

Automatic Recognition of Finger Gestures Using a Tri-Axis Accelerometer

Yinghui Zhou, Lei Jing, and Zixue Cheng

Abstract—More and more researchers devote themselves to system development of life support for elders such as health condition monitoring, appliances control, emergency call, and so on. Activity recognition using wearable devices is one of the ways of system implementation. In this paper, a system of gesture recognition is designed and implemented based on a tri-axis accelerometer. Six pairs of one-stroke finger gestures are defined according to the sensing characteristic of the accelerometer. Acceleration data was collected from 20 subjects under the supervision by researchers. Gesture information was segmented automatically from successive time signal. Mean, standard deviation, and the relationship of position sequence of wave crest and trough are extracted as features for training and testing of gestures classifying. Several classifiers were selected for gestures recognition. Results show the KNN classifier got the best recognizing result with an average accuracy of 91.58%. The recognition rate is not influenced by which hand being used to perform the gestures.

Index Terms—Wearable computing, gesture recognition, accelerometer, short time energy

I. INTRODUCTION

The world is rapidly ageing which will bring challenges for developed and developing countries. One common trend is more and more elders living alone and less able to benefit from the sharing of goods that might be available in a large household. Investigation from World Health Organization indicates, in Japan, the proportion of people living in 3-generation households has fallen from 46% in 1985 to 20.5% in 2006. Health care both physical and mental becomes an important problem to current society.

Health support system has been studied widely in recent years [1], [2]. Two kinds of main supporting ways focus on speech-based communication and activity-based recognition. The former provides a direct and effective way to know users intension, which has been used in the hospital and household [3], [4]. However voice signal is sensitive to environment sound such as a TV being on so that sometimes difficult to pick out useful speech signal. Even under certain circumstance, user is not able to make a voice to call for a help.

The other kind of system uses the technology of activity recognition to get users information. The Common method of gesture recognition can be divided two types: noncontact method and contact method. Noncontact method like using camera and ultrasonic had become popular method for

activity recognition [5], [6], which provides an application way of no burden to users with high recognition accuracy. However, some people do not like the way of being supervised, and some private spaces are inconvenient to be set camera such as in bathroom.

Contact method means the equipment is attached on the body of users to obtain activity information directly such as RFID tag and sensor. RFID technology is attractive for many applications since it can detect people's activities by simple ID and distance information [7]. But it is difficult to detect motion.

With the development of micro-electrical technology, sensor is used for human activity recognition [8], [9]. It can be attached on any part of human body to detect motion of body part. This paper attempts to recognize gestures of one finger by using a 3-axis accelerometer. One-stroke gestures are defined in 3D space according to the characteristic of accelerometer. Data collection and feature exaction are processed for gestures training and classification.

The paper is arranged as follows. The following section is to introduce how to define the one-stroke gestures in 3D space. The section III introduces data collection based on one accelerometer and signal processing. Gestures information is extracted automatically from successive signal for features computing. The section IV introduces experimental method and gives the evaluation. Conclusion is given in the last section.

II. GESTURE DEFINITION

The purpose of gesture definition is to give support to elders like controlling home appliances or calling for help. Therefore the gestures should be easy to be done for reducing physical load and easy to be understood and learned for reducing conscious load. Combining the way of controlling home appliances and the characteristic of 3-axis accelerometer, 6 pairs of one-stroke gestures are designed. One-stroke means a dynamic non-periodic gesture are executed only once during one period of data collection.

For example, "pushing" a button or "turning" a knob can be regarded as one-stroke gestures. The tri-axis accelerometer can detect the acceleration change of three directions in space as shown in Fig. 1.

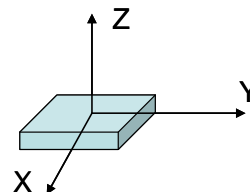


Fig. 1. Sensing direction of the tri-axis accelerometer

Manuscript received July 26, 2012; revised September 27, 2012.

The authors are with the Graduate School of Computer Science and Engineering, University of Aizu, Aizu-Wakamatsu, 965-8580, Japan (e-mail: d8131104@u-aizu.ac.jp, lejing@u-aizu.ac.jp, z-cheng@u-aizu.ac.jp).

