A New Effective Wearable Hand Gesture Recognition Algorithm With 3-axis Accelerometer

Zhe Ji, Zhi-Yi Li, Peng Li, MaoBo An
Telecom Network Security Division
National Computer network Emergency Response
technical Team Coordination Center of China (CNCERT/CC)
Beijing, 100029, China

***Abstract*—In this paper, a new simple and effective hand gesture recognition algorithm with 3-axis accelerometer sensor is proposed. It is different from much of the previous work that in our algorithm the preprocessed 3-axis waveforms from accelerometer are directly used as the input features and a fast dynamic time warping(FastDTW) is used as the recognizer. Our proposed algorithm is device-independent in that it can be very easily implemented and portable on any other smart wearable device equipped with 3-axis accelerometer. The experimental results show that the proposed algorithm can perform 99% correct rate with 5 training samples for each gesture in our dataset, which consists of 160 test samples from 16 different gestures. Besides, our algorithm can be 8 times faster than standard dynamic time warping(DTW) algorithm without performance decline.**

***Index Terms*—hand gesture recognition, 3-axis accelerometer, two-scaled normalization, fast dynamic time warping**

I. Introduction

With the successful application of smart phones and pads, it is no doubt that the development of smart wearable devices will play the main role in the next smart era. Their application will make our daily life more convenient and comfortable[1][2]. For more and more coming miniature wearable devices, gesture smart interface will become an important way to overcome their size and screen limitations[1]. At present, miniature 3-axis accelerometer can make gesture recognition on smart wearable possible under effective pattern recognition algorithm supporting. By measuring the acceleration changes of device, 3-axis accelerometer can capture more noticeable, finer or large-scale dynamic gestures than static gesture camera image. In our work, we will focus on the hand gesture recognition algorithm research based on the smart wearable device demo with 3-axis accelerometer equipped.

Preprocessing and identifying gestures are two crucial issues considered in all related research. In order to effectively preprocess the raw data collected from 3-axis accelerometer, a set of low pass filter is first used to clean the outliers and to make the waveform of data more smooth. Besides, a two-scale normalization method is also used to deal with above filtered data. To simplify the preprocess, we directly use the above preprocessed time series data as input features for the following classifier.

There are two typical existing algorithms in pattern recognition for classifying time-series feature. One is dynamic time warping(DTW) and the other is hidden Markov model(HMM)[3]. They are both initially successfully used in speech recognition[4] and have successful applications in gesture recogniton[5]. DTW can be well-used to determine similarity between two time series through finding optimal alignment between them. However, DTW has a quadratic time and space complexity that limits its use to only small time series data sets. HMM is a statistical model, which always needs much more examples than DTW algorithms to train model and needs too much times to do testing[3][5]. Considering the above shortcomings of DTW and HMM, we propose to use a new FastDTW algorithm as our classifier, which is actually a modified version of dynamic time warping algorithm. Through finding an approximation alignment of the optimal warp path between two time series, this algorithm can avoid the brute-force dynamic programming approach of the standard DTW algorithm and has a linear time and space complexity theoretically and experimentally as shown in [6]. Our proposed algorithm based gesture recognition system is device-independent. Therefore, It can be easily implemented and transported to any other wearable devices with 3-axis sensor equipped.

The paper is organized as follows. we first introduce the gesture library defined in our work, which includes 16 different typical hand gestures as shown in section II. Section III shows our system framework based on the proposed algorithm and our demo hardware device. In this section, we detail the data preprocessing algorithms and the fast dynamic time warping algorithm. Section IV demonstrates the experimental results. Section V concludes the paper and outlines the areas for future work.

II. Hand Gestures definition

Above all, we briefly introduce the gesture library used in our work. We define a gesture library which includes total 16 kinds of hand gestures in this section as shown in Fig.1.It can be seen that our library is bigger than lots of libraries in other work[7][8]. These gestures include space handwriting digits “2”, “3”, “4”, “5”, “6”, “7”, “8”, several capital letters “M”, “N”, “P” and several operation “left”, “right”, “up”, “down”, “ring”, “hook”.

