A New Feature Fusion Method for Gesture Recognition Based on 3D Accelerometer

Zhenyu He

Computer Center, Jinan University, Guangzhou, 510632, China
E-mail: hzy0753@126.com

Abstract: In this paper, a new feature fusion method for gesture recognition based on single tri-axis accelerometer has been proposed. The process can be explained as follows: firstly, the short-time energy (STE) features are extracted from accelerometer data. Secondly, the hybrid features which combines wavelet packet decomposition with Fast Fourier transform (WPD+FFT) are also extracted. Finally, these two categories features are fused together and the principal component analysis (PCA) is employed to reduce the dimension of the fusion feature. Recognition of the gestures is performed with Support Vector Machine (SVM). The average recognition results of seventeen complex gestures using the proposed fusion feature are 89.89%, which are better than using STE and WPD+FFT. The performance of experimental results show that gesture-based interaction can be used as a novel human computer interaction for consumer electronics and mobile device.

Key Words: Tri-axial accelerometer data, gesture recognition, feature fusion, human computer interaction, short-time energy

1. INTRODUCTION

Gesture recognition is becoming increasing popular as an input way for consumer electronics and mobile device human computer interaction (HCI). However, there are multiple technical challenges to gesture-based interaction [1]. First, unlike many pattern recognition problems, e.g. speech recognition, gesture recognition lacks a standardized or widely accepted ``vocabulary''. Therefore, it is often desirable and necessary for users to create their own gestures, or personalized gestures. For example, the the simple gestures such as Arabic numerals [2-4], simple linear movements and direction [5], tilt detection, shake detection [2-5] are usually study. Secondly, the targeted platforms for personalized gesture recognition are usually highly constrained in cost and system resources, including battery, computing power, and interface hardware, e.g. buttons. As a result, computer vision or ``glove'' based solutions are unsuitable. In recent years the availability of MEMS (Micro-Electromechanical System) tri-axial accelerometer allows for the design of an inexpensive mobile gesture recogniton system. These sensors are a low-cost, low-power solution to recognize gestures and can be used to recond the movements of a person.

In our work, a gesture recognition system based on single tri-axis accelerometer mounted on a cell phone is proposed. We present a novel human computer interaction for cell phone through recognizing seventeen complex gestures. Figure 1 shows an overview of the proposed framework for gesture interactive. When a user performs gestures on 3D space using the mobile phone, the movement is sensed by an accelerometer. Then the acquired data is processed and classified into a gesture through the gesture recognition algorithm. Finally, the corresponding function is executed and feedback to the users.

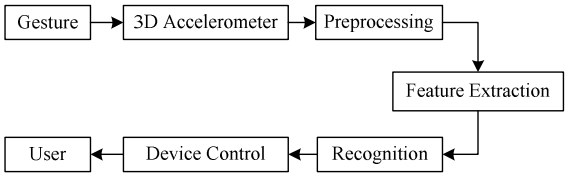


Fig. 1: Framework of Gesture Interactive for Cell Phone

As gesture recognition can be formulated as a typical classification problem and just like many pattern recognition problem, features extraction plays a crucial role during the recognition process. However, few works that extract effective features and make quantitative comparison of their quality are reported. To extract feature from the acceleration data, they convert three dimensional data into one dimensional vector using vector quantization [5]. Some work use acceleration, velocity, position and combination of acceleration with velocity respectively to recognize ten Arabic numerals [4]. Others work extracts the statistics of acceleration data such as local maximal or minimal point as feature [3]. Although gesture recognition using these simple obtain some success, the recognition results using these features can not get a higher accuracy because only using time-domain or frequency-domains features are not enough.

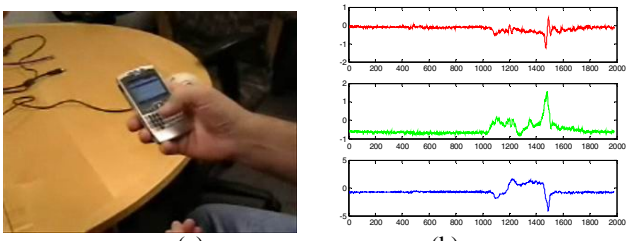
In this paper, a new feature fusion method for gesture recognition based on time-domain and frequency-domain is proposed. First of all, we extract the time-domain features from acceleration data, that is short-time energy. Secondly, we used the method described in [6] to extract the hybrid features which combine Wavelet Packet Decomposition (WPD) with

978-1-4244-7210-9/10/$26.00 ©2010 IEEE

Fast Fourier Transform (FFT). Finally, we fuse these two categories features as the input features of the SVM classifier. The classification of seventeen complex gestures shows encouraging results.