The six pairs of gestures can be divided into 3 pairs of gestures on X, Y and Z axis and 3 pairs of gestures on XY, YZ and XZ axial plane. These gestures are named as Crook and Unbend on X axis, Finger L-Shift and Finger R-Shift on Y axis, Finger Up and Finger Down on Z axis, Wrist L-shift and Wrist R-shift on XY plane, L-Rotate and R-Rotate on YZ plane, Wrist Up and Wrist Down on XZ plane. The modes of motion for the six pairs of gestures are shown in Fig. 2. According to the features on different axis and plane, the gestures will be detected and recognized.

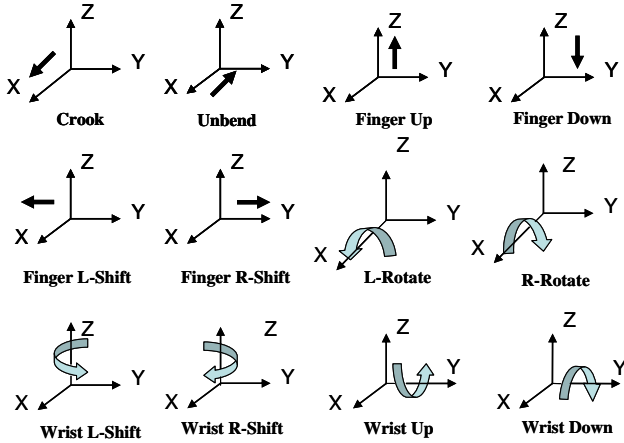


Fig. 2. Modes of the six pairs of gestures

III. DATA COLLECTION AND FEATURES EXTRACTION

A. Data Collection

Acceleration variance is detected by a tri-axis accelerometer MMA7361L from Freescale Semiconductor, Inc. with two sensitivity scale of $\pm 1.5g$ and $\pm 6.0g$. The sensitivity scale of $\pm 1.5g$ is selected since it is sufficient for detecting the predefined gestures.

The accelerometer is embedded in a circuit board which is connected with a MCU board and a power board to form a system like a ring in shape for data collection and signal processing.

The system is attached on the middle phalanx of finger since it is the most flexible in all fingers. Gestures data is collected as sampling 50Hz. The digital signal is stored in a PC for data analysis, features extraction and gestures classification.

B. Extraction of Activity signal

Extraction of activity signal means to separate the data related to gesture from a successive signal, namely to detect start and end point of a gesture. Due to the acceleration changing obviously with a transition from inactivity to activity, the short time energy (STE) is computed to determine the change of state. The STE for a sampling signal is defined as

$$E_n = \sum_{m=-\infty}^{\infty} X^2(m)h(n-m) \quad (1)$$

where n is frame index, $X(m)$ is the signal in time domain and $h(n)$ is a window function. In this paper, a rectangular window function is adopted which is defined as

$$h(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where N is the length of the window in samples.

There are three channels of output for the tri-axis accelerometer as shown in Fig. 3(a). Assuming STE of signal in each axis is $E_{ni}, i = x, y, z$, the total STE is computed.

$$E_{nt} = \sum E_{ni} \quad i = x, y, z \quad (3)$$

To reduce the influence of signal noise, the signal duration of STE greater than a target value is considered.

When total STE E_{nt} is greater than a threshold T_e and duration is long enough, we think a gesture starts to be performed. In this paper, the duration is substituted for the points of $E_{nt} \geq T_e$, which is defined as N_{nt} . So if $E_{nt} \geq T_e$ and N_{nt} is larger than a threshold T_n , it is regarded as gesture starting. At the same way, the condition of gesture ending is $E_{nt} < T_e$ and $N_{nt} \geq T_n$, N_{nt} here represents the points of $E_{nt} < T_e$. Fig. 3(a) shows the acceleration curve of gesture *Finger L-shift* and Fig. 3(b) shows its total STE curve. The horizontal ordinate represents the time sequence, and the vertical ordinate represents amplitude. The amplitude of total STE is reduced ten thousand times scale. The part marked by two dotted lines shows the changes of acceleration and total STE during the gesture being performed.

