Accelerometer Based Gesture Recognition Using Fusion Features and SVM

Zhenyu He
Computer Center, Jinan University, Guangzhou, China
Email: hzy0753@126.com

Abstract—**In this paper, 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. 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 extract the hybrid features which combine Wavelet Packet Decomposition with Fast Fourier Transform. Finally, we fuse these two categories features together and employ the principal component analysis to reduce dimension of fusion features. The Classifier we used is Multi-class Support Vector Machine. The average recognition results of seventeen complex gestures using the proposed fusion feature are 89.89%, which better than previous works. The performance of experimental results show that gesture-based interaction can be used as a novel human computer interaction for mobile device and consumer electronics.**

Index Terms—**Gesture recognition, Tri-axial accelerometer, fusion features, human computer interaction, short-time energy**

I. Introduction

Context awareness is an emerging application area with the aim of easing human computer interaction (HCI). In the case of a mobile device the HCI can be tedious given the physical size limitation both in the keyboard and screen [10]. If the mobile terminal can aware of the user's current context then it could react in some appropriate manner to suit the user without the need of user interaction.

Since gestures are commonly used in daily life, gesture-based interaction can be one of novel interaction ways that users want. To implement the gesture-based interaction, many different techniques, such as vision-based gesture interaction, touch-based gesture interaction have been utilized [2]. In recent years, a new kind of interaction technology that recognizes users' movement has emerged due to the rapid development of sensor technology. An accelerometer measures the amount of acceleration of a device in motion. analysis of acceleration signals enables a new gesture interaction methods.

Although in the literature there are already exist some approaches of using acceleration signals for gestures recognition, 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. Nowaday, 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.

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 our work, 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 [6] 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 features. Recognition of the gestures is performed with Support Vector

doi:10.4304/jsw.6.6.1042-1049

Machine (SVM). The classification of seventeen complex gestures shows encouraging results.

The rest of this paper is organized as follows. In section II, we introduce the recognition system and data collection. Section III presents the detailed information about the feature extraction, including Short-time energy, WPD+FFT and feature fusion. Section IV introduce classification method and experiment results is given in section V. Finally, conclusions are given in section VI.

II. Recognition System and Data Collection

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.

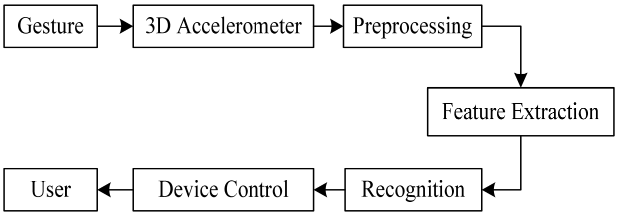

Figure 1. Framework of Gesture Interactive for Cell Phone

As shown in Figure 2, 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 seventeen 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 3 shows the example of raw data.

Figure 2. Setup of data collection

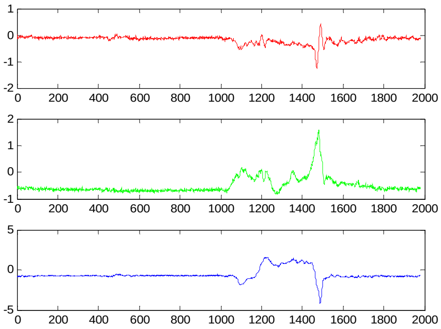

Figure 3. Example of Raw Data

TABLE I.
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

III. 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. 4 and the details of the methods is presented as follows.

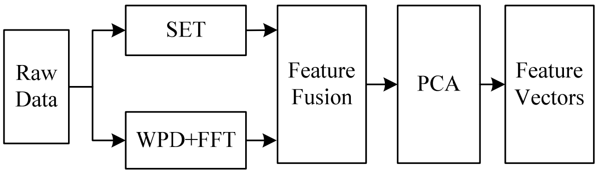

Figure 4. Block diagram of our feature fusion method

A. *Short-Time Energy*

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

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

This expression can be written as

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

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

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

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

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 5 shows the example of short-time energy of accelerometer signal.

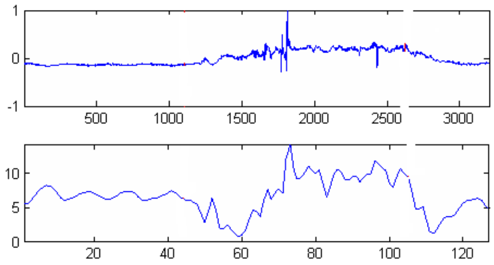

Figure 5. Y-axis data and its short-time energy

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 6, 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.