2. Data Collection

As shown in Figure 2(a), a single tri-axis accelerometer is mounted on a cell phone to collect different gestures data. Sixty-seven subjects held the cell phone in hand and performed 17 different gestures in different days. The exact sequence of gestures is listed in Table 1. The output signal of the accelerometer is sampled at 300 Hz. Since acceleration signals are sampled in equal-time interval, the length of raw data is variable according to different gesture and different input speed. Data from the accelerometer has the following attributes: time, acceleration along X-axis, Y-axis and Z-axis. Figure 2 (b) shows the example of raw data.



(a) (b)

Fig. 2: (a) Setup of data collection and (b) example of raw data

Table 1: Gestures labels

Class	Gestures
1	Tilt phone to left & back, then to right & back (>15 deg)
2	Tilt phone towards & then away from you (>15 deg)
3	Slowly tilt phone 90 deg to the left & back, then to right & back
4	Slowly tilt phone 90 deg towards & then away from you
5	Shake phone with no specific direction once
6	Shake phone to the left & back, then to right & back
7	Shake phone towards you & back, then away from you & back
8	Pan phone upward & downward & right & left
9	Tap phone on top left & right, then bottom left & right corner
10	Pick up phone from table, hold to view, & back to table
11	Pick up phone from table & bring it to ear & back to table
12	Bring phone from holding for viewing to ear & back to viewing
13	Take phone off belt clip & hold & put it back
14	Phone in the pocket (no intentional motion)
15	Rotate phone from portrait to landscape & back to portrait
16	Roll phone to left & back, then to right & back
17	Move phone towards, then away from your face

3. Feature Extraction

Feature extraction is the elementary problem in the area of pattern recognition. For gesture recognition task, extraction of effective gesture features is a very important step which will greatly improve the performance of the gesture recognition system. Therefore, we proposed a effective features fusion method from acceleration data in this paper. The block diagram of our proposed feature extraction is shown in Fig. 3 and the details of the methods is presented as follows.

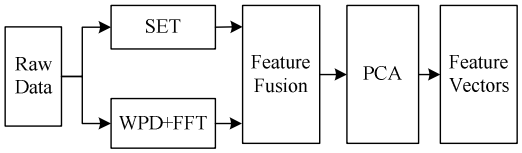


Fig. 3: Block diagram of our feature fusion method

3.1. Short-Time Energy

In general, we can define the short-time energy as [7]

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (1)$$

This expression can be written as

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

Where $$h(n) = w^2(n) \quad (3)$$

Equation (2) can thus be interpreted as the signal $x^2(n)$ is filtered by a linear filter with impulse response $h(n)$. The choice of the impulse response, $h(n)$, or equivalently the window, determines the nature of the short-time energy representation. In this paper, the rectangular window is chosen as impulse response and it is defined as

$$\begin{aligned} h(n) &= 1 \quad 0 \le n \le N-1 \\ &= 0 \quad otherwise \end{aligned} \quad (4)$$

The rectangular window corresponds to applying equal weight to all the samples in the interval $(n-N+1)$ to n. Moreover, the selection of the window length N is a critical problem. That is, we wish to have a short duration window (impulse response) to be responsive to rapid amplitude changes, but a window that is too short will not provide sufficient averaging to produce a smooth energy function. One simple way is to choose the window length N after some comparison. Figure 4 shows the example of short-time energy of accelerometer signal.

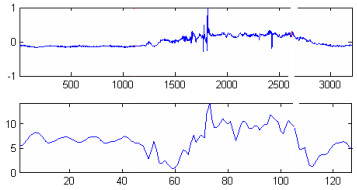


Fig. 4 Y-axis data and its short-time energy

Since the length of data is variable according to the different gesture and different subject's input speed, it is necessary to change the sampling rate of the accelerometer single before Short-Time Energy feature extraction. By combining decimation and interpolation, it is possible to change the sampling rate by a noninteger factor[8]. Specifically, consider Figure 5, which shows an interpolator that decreases the sampling period from *T* to *T/L*, followed by a decimator that increases the sampling period by *M*, producing an output sequence $\tilde{x}_d[n]$ that has an effective sampling period of $T' = TM / L$. By choosing *L* and *M* appropriately, we can approach arbitrarily close to any desired ratio of sampling periods.