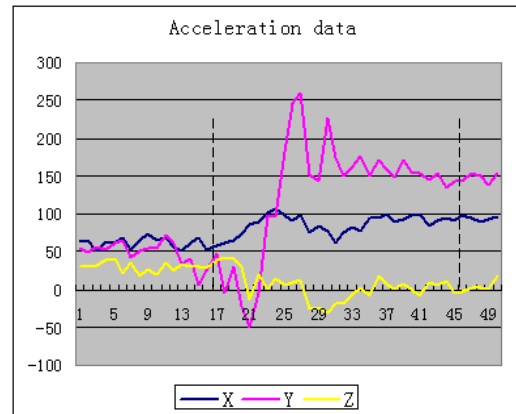


Fig. 3. (a) Acceleration curve of three axis

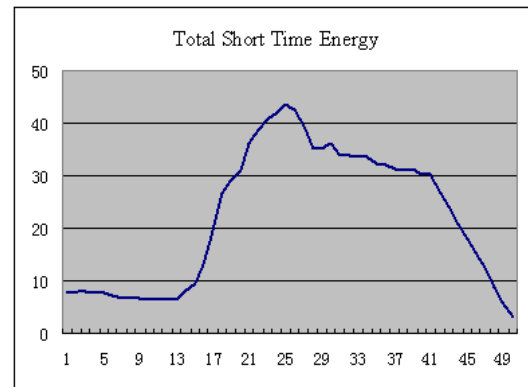


Fig. 3. (b) Total STE curve

The Fig. 3 shows that when a gesture is being performed, the total energy of the part of signal is obviously higher than that of static period. Therefore, the segment of signal is regarded as a meaningful gesture.

C. Feature Selection

It can be seen from the Fig. 3, the acceleration changing in Y axis is more obvious than that in X and Z axis. Therefore the mean and standard deviation (SD) of signal in each axis are selected as features, they are meanX, meanY, meanZ, stdX, stdY, and stdZ. To distinguish the gestures of different direction in same axis, the relationship of position sequence between wave crest and trough in one axis is extracted as XmaxNo>XminNo, YmaxNo>YminNo, ZmaxNo>ZminNo. Therefore, nine features: average values, SD values and position relation of wave crest and trough on three axes consist of the feature vector for gesture recognition.

IV. EXPERIMENT AND EVALUATION

Totally 20 students in the university (16 males and 4 females, average age 25.8 ± 7.8) volunteered for the experiment of data collection under the supervision of researchers. They were required to perform the predefined 12 gestures by forefinger in easy attitude. The gestures start from horizontal state and finished in horizontal state. Each gesture was repeated 5 times per people and 100 times for 20 people totally.

We recorded the duration of each gesture per subject. Results show that it roughly ranges from 200 to 800ms. Because the sampling rate is 50Hz, the window size is compromised with 10 samples. The threshold of energy is 20 and points' threshold is 5.

A. Gesture Recognition

Gestures are recognized based on the algorithm of Naïve Bayes, C4.5 decision tree and KNN performed in the Weka, a kind of machine learning toolkit. All gestures data of 20 subjects are used for training and testing. Training results are analyzed using 10 fold LOOCV (Leave One Out Cross Validation) method. That means the whole dataset are averagely divided into 10 subsets. All subsets are used for training except one for testing. Each subset is chosen once as the testing dataset so that the process will be repeated 10

times. The classifying result is the average of 10 times testing.

For the three algorithms, KNN ($k=3$) shows the highest recognition accuracy with 91.58%. The second most accurate is Naïve Bayes with 88.92% and then is C4.5 decision tree with 87.58%. Table 1 shows the results of classifying the 12 gestures using KNN algorithm.

B. Evaluation

It can be seen from the Table 1, the gesture of Wrist Down is easy to be confused with Crook that error recognizing rate is 13%. The reason is that both of two gestures can reflect acceleration change of X and Z axis, namely they are similar in space. The same situation also happens between Finger R-shift and Wrist R-shift. That shows the similarity of gestures would affect the recognition accuracy directly.

Among the 20 subjects, 15 subjects performed the gestures using right hand, and the rest 5 subjects used left hand. Based on the algorithm of KNN, the recognition accuracy of using right hand is 91.56% which is close to 89.67% of using left hand. That indicates the recognition accuracy is almost unrelated to which hand being used.

V. CONCLUSION

A system of finger gestures recognition based on a tri-axis accelerometer is proposed which can classify six pairs of one-stroke finger gestures. The gestures is predefined in 3D space considering the sensing characteristic of the accelerometer. The system can find gesture information automatically from successive acceleration signal, and then extract feature value for gesture classification.