B. *WPD+FFT feature*

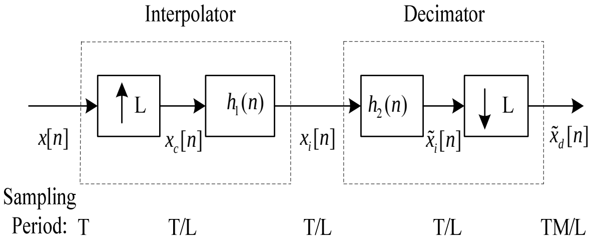

Figure 6. System for changing the sampling rate by a noninteger factor

Fast Fourier transform (FFT), is a typical signal processing approach which can be used to transform the signal from spatial domain to frequency domain. Figure 7 is an example of Y-axis accelerometer data and its FFT coefficients. As shown in Fig. 7, lots of frequency component of our gesture acceleration data are centralized at the low-frequency. Most of the visually significant information is concentrated in just a few FFT coefficients. Therefore, we discard the high-frequency FFT coefficients, and select the low-frequency FFT coefficients as gestures features. In this paper, we extract the first 128 magnitude of FFT coefficients from each axis acceleration data for features. The experiment has shown that using these low-frequency features not only hold the primary information, but also reduce the dimensions of data.

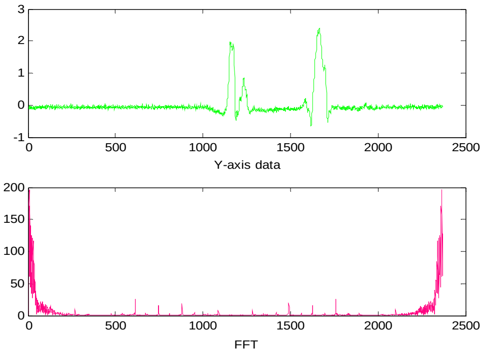

Figure 7. Y-axis data and its FFT

In order to analyze acceleration data more accurately, we employ wavelet packet decomposition (WPD) [11]. It works by generalizing the link between multiresolution approximation and wavelet bases. Compared to wavelet transform, WPD not only decomposes the approximation coefficients, but also the detail coefficients .

A signal space V_j of a multiresolution approximation is decomposed in a lower resolution space V_{j+1} plus a detail space V_{j+1} . The decomposition is achieved by dividing the orthogonal basis $\{\phi(t-2^j n)\}_{n\in Z}$ of V_j into two new orthogonal bases $\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in Z}$ of V_{j+1} and $\{\phi_{j+1}(t-2^{j+1}n)\}_{n\in Z}$ of W_{j+1} , were $\phi(t)$ and $\varphi(t)$ φ(t) are scaling and wavelet functions respectively [12].

The decomposition for WP can be implemented by using a pair of Quadrature Mirror Filter (QMF) bank that divides the frequency band into equal halves. Due to the decomposition of the approximation space (low frequency band) as well as the detail space (high frequency band), the frequency division of the MES signal take place on both the lower and higher sides. This recursive splitting of vector space is represented by admissible WP tree.

A WPD is shown in Fig. 8, where $s(0,0)$ denotes the original signal space, $s(j,k)$ denotes the decomposed subspace, *j* is the decomposition level, and *k* is the index of the subspace occurring at the *j*th level. Therefore, wavelet packet decomposition can decompose the signals to different frequency range ideally. By using WPD, we can obtain decomposed signals that can efficiently represent the features of signal patterns.

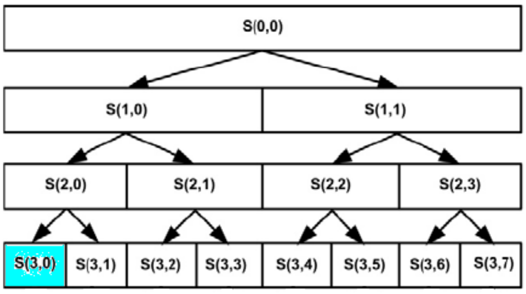

Figure 8. Structure of WPD

In wavelet analysis, different basis functions may be suitable for different signals, and appropriate selection of the wavelet basis for signal representation can result in maximal benefits. One simple way is to choose a basis available after some comparison, although such a result is not optimal. In our experiment, Daubechies 3 wavelet was selected by comparing the decomposition level required while keeping the energy as much as possible.

As discuss above, changes in gestures are characterized mainly by low-frequency of signal. Thus, the low-frequency component, which includes gesture information, can discriminate the different gestures efficiently. Therefore, we decompose the original signals three level using Daubechies 3 and then we obtain wavelet packet coefficients of node s(3,0) which represents the low-frequency of signal. Figure 9 shows an example of Y-axis accelerometer data and its wavelet packet coefficients of node s(3,0). It can be seen that wavelet packet decomposition not only extracts the primary information effectively, but also remove the high-frequency noise and random dithering of signal. After that, we transform wavelet packet coefficients using FFT and extract the firs 128 FFT magnitude of coefficient as features.