The gestures in our gesture library are discriminative between each other and simple for convenient use. Following this rule, we also exclude several other digits and characters. For example, the digit “1” is excluded because of its very similar

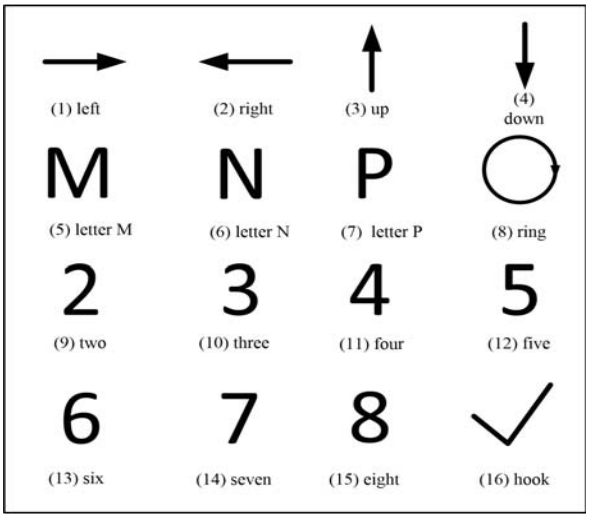

Fig. 1. The total 16 defined gestures in our work.

gesture trace with "down", the letter "Z" is also excluded because of its very similar gesture with "2" etc. Besides, we also exclude the characters with multi-stroke, like "Y", "T etc. Beside these defined gestures, some other characters can also be added in gesture library, such as character "C" etc.

III. Proposed Hand Gesture Recognition System Framework

As shown in Fig.2, the system framework based on our proposed gesture recognition algorithm consists of two main parts, data preprocessing and classification. We will detail the algorithms as following respectively.

In preprocessing modular, a set of low pass filters is used to remove outliers from original sensor data and to make the trace waveforms more smooth. Then, a new two-scale normalization method is proposed to normalize the filtered data from both the amplitude and the length scales. In classification modular, FastDTW algorithm is used as our recognizer engine.

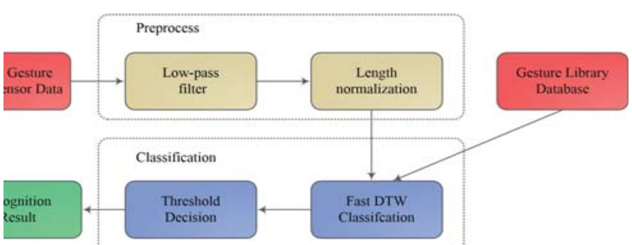

Fig. 2. The gesture recognition system framework based on proposed algorithm.

As shown in Fig.3, we make a wearable demo hardware device. In our demo device, we use MPU6050 device as our accelerometer with a programmable full-scale range of of $\pm 2g$, $\pm 4g$, $\pm 8g$ and $\pm 16g$. In this paper, we choose the $\pm 2g$ range for higher sensitivity. MPU6050 also has three more digital-output X-, Y- and Z-Axis angular rate sensors (gyroscopes) with several user programmable full-scale ranges. For more precision and stable performance, these sensor data can be also compatible in our algorithm.

Our demo device size is 20mm*20mm, so it can be easily used as a modular part in wearable smart band or other smart devices, which also has much more additional functions like time display, date display and other biometric monitoring, etc.

