# A Novel Feature Reduction Method for Real-Time EMG Pattern Recognition System

Peipei Yang, Kexin Xing, Jian Huang, Yongji Wang

Institute of Automation, Chinese Academy of Science, Beijing, 100190, China
College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310032, PRChina
Department of Control Science and Technology, Huazhong University of Science and Technology, Wuhan, China

***Abstract*— This paper proposes a novel feature reduction approach for real-time electromyogram (EMG) pattern recognition. This study extracts time and frequency information by wavelet packet transform (WPT) coefficients and uses the node energy as the feature to overcome the translation-invariant property of WPT. Then the non-parametric discriminant analysis (NDA) is used for feature reduction. Because of some inherent properties of the packet node energy, the within-class scatter matrix is usually singular in this approach, which makes feature project unavailable. To solve this problem, a recursive algorithm is proposed to discard some feature components that lead to singularity and contain relatively less discriminant information. Finally, the support vector machine (SVM) is used as the classifier and gives the recognition result. The corresponding pattern of the action could be recognized in a millisecond (ms). The experimental results show that the proposed method has strong robustness and good real-time performance.**

***Index Terms*—EMG; real-time pattern recognition; wavelet packet; non-parametric weighted feature extraction; SVM**

## 1. INTRODUCTION

The measurement of the intention of muscle activity plays an important role and has been widely applied to many man-machine interface applications [1].

However, the surface EMG signals are often mixed with signals generated by many muscles due to crosstalk and contain much noise [2, 3], which makes identifying the intention of motivation from EMG signals difficult. There are many kinds of features that can be extracted from EMG signals [4]. Among them, the wavelet packet transform (WPT) has become a popular tool to extract features from the EMG signal [5-7]. However, WPT inherits a deficiency from the wavelet mother function that it lacks of a translation-invariant property. To overcome this problem, Gary G. Yen and Kuo-Chung Lin [8] suggested constructing the feature vector with wavelet packet node energy. Therefore, instead of the original coefficient, the wavelet packet node energy is used to construct the feature vector in this paper.

Normally, the dimension of feature vector is very high. In order to improve the performance and reduce the calculation, the dimension of the feature vector should be reduced by discarding the redundant information. Linear projection is a popular feature reduction method due to its flexibility and effectively and contains much famous methods as special cases. Principal Component Analysis (PCA) is a widely-used feature projection method [9]. Englehart et al. [10] extracted WPT coefficients from four-channel EMG signals, and performed dimensionality reduction by using a PCA. Unfortunately, PCA is designed for signal representation rather than classification, and thus could not usually give the optimistic result for pattern recognition. While linear discriminant analysis (LDA) provides us an approach of feature projection in respect of classification, Bor-Chen Kuo and David A. Landgrebe proposed an improved approach of discriminant analysis called nonparametric weighted feature extraction (NWFE) [11] for classification which has a better performance than the traditional ones. To find the best project method, Chu et al. [7] applied respectively LDA, PCA, NLDA (nonlinear discriminant analysis) and SOFM four methods to these features which has been extracted from four-channel EMG signals by WPT. The result shows that LDA has best classification accuracy and real-time performance. However, calculating the within-class scatter matrix $S_W$ is necessary for LDA or NDA method. If there is a high degree of redundancy among features, $S_W$ will be singular, and thus these methods are unavailable. To solve this problem, a combination of PCA and a self-organizing feature map (SOFM) is applied to myoelectric control and it was reported to produce better performance than PCA alone [12]. Zhang et al. implemented PCA and LDA projection to reduce the feature of EMG [13]. These methods depend on the capability of PCA to eliminate the redundancy and noise in EMG data, but they have much higher computational cost. Thus, in this work the FS and FP were united for feature dimension reduction. Furthermore, we proposed a recursive feature selection algorithm, which could discard the components that make the within-class matrix singular and contain relatively less classification information.

The rest of this paper is organized as follows: Section II describes the data acquisition system and illustrates how the EMG signal is converted to time series. In section III, we present the algorithms for EMG pattern recognition, including feature extraction, feature selection, feature projection, the method to avoid a singular within-class matrix, and the classifier. In Section IV, the experiment result shows that our algorithm has good performance to recognize the pattern of EMG. Finally, the conclusion of current work and the possible future improvement are described.