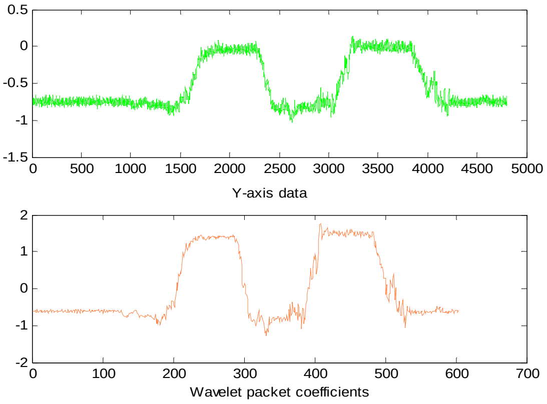

Figure 9. WP coefficients of node (3,0)

C. 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.

D. Dimension Reduction

One approach to coping with the problem of excessive dimensionality is to reduce the dimensionality by linear combining features [13]. In effect, linear methods project the high-dimensional data onto a lower dimensional space, we call it feature compression. One approach-known as Principal Component Analysis or PCA [14][15]. PCA seeks a projection that best represents the original data in a least-squares sense.

Let us consider a set of N sample $\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_N\}$ represented by *t*-dimensional feature vector. The PCA can be used to find a linear transformation mapping the original *t*-dimensional feature space into an *f*-dimensional feature subspace, where normally $f << t$. The new feature vector $y_i \in \Re^f$ are defined by

$$y_i \in W_{pca}^T x_i (i = 1,2,\cdots,N) \tag{5}$$

Where W_{pca} is the the linear transformations matrix, i is the number of sample images.

The columns of W_{pca} are the f eigenvectors associated with the f largest eigenvalues of the scatter matrix S_T , which is defined as

$$S_T = \sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T \tag{6}$$

Where $\mu \in \Re^t$ is the mean image of all samples.The scatter matrix can then be represented by a set of special vectors, which satisfy the following equation:

$$Ce_k = \lambda_k e_k \tag{7}$$

These vectors are called eigenvectors, each eigenvector, e_k , has an associated eigenvalue λ_k . The largest eigenvalues of the correlation matrix, represent the largest inherent variation in the original data set and tell us most about the original data. The most popular way to derive the eigenvalues and vectors is to use Singular Value Decomposition (SVD), although may other techniques are also possible. The technique for computing these values has been outlined in Numerical Recipes [16] and will not be discussed any further here.

Once the principle components have been calculated, the next step is to decide how many of the PCs should be kept, in order to maintain a correct and accurate representation of the original data. One method is to define a threshold value, such that the total number of principle components kept should be greater than this value usually 80-90%. The total proportion of the variance of the original variables, accounted by principle components is given by:

$$v = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{i=1}^{N} \lambda_i} \tag{8}$$

It is normally the case that just the first three or four principle components will be kept, as the first PC is usually very large and the drop off of variance representation is quite steep.

The final step is to project the feature vectors into the new eigenvector space, these projected points can then be used for classification. The feature vectors x_i can be projected into eigenspace by multiplying the feature vectors by the newly calculated eigenvectors:

$$y_i = [e_1, e_2, \cdots, e_{k-1}]^T x_i \tag{9}$$

IV. THE SVM CLASSIFICATION

Automatic classification of different gestures is a challenging work, which requires using a classifier to map the features of the acquired signals to the gestures patterns. Design and implementation of various types of classifiers, ranging from linear methods (Linear Discriminant Analysis or LDA) to nonlinear methods (ANN or SVM, etc.), has been addressed in machine learning theories [17-19], whereas selection of classifier for a given problem is an empirical and experimental work.

Comparisons of different types of classifier have been well studied. It is commonly agreed that linear classifier is more straightforward but poor flexibility, and its performance mainly depends on the nature of the problem and the construction of feature space. Nonlinear classifier is relatively adjustable, and therefore it is eligible for a broader range of problems. The difficulty in using nonlinear method, such as ANN, lies in the determination of their capacity. Improper design may also cause either under fitting or over fitting problems. On the contrary, SVM is a new type of learning machine that can automatically adjust its capacity according to the scale of a specific problem by maximizing the width of the classification margin [17]. One of the benefits is its ability to explore more information from the given data by using a nonlinear function to map the original features into a high-dimensional space as shown in the following description.