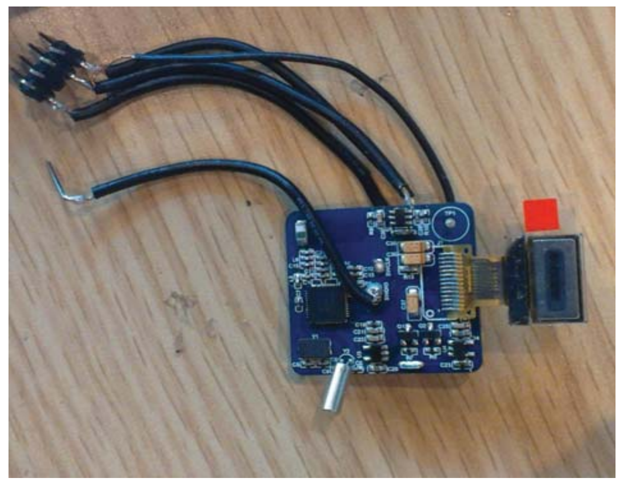

Fig. 3. The demo hand gesture recognition modular for smart wearable device

A. Preprocessing of raw sensor data

Next, we will detail our methods in preprocessing step. Supposing that 1-D sensor data from x-axis is $G_x(i)$ and y-axis and z-axis can be represented as $G_y(i)$ and $G_z(i)$ in same way, we first filter the raw gesture sample data to generate a new sample data $G_x^1(i)$ with median low-pass filter as following.

$$G_x^1(i) = median\{G_x(i-k), \cdots, G_x(i), \cdots, G_x(i+k)\} \quad (1)$$

where the parameter k of filter is set to 5 for all three dimensions. Then, we use the same median low-pass filter to $G_x^1(i)$ to generate another new data sample $G_x^2(i)$ with different parameter $k = 3$. In addition, we make use of Hanning window low-pass filter to $G_x^2(i)$ to generate the final data sample $G_x^3(i)$ as following.

$$G_x^3(i) = G_x^2(i-1)/4 + G_x^2(i)/2 + G_x^2(i+1)/4 \quad (2)$$

For more accurate classification, we also propose a two-scale normalization method following the above filtering process. At first, we subtract the mean value of $G_x^3(i)$ and then normalize

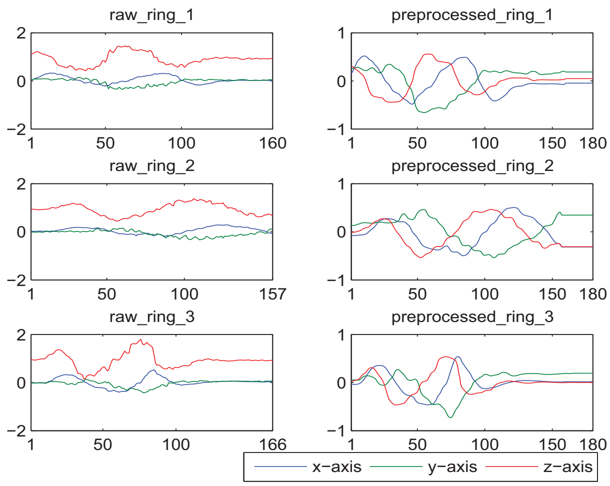

Fig. 5. The 3-axis ring gesture waveform before and after preprocessing block.

its amplitude to generate $G_x^4(i)$. And then, we resample the data samples to a fixed length such as shown in [9][10]. This is because the gesture input duration and movement speeds can vary between users, even for the same intended gesture. In our system, N is experimentally set to be 180 , which is slightly above the average amount 160 of sensor output data. Setting N too low, it will decrease the gesture recognition precision, while choosing N too high will just increase the computation time for following gesture recognition, without a significant gain in accuracy[9][10]. Our resampling method is performed using piecewise linear interpolation, in which the resampled data consists of N equidistant points x_n. The locations of the x_n are built up by successive addition of the points in the original data sample G_x^4. As shown in Fig.4

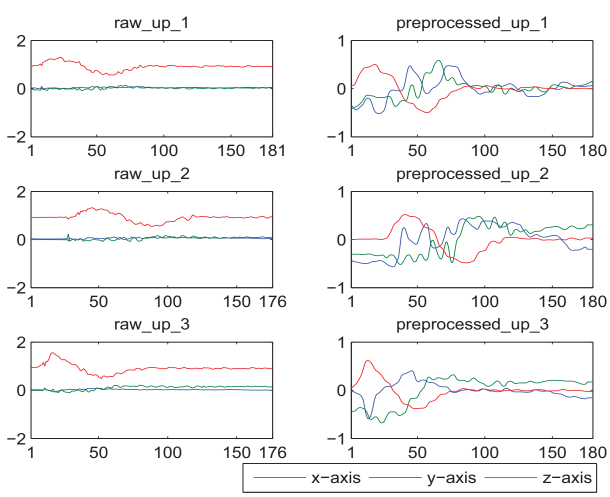

Fig. 4. The 3-axis up gesture waveform before and after preprocessing block.