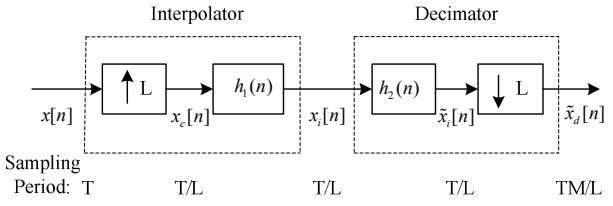


Fig. 5: System for changing the sampling rate by a noninteger factor

3.2. WPD+FFT feature

A hybrid feature extraction methods, namely combine wavelet packet decomposition (WPD) with Fast Fourier transform (FFT) are proposed in [6]. Since the useful information for gesture recognition locates in some specifically signals frequency band, wavelet packet decomposition are used to decompose signals to different frequency band. As discuss in [6], the changes in gestures are characterized mainly by low-frequency of signal. In other words, the low-frequency component, which includes gesture information, can discriminate the different gestures efficiently. Therefore, we decompose the original signals three level using Daubechies wavelet of order 3 and then we obtain wavelet packet coefficients of node s(3,0) which represents the low-frequency of signal. After that, we transform wavelet packet coefficients using FFT and extract the firs 128 FFT magnitude of coefficient as features.

3.3. Feature Fusion

In general, the feature fusion techniques for pattern classification can be subdivided into two basic categories [9]. One is feature selection based, and the other is feature extraction based. In the former, all feature sets are first grouped together and then a suitable method is used for feature selection. In this paper, we adopt the feature selection based method for feature fusion.

Suppose A and B are representation Short-Time Energy and WPD+FFT feature spaces respectively and they defined on pattern sample space Ω. For an arbitrary sample $\xi \in \Omega$, the corresponding two feature vectors are $\alpha \in A$ and $\beta \in B$. The. The serial combined feature of γ is defined by $\gamma \in (\alpha, \beta)^T$. Obviously, if feature vector α is n dimensional and β is m dimensional, then the serial combined feature γ is $n+m$ dimensional. All serial combined feature vectors of pattern samples form a $n+m$ dimensional combined feature space.

In order to reduce the dimension of the feature set, we use Principal component analysis (PCA) [10] to find out important features which are helpful to recognition. PCA is one of the well-known dimension reduction methods. The basic idea of the PCA is to seek a projection that transforms the original features into a lower dimensional space and preserves most information of the original data..

4. Classification

The classification algorithm we used is Multi-class Support Vector Machine. Five-fold cross-validation was used for classifier assessment. The data was randomly divided into five groups with the same number of samples for different classes. The classifier was built five times. Each time one group in turn was excluded from the training and used solely as a test set. The cross-validated classification result is the average of the five testing results.

5. Experiment Results and Discussion

This section describes experiments with the developed activity recognition system.

For the first experiment, we extract Short-Time Energy of accelerometer signal as the input features of the SVM classifier. The preprocessing step consists of two sub steps: removing the gravity components and resampling the signal. The sensed acceleration signal contains not only the gesture movement acceleration but also earth gravity. The gravity amounts are different according to the posture of the sensor in the 3D space. The gravity components are approximately removed by subtracting the mean of accelerations at each time. After that, we resampling the different accelerometer signals as equal length (1024 samples) by combining decimation and interpolation.

According to section 3.1, Short-Time Energy features were extracted from the accelerometer data using a rectangular window length of *N*. For each window, we computed the Short-Time Energy as feature. As discuss above, the selection of the window length *N* is very important. In order to determine the optimal widow length, we test the recognition performance with different window length. Experimental results are summarized in Table 2. It can be seen that the best performance of our system is obtained when *N=64*. We extract the total of 45 dimensions features form three axis acceleration data when *N=64*.

Table 2: Accuracy Versus Different Window Length

N	32	64	128	512
Accuracy	86.22	**86.74**	85.95	84.28

For the second experiment, the fusion feature, that is fusing the short-time energy and WPD+FFT feature are chosen as the input features of the SVM classifier. According to Section 3.3, short-time energy and WPD+FFT feature are firstly combined into one set of vectors and giving a total of 429 dimension features. In order to reduce the dimension of the fusion features, the PCA is employed to extract the most discriminating features for recognition. According to our experiment, the first 30 components of PCA from fusion feature are enough to obtained high recognition accuracy. In order to compare the performance of our fusion features against short-time energy feature and WPD+FFT feature, we carry out experiments under same experimental conditions. In the experiments, we carried out five-cross-validation procedure to validate the effectiveness of the proposed features. The recognition results of short-time energy feature, WPD+FFT feature and fusion features are given in Table 3.