Let the training set D be $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, with each input xi $x_i \in \Re^m$ and the output label $y_i \in \{\pm 1\}$. With the nonlinear function ϕ , input vector $\mathbf{x}$ is mapped to $\phi(\mathbf{x})$. The optimal classifier is obtained by solving a quadratic optimization problem [17]:

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\left[\alpha_i \alpha_j y_i y_j \cdot \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)\right] \tag{10}$$

$$\text{with } 0 \le a_i \le C, \qquad i = 1,\cdots,l ,$$

in which C is the regularization parameter that controls the trade-off between model complexity and empirical risk. According the Kuhn–Tucker theorem, samples that have $a_i > 0$ must lie along the margins of the decision boundary, which are called support vectors, To avoid computation of the inner product $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$ in a high-dimensional space, only those functions that can satisfy Mercer's condition, $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$, is considered to be the kernel. With the derived support vectors, the decision function for a new sample $\mathbf{x}$ is expressed as

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{\text{support vectors}} y_i a_i K(\mathbf{x}_i \cdot \mathbf{x}) \right) \tag{11}$$

Typical kernel functions include linear, polynomial and RBF. Although no analytical study exists about the optimal choice of kernel function, RBF is widely used as the kernel function in gait classification studies [18]:

$$K(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp\left(-\frac{\left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2}{2\sigma^2} \right) \tag{12}$$

where σ controls the width of the RBF kernel. Considering the optimal selection of kernel function, Kamruzzaman et al. compared different kernel functions for the diagnosis of cerebral palsy gait and reported that RBF and polynomial kernel function obtained more than 96% overall accuracy [19]. In this study, a RBF-kernel SVM is used to modeling the different gestures based on the signal of accelerometer.

In general, during the implementation of a RBF-based SVM, two parameters, the width of RBF kernel (σ) and the regularization parameter (C) are needed to be determined. However, in our experiment, we use a new fast and parameter-free training algorithm for SVM [20]. The experiment [20] shows that this new algorithm is faster than the Sequential Minimal Optimization(SMO) with Least Recent Use(LRU) cache method. Further, the upper bound of generalized error derived from Statistical Learning Theory(SLT) is found to be better than estimated error by cross-validation as a minimization criterion for free parameters selection. At last, by using ZQP idea and parallel method, the free parameters searching process can be speed up dramatically.

As SVM was originally designed for binary classification, it cannot deal with multi-class classification directly. The multi-class classification problem is usually solved by decomposition of the problem into several two-class problems. In this paper, we used One-versus-One Strategy (OVO), where a set of binary classifiers are constructed using corresponding data from two classes. While testing, we used the voting strategy of “Max-Wins” to produce the output. The SVM algorithm is implemented in Visual C++.

V. Experimental design and results

This section describes experiments with the developed activity recognition system. 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.

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 III, 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 II
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 III, 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 FFT, short-time energy feature, WPD+FFT feature and fusion features are given in Table III.

It can be seen from Table III that all these 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 FFT, Short-time energy, WPD+FFT and fusion feature are 86.92%, 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.

TABLE III
RECOGNITION PERFORMANCE COMPARISON OF DIFFERENT FEATURES

Class	Accuracy			
	FFT [6]	STE	WPD+FFT [6]	Fusion features
1	86.59	83.52	89.56	88.13
2	92.53	83.96	92.53	91.10
3	82.20	77.80	82.20	83.74
4	87.91	84.95	87.91	87.91
5	70.44	82.2	71.98	83.74
6	80.55	82.09	82.09	87.91
7	88.24	82.20	88.24	91.21
8	85.38	94.18	85.38	88.35
9	94.29	92.75	94.29	95.71
10	92.64	92.53	92.64	94.18
11	85.16	83.63	85.16	88.24
12	89.67	87.91	89.67	90.98
13	85.05	85.16	85.05	89.56
14	81.98	89.34	81.98	86.48
15	94.06	92.64	94.06	94.07
16	94.06	94.06	95.49	97.03
17	88.35	87.03	88.35	89.89
Average	86.92	86.82	87.36	89.89

In order to find out which gestures are relatively harder to be recognized, we analyzed the confusion matrices. Table IV shows the aggregate confusion matrix for our feusion features. It can be seen that the third and fifth gesture is hard to recognize. Becase the third gesture often confuse with first gestue while the fifth gesture often confuse with the sixth gesture and the seventh gesture. This result is reasonable, because the raw signals of the fifth gesture are similar to the sixth gesture and the seventh gesture. An example of the fifth and sixth gesture is shown in Fig. 10 and Fig. 11 while the raw signal of the seventh gesture is shown in Fig.3.