and Fig.5, we compare three waveforms before and after our preprocessing block of both "up" and "ring" gesture. It can be seen that raw waveforms have different amplitude ranges and wavelengths while the processed waveforms have the same amplitude range and the same wavelength. Through above a series of preprocesses, the waveforms of output accelerometer data become more regular and smooth.

B. FastDTW:fast dynamic time warping

Due to time-series property of hand gesture data and the speed requirements of algorithm in smart wearable devices, we propose to adopt FastDTW algorithms as our classifier, which is a devised version of traditional DTW algorithm. DTW algorithm is able to complete the optimal alignment between two time series. Therefore, it is a good way to evaluate the similarity of two time series from hand gesture sensor and to implement the classification. It doesn't need lots of samples to train a robust model like HMM algorithm. Therefore, we use DTW to overcome the limitation of two identical gesture sample data with slight shift along the time axis.

The standard dynamic time warping problem can be described as follows: Given two time series X and Y with length $|X|$ and $|Y|$,

$$X = x_1, x_2, \cdots, x_i, \cdots, x_{|X|} \quad (3)$$
$$Y = y_1, y_2, \cdots, y_j, \cdots, y_{|Y|} \quad (4)$$

find an optimal warp path W, subject to

$$W = w_1, w_2, \cdots, w_K, max(|X|, |Y|) \leqslant K \leqslant (|X| + |Y|) \quad (5)$$

where K is the length of the warp path and the k^{th} element of the warp path is $w_k = (i, j)$, where i is an index of time series X, and j is an index of time series Y. All the possible warp paths must start at the beginning of each time series at $w_1 = (1, 1)$and finish at the end of both time series at $w_K = (|X|, |Y|)$. This can make sure that every index of both time series will be used in the warp path and there is also a constraint on the warp path which forces i and j to be monotonically increasing in the warp path, every index of each time series must be used, stated more formally as:

$$w_k = (i, j), w_{k+1} = (i, j) i \leqslant i \leqslant i + 1, j \leqslant j \leqslant j + 1 \quad (6)$$

The optimal warp path is the minimum-distance warp path, where its distance is

$$Dist(W) = \sum_{k=1}^{K} Dist(w_{ki}, w_{kj}) \quad (7)$$

A dynamic programming approach can be used to find this minimum-distance warp path. However, traditional dynamic programming approach implement has a quadratic time and space complexity $O(N^2)$ that limits its use to only very small time series data sets in smart wearable device. Based on these considerations, we adopt the speed-up version of DTW algorithm, which is named FastDTW, as our recognizer engine[6].

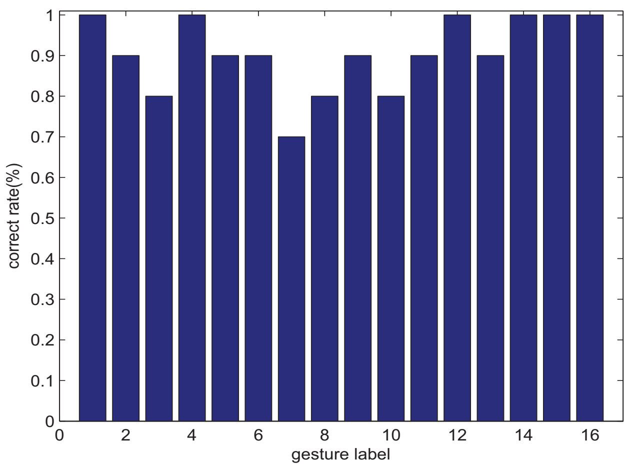

Fig. 6. The correct rate of 16 different gestures with 1 train sample.