Due to low physical load and conscious load of finger gestures, the system could be applied for life supporting of elder, for example controlling appliances, calling for a help or communicating with family far away from them.

In the future, system need be improved for features extraction and gestures recognition to increase the accuracy and robustness of the system.

TABLE I: CLASSIFICATION RESULTS OF 12 GESTURES USING KNN ALGORITHM

Classified as	a	b	c	d	e	f	g	h	i	j	k	l	91.58%
a=Crook	86	0	2	9	0	2	0	0	0	0	1	0	86%
b=Unbend	0	100	0	0	0	0	0	0	0	0	0	0	100%
c=Finger Down	1	0	96	2	0	0	0	0	0	1	0	0	96%
d=Wrist Down	13	0	9	77	0	0	0	0	0	0	1	0	77%
e=L-Rotate	0	0	1	0	93	0	4	2	0	0	0	0	93%
f=Finger L-Shift	0	0	0	0	0	97	2	0	1	0	0	0	97%
g=Wrist L-Shift	0	0	2	0	0	4	94	0	0	0	0	0	94%
h=R-Rotate	0	0	0	1	4	0	0	91	1	3	0	0	91%
i=Finger R-Shift	0	0	0	0	0	2	0	0	91	5	2	0	91%
j=Wrist R-Shift	0	0	0	0	0	1	1	0	11	87	0	0	87%
k=Finger Up	0	0	0	0	0	0	0	0	1	0	96	3	96%
l=Wrist Up	0	0	0	0	0	0	0	0	1	0	8	91	91%

ACKNOWLEDGMENT

The authors are grateful to all the volunteers for their participation in the experiment.

REFERENCES

[1] S. Consolvo, P. Roessler, *et al.*, "Technology for care networks of elders," *Pervasive*, vol. 3, no. 2, pp. 22-29, 2004.

[2] V. Stanford, "Using pervasive computing to deliver elder care," *IEEE Pervasive Computing*, 2002, vol. 1, no. 1, pp. 10-13.

[3] K. Jusczy, W. Peter, and L. A. Paul "Speech perception and spoken word recognition: Past and present," *Ear and Hearing*, vol. 23, no. 1, pp. 2-40, 2002.

[4] A. Zafar, J. Overhage, and C. McDonald, "Continuous speech recognition for clinicians," *Journal of American Medical Informatics Association*, vol. 6, no. 3, pp. 195-204, 1999.

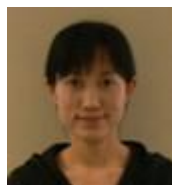
[5] S. Mitra and A. Tinku, "Gesture recognition: A survey," *IEEE Transactions on System, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 37, no. 3, pp. 311-323, 2007.

[6] S. Helal, B. Winkler, *et al.*, "Enabling location-aware pervasive computing applications for the elderly," *First IEEE International Conference on Pervasive Computing and Communications (PerCom'03)*, Fort Worth, Texas, March, 2003

[7] J. R. Smith, K. P. Fishkin, *et al.*, "RFID-based techniques for human-activity detection," *Communication of the ACM*, vol. 48, no. 9, pp. 39-44, 2005.

[8] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive 2004*. Springer: Linz, Vienna. pp. 1-17, 2004.

[9] P. Juha, *et al.*, "Activity classification using realistic data from wearable sensors," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119-128, 2006.



Y. Zhou received her B.E. degree and M.E. degree in Computer Science and Engineering from Jiamusi University and Yanshan University, China in 2001 and 2004 respectively. Now she is pursuing her ph.D. Degree in University of Aizu. Her research is concerned with ubiquitous learning, wearable computing and pattern recognition



L. Jing received his B.Eng. degree from Dalian University of Technology, China, in 2000, M.Eng. degree from the Yanshan University, China, in 2003, and Ph.D from University of Aizu, Japan, in 2008. Currently he is an associate professor at the University of Aizu. His research interests include sensor networks, wearable computing, and ubiquitous learning.



Z. Cheng received his B.Eng. degree from Northeast Institute of Heavy Machinery in 1982, his Master degree and Ph.D degree from Tohoku University, Japan in 1990 and 1993, respectively. Currently, he is a full professor the School of Computer Science and Engineering, the University of Aizu, Japan. His current interests include distributed algorithms, ubiquitous learning, context-aware platforms, and functional safety for embedded systems