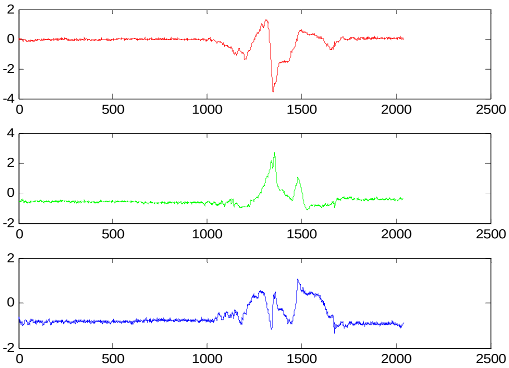

Figure 10. Raw data of the fifth gesture

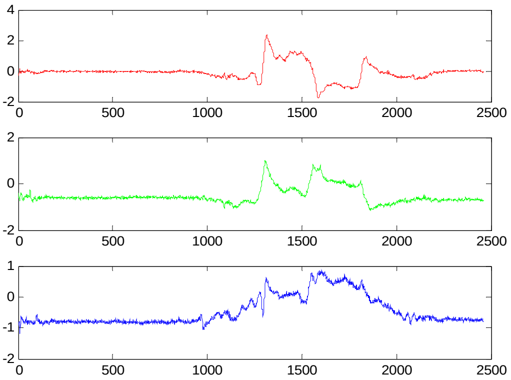

Figure 11. Raw data of the sixth gesture

TABLE IV
THE CONFUSION MATRIX FOR FUSION FEATURES

class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	59		**5**					1	1		1						
2		61		2						1		1	1				1
3	**7**		56	1							1				2		
4	1	6		59										1			
5		1			56	3	**5**			1		1					
6	1		1		**5**	59					1						
7		1		1	2		61	1	1								
8					3			59	1			1	1				2
9						1			64						1		1
10							1		1	63			1	1			
11								2			59	2		3	1		
12								1	1		2	61		2			
13								1		1	2		60	3			
14										1	3	4		58		1	
15			1						1	2					63		
16								1					1			65	
17		2							4			1					60

VI. 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 personalized gesture recognition based on single tri-axis accelerometer can provides a novel human computer interaction.

ACKNOWLEDGMENT

This paper is partially supported by the following fundings: Scientific Research Foundation for the Introduced Talents of Jinan University.

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] Flanagan J.A. and Mantyjarvi J., “Unsuperised clustering of symblo strings and context recogniton”.*ICDM, Maebashi,Janpan*. pp.171-178.

[11] R. R. Coifman, Y. Meyer, and M. V. “Wickerhauser, "Wavelet analysis and signal processing”, in *Wavelets and Their Applications*, M. B. Ruskai, Ed. Boston: Jones and Bartlett, 1992.

[12] L. Deqiang, W. Pedrycz, and N. J. Pizzi, “Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification”, *IEEE Transactions on Biomedical Engineering,* vol. 52, pp.1132-1139, 2005.

[13] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification. Wiley, New York, 2001.

[14] Hong-Bo Deng, Lian-Wen Jin, et. al, “A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA”, *International Journal of Information Technology,* Vol. 11 No. 11, 2005.

[15] Dawson M. R., “Gait Recognition” Final Thesis Report, Department of Computing, Imperial College of Science, Technology & Medicine, London, 2002.

[16] Numerical recipes in C: The art of scientific computing (second edition). William H. Press, Saul A. Teukolsky, William T. Vetterling & Brian P. Flannery， Cambridge university press, 1988-1992.

[17] Hong-Yin Lau, Kai-Yu Tong, Hailong Zhu, “Support vector machine for classification of walking conditions using miniature kinematic sensors”, *Med Biol Eng Comput* vol.46, pp.563–573, 2008.

[18] Begg R, Kamruzzaman J., “A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data”, *J Biomech* 38(3):401–408,2005.

[19] Kamruzzaman J, Begg RK, “Support vector machines and other attern recognition approaches to the diagnosis of cerebral palsy gait”, *IEEE Trans Biomed Eng, vol* 53, pp. 2479–2490, 2006.

[20] Zhi-Jie He, Lian-wen Jin, “A new fast training algorithm for SVM”, *IEEE Int. Conf. on Machine Learning and Cybernetics,* pp. 3451 - 3456, 2008

Zhenyu He received his Ph.D. degree in Communication and Information System from South China University of Technology in 2009. Now he is a teacher at college of information science and technology, Jinan University, Guangzhou, China. His current interests include pattern recognition, machine learning, signal processing and intelligent system.