FastDTW is a fast dynamic time warp algorithm which tries to find a nearly optimal alignment between two time series in linear time and space complexity. Through using a multilevel approach, time series are initially sampled down to a very low resolution. A warp path is found for the lowest resolution and projected onto an incrementally higher resolution time series. Then the projected warp path is refined and projected recursively again to a higher resolution. The process of refining and projecting is continued until a warp path is found for the full resolution time series[6]. FastDTW algorithm mainly includes three crucial steps: (1) coarsening,(2) projection, (3) refinement.

As described in [6], the coarsening step mainly shrinks a time series into a smaller time series, which represents the same curve as accurately as possible with fewer data points. The projection step mainly finds a minimum-distance warp path at a lower resolution, and uses that warp path as an initial guess for the following higher resolutions minimum-distance warp path. The third refinement step is to refine the warp path projected from a lower resolution through local adjustment of the warp path. It can also be proved that FastDTW algorithm has linear time and space complexity $O(N)$ both theoretically and experimentally [11] and more details of FastDTW can be found in [6] .

IV. Experimental Results

In this paper, all experiments are carried out on self-built dataset obtained by our demo device as shown in Fig.3. Experimental data set contains 16 kinds of gestures as shown in Fig.1. We collect the amount of 5 train samples and 10 test samples for each gesture in that there are totally 80 train samples and total 160 test samples in our data set.

At first, we use only 1 train sample as gesture model and 160 sample sized test set. The recognition results are shown in Table I and we visualize this result as shown in Fig.6.

TABLE I
The correct rate of 16 different gestures with 1 train sample

Gesture label	Correct rate	Gesture label	Correct rate
1	1.00	9	0.90
2	0.90	10	0.80
3	0.80	11	0.90
4	1.00	12	1.00
5	0.90	13	0.90
6	0.90	14	1.00
7	0.70	15	1.00
8	0.80	16	1.00

It can be seen that the average recognition correct rate of the whole test set can achieve as high as 90.6% correct rate while the recognition correct rates range from about 70 % to 100 %. For some "bad" performed gestures, we guess that more training sample would be a good solving method. Therefore, we evaluate the effects of performances with different number of training samples.

In addition, we evaluate the effect of train sample number to performance using both two classifiers: FastDTW and DTW. As shown in Table.II, we evaluate the correct rate by using 1 to 5 samples to train each gesture, respectively. It can be obviously seen that the recognition performances range from 90.6 % to 99.4 % by using FastDTW and range from 91.8 % to 99.4 % by using DTW. We also visualize this result as shown in Fig.7

TABLE II
The performance of FastDTW and DTW algorithm using 1 to 5 train samples

Training samples	FastDTW	DTW
1	0.906	0.918
2	0.938	0.950
3	0.956	0.956
4	0.981	0.981
5	0.994	0.994

In that, we can conclude that the more train samples are used, the better the performance can be obtained. Using 5 samples to train gesture model, our algorithm can achieve a high enough correct rate about 99.4 %.

In addition, we also compare and evaluate the speed of both two algorithms by recording the testing time of the same testing task. Under the condition of using the same hardware device and testing task, FastDTW algorithm only spends 17.2 seconds on testing while traditional DTW algorithm spends more than 137.1 seconds as shown in Table. III.

From the comparison result of running time of two algorithms in Table.III, it can be obviously seen that FastDTW in our system can be about 8 times faster than traditional DTW algorithm. In that, it is well proved that the algorithm we proposed to use is effective.