Table 3: Recognition Performance Comparison of Different Features

Class	Accuracy		
	STE	WPD+FFT [6]	Fusion feature
1	83.52	89.56	88.13
2	83.96	92.53	91.10
3	77.80	82.20	83.74
4	84.95	87.91	87.91
5	82.2	71.98	83.74
6	82.09	82.09	87.91
7	82.20	88.24	91.21
8	94.18	85.38	88.35
9	92.75	94.29	95.71
10	92.53	92.64	94.18
11	83.63	85.16	88.24
12	87.91	89.67	90.98
13	85.16	85.05	89.56
14	89.34	81.98	86.48
15	92.64	94.06	94.07
16	94.06	95.49	97.03
17	87.03	88.35	89.89
Average	**86.82**	**87.36**	**89.89**

It can be seen from Table 3 that all three features can recognize the 17 complex gestures based on single tri-accelerometer. Particularly, the proposed fusion feature outperforms the others while the performance of using short-time energy is only slightly lower. The average recognition results for Short-time energy, WPD+FFT and fusion feature are 86.82%, 87.36% and 89.89% respectively. Experimental results show that the fusion feature which combine Short-time energy and WPD+FFT is obviously effective. In fact, the short-time energy is time-domain feature and WPD+FFT is frequency-domain feature. For gesture recognition, time-domain feature and frequency-domain have their own advantages. Thus it is really reasonable to fuse these two categories features to improve the recognition accuracy. Besides, The performance of experimental results also shows that using PCA not only hold the primary information, but also reduce the dimensions of data efficiently.

6. Conclusion

A new feature fusion method for gesture recognition based on a single tri-axis accelerometer mounted on a cell phone has been proposed in this paper. The fusion features combine the short-time energy (STE) with the hybrid features which integrate wavelet packet decomposition and Fourier transform (WPD+FFT). In order to reduce the dimension of the fusion features, the principal component analysis is employed to extract the most discriminating features for recognition. Gesture recognition results are based on acceleration data collect from 67 subjects. The experimental results indicate that the classification accuracy is increased obviously under the proposed fusion feature and demonstrate that the developed fusion feature is more effective than only using the STE or WPD+FFT feature. The encouraging results indicate that gesture recognition based on single tri-axis accelerometer can provides a novel human computer interaction.

REFERENCES

[1] Jiayang Liu, Lin Zhonga, et. al,, uWave: Accelerometer-based personalized gesture recognition and its applications, *Pervasive and Mobile Computing,*Vol 5, Issue 6, pp. 657-675, 2009

[2] Eun-Seok Choi, Won-Chul Bang, et. al, Beatbox Music Phone: Gesture Interactive Cell phone using Tri-axis Accelerometer, *IEEE Int. Conference on Industrial Technology*, 2005.

[3] Sung-Jung Cho, Eunseok Choi, et. al., Two-stage Recognition of Raw Acceleration Signals for 3-D Gesture- Understanding Cell Phones, *10th IWFHR,* La Baule, France, Oct. 2006.

[4] Sung-Do Choi, A.S. Lee, On-Line Handwritten Character Recognition with 3D Accelerometer, *IEEE Int. Conference on Information Acquisition*, pp.845-850,2006.

[5] S. Kallio, J. Kela and J.Mantyjarvi, Online gesture recognition system for mobile interaction, *IEEE Int. Conference on Systems, Man and Cybernetics,* vol 3, pp.2070-2076，2003

[6] Zhenyu He, Lianwen Jin, et. al. Gesture recognition based on 3D accelerometer for cell phones interaction, *IEEE Asia Pacific Conference on Circuits and Systems*, PP.217-220, 2008

[7] L.R.Rabiner, R.W.Schafer, *Digital Processing of speech signals*, Prentice Hall, 1978.

[8] Alan V. Oppenheim, Ronald W.Schafer and John R.Buck, *Discrete-time signal processing(2en ed.)* Prentice Hall, 1999

[9] Jian Yang, Jing-yu Yang, et. al., Feature fusion: parallel strategy vs. serial strateg, *Pattern Recohnition*, vol 3, pp. 1369-1381, 2003

[10] J. Mantyla, J. Himberg, T. Seppanen, Recognizing Human Motion with Multiple Acceleration Sensors, *IEEE Int. Conference on Systems, Man and Cybernetics*, Tucson, USA, vol. 3494, pp. 747-752, 2001..

A New Feature Fusion Method for Gesture Recognition Based on 3D Accelerometer

Zhenyu He

Computer Center, Jinan University, Guangzhou, China, 510632,
E-mail: hzy0753@126.com

Abstract: In this paper, a new feature fusion method for gesture recognition based on single tri-axis accelerometer has been proposed. The process can be explained as follows: firstly, the short-time energy (STE) features are extracted from accelerometer data. Secondly, the hybrid features which combines wavelet packet decomposition with Fast Fourier transform (WPD+FFT) are also extracted. Finally, these two categories features are fused together and the principal component analysis (PCA) is employed to reduce the dimension of the fusion feature. Recognition of the gestures is performed with Support Vector Machine (SVM). The average recognition results of seventeen complex gestures using the proposed fusion feature are 89.89%, which are better than using STE and WPD+FFT. The performance of experimental results show that gesture-based interaction can be used as a novel human computer interaction for consumer electronics and mobile device.

Key Words: Tri-axial accelerometer data, gesture recognition, feature fusion, human computer interaction, short-time energy