978-1-4673-5534-6/13/$31.00 

## 2. ALGORITHM DESCRIPTION

The EMG signal pattern recognition system is composed of feature extraction, dimensionality reduction and classification three main parts as shown in Fig.1.

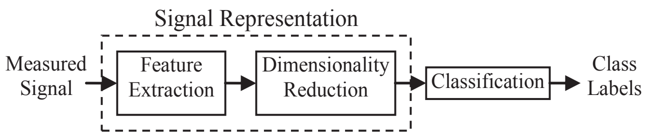


Fig.1. Structure of the Recognition System

### 2.1. Feature Extraction

As previously mentioned, we choose WPT to extract the feature of EMG. In practice, for convenience of processing with computer, discrete wavelet transform (DWT) is widely used and can be implemented using multiresolution analysis introduced by Mallat [14]. Each wavelet packet coefficient is given by

$$w_{j,n,k} = < f, W_{j,k}^{n} > = < f, 2^{j/2} W^{n}(2^{j} t - k) > , \quad (1)$$

where $j$ is a scaling parameter, $k$ is a translation parameter, and $n$ is an oscillation parameter.

Considering the deficiency that lacks of a translation-invariant property, the wavelet packet node energy is adopted to construct the feature vector in our work. Then the wavelet packet node energy is defined as

$$e_{j,n} \equiv \sum_{k} w_{j,n,k}^{2} , \quad (2)$$

which measures the signal energy contained in some specific frequency band indexed by parameters $j$ and $n$.

### 2.2. Feature Selection

To implement a feature selection, the feature components are ranked according to their abilities to discriminate between different classes measured by a criterion function. One simple and widely-used criterion function is Fisher's criterion [15], which is defined as (3) in the case of two classes.

$$J_{f_k}(i,j) = \frac{\left| \mu_{i,f_k} - \mu_{j,f_k} \right|^{2}}{\sigma_{i,f_k}^{2} + \sigma_{j,f_k}^{2}} , \quad (3)$$

where $\mu_{i,f_k}$, $\mu_{j,f_k}$ are the mean values of the $k$-th feature, $f_k$, for class $i$ and $j$, and $\sigma_{i,f_k}^{2}$, $\sigma_{j,f_k}^{2}$ are the variance of the $k$-th feature for class $i$ and $j$ correspondingly. For the case of more than two classes, the criterion function is defined as

$$J_{f_k} = \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} J_{f_k}(i,j) , \quad (4)$$

where $L$ represents the number of classes.

PWM [8] method is a modification of the basic feature selection approach that select a feature subset based on the ability to separate each possible pair of classes instead of the ability to separate all classes. The discriminant power of each feature component is needed in the next step and it can be calculated by implementing a feature selection. The discriminant power is based on the two criteria below.

(1) One feature component that has been selected by more pairs of classes in PWM method could separate more pairs of classes and thus has a better discriminant power.

(2) To a certain pair of classes, the feature component with a better separate ability of this pair has a better discriminant power.

Based on the criteria above, the modified PWM algorithm is described as follows.

**Algorithm 1**:

Supposing the feature vectors are composed of $n$ components and separated into $L$ classes, the input is the set of training samples $\{x_k^{(i)} \in R^n \mid k = 1,2,\cdots,N_i, i = 1,2,\cdots,L\}$, and the outputs are the selected feature component subset $F_{final}$ and the discriminant power of each feature component $P = \{P(f_1), P(f_2), \cdots P(f_n)\}$.

The steps of this algorithm are:

1. Initialize the discriminant power of each component $P = \{0,0,\cdots,0\}$
2. For each possible pair of classes $\{(i,j) \mid i = 1,2,\cdots,L-1, j = i+1, i+2,\cdots L\}$, calculate the discriminant criterion in equation(3) for each feature component $f_k$,
3. For each class pair $(i,j)$, sort $J(i,j)$ such that

$$J_{f_1}(i,j) \geq J_{f_2}(i,j) \geq \cdots \geq J_{f_d}(i,j) \geq \cdots \geq J_{f_n}(i,j) . \quad (5)$$

Determine the feature subset $F_{i,j}$ for each class pair by selecting $d$ feature components that have maximum $J_{f_k}(i,j)$ value

$$F_{i,j} = \{f_k \mid k = 1,2,\cdots,d\}; i = 1,2,\cdots,L-1; j = i+1,\cdots,L . \quad (6)$$