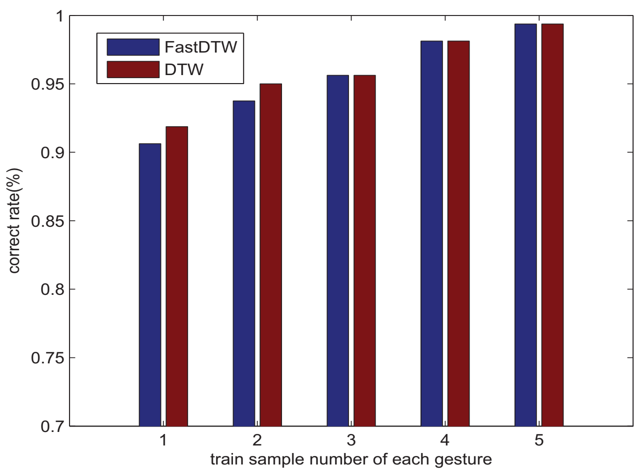

Fig. 7. The performance of FastDTW and DTW algorithm using 1 to 5 train samples

TABLE III
THE RUNNING TIME OF FASTDTW AND DTW ALGORITHMS SPENDING ON THE SAME TESTING TASK AND HARDWARE DEVICES

Algorithm	Running time(second)
DTW	137.1
FastDTW	17.2

V. CONCLUSION AND FUTURE WORK

In our work, we propose an easy and effective hand gesture recognition system for wearable smart devices with 3-axis accelerometer. Experimental results show that our proposed method can perform a correct rate of about 99.4% on the testbench dataset which contains 160 test samples from 16 different gestures. Besides, FastDTW algorithm for gesture classification can perform about 8 times faster than traditional DTW algorithm. In addition, our proposed algorithm is device-independent. Therefore, It can be easily implemented and transported to any other smart wearable devices with 3-axis sensor equipped. In the future, we will mainly focus on how to make automatic valid active gesture detection instead of hard switch.

REFERENCES

[1] Qiang Li Zhiyuan Lu, Xiang Chen and Xu Zhang, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices," *Human-Machine Systems, IEEE Transactions on*, vol. 44, no. 2, pp. 293–299, 2014.

[2] Yong-Ki Son Il-Yeon Cho, John Sunwoo and Cheol-Hoon Lee, "Development of a single 3-axis accelerometer sensor based wearable gesture recognition band," *Springer Berlin Heidelberg*, pp. 43–52, 2007.

[3] LR Rabiner, Ed., *A tutorial on hidden Markov models and selected applications in speech recognition.*, vol. 77. Proceedings of the IEEE, 1989.

[4] Paul R. Cohen Tim Oates, Laura Firoiu, "Clustering time series with hidden markov models and dynamic time warping," *In Proceedings Of The IJCAI99 Workshop On Neural, Symbolic And Reinforcement Learning Methods For Sequence Learning*, 1999.

[5] Timo Pylv, "Accelerometer based gesture recognition using continuous hmms," *Lecture Notes in Computer Science*, pp. 639–646, 2005.

[6] Stan Salvador and Philip Chan, "Toward accurate dynamic time warping in linear time and space.," *INTELLIGENT DATA ANALYSIS*, vol. 11, no. 5, pp. 561–580, 2007.

[7] Shengli Zhou Ruize Xu and Li W.J., "Mems accelerometer based nonspecific-user hand gesture recognition," *Sensors Journal IEEE*, vol. 12, no. 5, pp. 1166–1173, 2012.

[8] CHENG Zixue JING Lei, ZHOU Yinghui and WANG Junbo, "A recognition method for one-stroke finger gestures using a mems 3d accelerometer," *IEICE Transactions on Information and Systems*, , no. 5, pp. 1062–1072, 2011.

[9] Sven Kratz and Michael Rohs, "The $3 recognizer: Simple 3d gesture recognition on mobile devices," in *In proceedings of the 15th International Conference on Intelligent User Interfaces*, New York, NY, USA, 2010, IUI '10, pp. 419–420, ACM.

[10] Jacob O. Wobbrock, "Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes," 2007.

[11] P Chan S Salvador, "Fastdtw: Toward accurate dynamic time warping in linear time and space," *3rd Workshop on Mining Temporal and Sequential Data, ACM KDD'04*, vol. 11, no. 5, pp. 561–580, 2004.