ to $3g$ and noise below $3.5mg$. However, our accelerometer chip, CMA3000 has a measurement range of $-2g$ to $2g$ with average noise of $13mg$ over three axes. Thus, our system has a lower data bandwidth and a lower SNR, and the unsupervised DTW algorithm does not perform well in our scenario.

We use our prototype to record accelerometer data in our gesture recognition experiments. We used buttons on this device to enable and disable gesture recording. We performed the eight types of gestures from a Nokia study [3], as illustrated in Figure 1. Thirty samples were recorded for each type of gesture.

B. Results

We evaluate the performance of our algorithm, and compare it with unsupervised DTW developed in [2]. First, for each of eight types of gestures, we randomly choose one sample from thirty samples as measured online data. We then use the remaining sets of twenty-nine samples as offline training data to select templates for each of the eight gestures. Then we run our algorithm to classify the eight online gestures, and record classification results. We performed the above procedure 1000 times, and obtained an average classification rate of 92.3% for this dataset. We also test using fewer samples as training data, and our algorithm achieved similar performance.

Table I shows the confusion matrix from our method, and Table II shows the matrix from the unsupervised DTW method. For our method, the classification rates for three gestures (6, 7 and 8) are 100% from 1000 trials. The unsupervised DTW method (without template selection) never achieves 100% accuracy. For gesture type 4, the classification rate from the unsupervised method is only 65%, while our new method has more than 72% accuracy for all gesture types. Our method achieves more robust performance than the unsupervised DTW method.

	1	2	3	4	5	6	7	8
1	94.0	2.5	0	0	0	0	0	3.5
2	21.3	72.3	0	3.2	0	0	0	3.2
3	7.6	0	92.4	0	0	0	0	0
4	7.4	6.9	0	85.7	0	0	0	0
5	0	0	0	0	93.8	6.2	0	0
6	0	0	0	0	0	100	0	0
7	0	0	0	0	0	0	100	0
8	0	0	0	0	0	0	0	100

TABLE I: Confusion matrix for eight gestures from our algorithm (Columns are recognized gestures and rows are actual gestures).

	1	2	3	4	5	6	7	8
1	89.4	2.1	1.1	2.3	0	0	0	5.1
2	6.8	89.1	0	2.1	0	0	0	2.0
3	11.6	0.6	86.8	1.0	0	0	0	0
4	21.6	11.3	0	65.6	0	0	0	1.5
5	0	1.0	0	0	96.1	2.9	0	0
6	0	0.5	0.1	0.1	1.0	98.3	0	0
7	0	0	0	0	0	0	97.6	2.4
8	0.8	0	0	0	0	0	0	99.2

TABLE II: Confusion matrix for eight gestures from [2].

IV. Conclusion

We propose a new recognition algorithm for hand gesture recognition with low bandwidth, low SNR signals. We use a simple training step to perform template selection before performing dynamic time warping. Our experimental results show that this new algorithm achieves more than 92% accuracy for eight gestures without much training by using limited numbers of low quality signals.

References

[1] T. W. Parsons *et al.*, *Voice and speech processing*. McGraw-Hill New York, 1986, vol. 9.

[2] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," in *IEEE International Conference on Pervasive Computing and Communications*, March 2009.

[3] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca, "Accelerometer-based gesture control for a design environment," *Personal and Ubiquitous Computing*, vol. 10, no. 5, pp. 285–299, 2006.

[4] TI CC2540 Mini Development Kit website. http://www.ti.com/tool/cc2540dk-mini.