For each class pair $(i,j)$, if feature component $f_r$ is chosen as an element of $F_{i,j}$, it indicates that this component has a better discriminant power for class $i$ and $j$, and the smaller the index $r$ is, the better the discriminant power of this feature component is, so increase the corresponding value by letting

$$P(f_r) = P(f_r) + d + 1 - r . \quad (7)$$

4. Form the final feature set by taking the union of each feature subset

$$F_{final} = \left\{ \bigcup_{i=1}^{L-1} \bigcup_{j=i+1}^{L} F_{i,j} \right\} . \quad (8)$$

### 2.3. Feature Projection

In fact, the subset of the feature components obtained above that could be regarded as a special combination of features. The dimension of the features is still high. The classification property could be further improved by feature projection method and we use NWFE in our system.

A *within-class scatter matrix* [11] showing the scatter of samples around their respective class expected vectors is defined as

$$S_w = \sum_{c=1}^{L} p_c \left\{ \frac{1}{N_c} \sum_{i=1}^{N_c} (x_i^c - m_c)(x_i^c - m_c)^T \right\}, \quad (9)$$

where $x_i^c$ is a sample in class *c*, $N_c$ is the number of samples of class *c*, $p_c$ is the prior probability of class *c*, and $m_c$ is the mean of the samples of class *c* which could be defined as

$$m_c = \frac{1}{N_c} \sum_{i=1}^{N_c} x_i^c . \quad (10)$$

Define the *local mean* as

$$M_j(x_k^{(i)}) = \sum_{l=1}^{N_j} w_{kl}^{(i,j)} x_l^{(j)} , \quad (11)$$

where

$$w_{kl}^{(i,j)} = \frac{dist(x_k^{(i)}, x_l^{(j)})^{-1}}{\sum_{t=1}^{N_j} dist(x_k^{(i)}, x_t^{(j)})^{-1}} . \quad (12)$$

Then the *between-class matrix* [11] showing the scatter of the expected vectors around the mixture mean is defined as

$$S_b^{NW} = \sum_{i=1}^{L} p_i \sum_{\substack{j=1 \\ j \neq i}}^{L} \sum_{k=1}^{N_i} \frac{\lambda_k^{(i,j)}}{N_i} (x_k^{(i)} - M_j(x_k^{(i)}))(x_k^{(i)} - M_j(x_k^{(i)}))^T , \quad (13)$$

where

$$\lambda_k^{(i,j)} = \frac{dist(x_k^{(i)}, M_j(x_k^{(i)}))^{-1}}{\sum_{l=1}^{N_i} dist(x_l^{(i)}, M_j(x_l^{(i)}))^{-1}} . \quad (14)$$

If $\bar{x} = A^T x$ represents a linear transform of the original vector, the *within-class* and *between-class scatter matrices* are $\bar{S}_w = A^T S_w A$ and $\bar{S}_b^{NW} = A^T S_b^{NW} A$ respectively. The goal is to find a proper subspace to make a maximum ratio of $\bar{S}_b^{NW}$ and $\bar{S}_w$. In order to formulate criteria for class separability, the matrices are converted to a number that could reflect the *within-class* and *between-class scatter*. One effective criterion is shown as

$$J_1 = tr(S_w^{-1} S_b^{NW}) , \quad (15)$$

which is chosen to evaluate the reduced subspace. Then the problem could be formulated to find such a projection matrix

$$\bar{A} = \arg\max_A tr(S_w^{-1} S_b^{NW}) . \quad (16)$$

The solution of this problem is given by the *K* eigenvectors of $S_w^{-1} S_b^{NW}$ where *K* is the dimension of the projection space [9]. Once $\bar{A}$ is obtained, the projection of the feature vector in the space with a lower dimension $\bar{A}^T x$ could be computed for each sample.

### 2.4. Avoid the singularity of within-class matrix

An important challenge of implementing feature projection on the feature vector is that the *within-class matrix* might be singular. By a further study, it can be found that the essential reason for the singularity of *within-class matrix* is the property of wavelet packet node energy rather than the size of training set.

**Theorem 1** If the feature vector is composed of the node energy values of all frequency subbands generated by wavelet packet decomposition, the within-class matrix will be singular. Because of space constraints, the proof is discarded.

In order to implement a feature projection properly, the nodes that are linear correlated with others and contain relatively less discriminant information should be discarded before implementing a feature projection. To solve above problem, a recursive feature selection algorithm is proposed in the work. Firstly, a *state tree* that has the same structure as the wavelet packet decomposition tree is created. Each node of this tree could be marked as one of the three states shown below and this algorithm is implemented by updating the state of each node step by step.

(a) *Engaged*: The nodes marked as *engaged* are ones that have been selected and used as components of the feature vector. Therefore, nodes that are linear correlated with them shouldn't be selected in the following steps.

(b) *Available*: The nodes marked as *available* are ones that haven't been selected as a component of the feature vector and are not linear correlated with the nodes that are marked as *engaged* or *unavailable*, so they could be selected as a component of the feature vector at next step.

(c) *Unavailable*: The nodes marked as *unavailable* are not selected as components of the feature vector, but the nodes that are linear correlated with them shouldn't be selected either. These *unavailable* nodes are always linear correlated with some *engaged* nodes, so if one node is correlated with some *unavailable* nodes, it is also correlated with some *engaged* nodes and this make the within-class matrix singular.

At the initial state, all nodes are marked as *available*, indicating that no nodes have been selected at current step. So in order of their discriminant power, every node is tested and tries to be selected as a component of the vector with recursive function *Mark*. The nodes with better classification property are tested priorly to ensure that they have more opportunities to be selected in the feature vector. At the same time, the recursive algorithm also ensures that there're no linear correlated nodes. In this algorithm, evaluations of the nodes with better classification property are prior to others and thus they are more liable to be selected as features. Meanwhile, the algorithm ensures that there are no linear correlated features after selection. Assume that $F_n$ is the feature vector set at step *n*, and then the current state of node *a* is firstly tested. If the state is *available*, the function will mark node *a* as *engaged* and update the set of feature vector by $F_{n+1} = F_n \bigcup \{a\}$. There might be node *b* linear independent with the elements of $F_n$ but correlated with the elements of $F_{n+1}$, and such node should not be selected in the feature vector in the following steps. Therefore, node *b* should be marked as *unavailable*, and this implies that node *b* has actually been selected in set *F* in a linear respect, which might make another node *c* linear independent with elements of

$F_n$ correlate with ones of $F_{n+1}$. So we should then mark node *c* as *unavailable*, and like node *b*, such case will continue until the change of state of one node from *available* to *unavailable* no longer makes the same change happen on any other node.

Such procedure is implemented by a recursive function whose prototype is *Mark(tree, node, value)*. The parameter *tree* is the current state tree, *node* is the node to be marked, and *value* is the state that *node* is to be marked as. This function ensures all "*engaged*" nodes to be mutually linear independent and the algorithm could be described as the pseudo code following. The variable *father*, *brother*, *sonA*, *sonB* represent the father node, brother node, two son nodes of the formal parameter *node* of this function respectively, and *tree*[*node*] is the current state of *node* in the state tree.

**Algorithm 2:**

```
1. tree[node]=value;
2. IF node==root THEN
      GOTO 3;
   ELSE IF tree[father]==available &&
tree[brother]!=available THEN
      Mark(tree, father, unavailable);
   ELSE IF tree[brother]==available &&
tree[father]!=available THEN
      Mark(tree, brother, unavailable);
   END IF
3. IF node==leaf THEN
      GOTO 4;
   ELSE IF tree[sonA]==available && tree[sonB]!=available
THEN
      Mark(tree, sonA, unavailable);
   ELSE IF tree[sonB]==available && tree[sonA]==available
THEN
      Mark(tree, sonB, unavailable);
   END IF
4. RETURN
```

Feature selection is only based on training samples, so the nodes selected to compose the feature vector are invariable when the recognition system works and only the WPT coefficients need to be calculated at the selected nodes, which could prevent the redundant work and improve the efficiency.

### 2.5. *Classifier*

At last, the reduced feature vector is sent to a SVM classifier. However, SVM is originally designed for binary classification while our goal is to classify six kinds of EMG signals corresponding to different motions. Two types of methods have been proposed to classify samples of more than two classes. One is by combining several binary classifiers, such as "one-against-all" [17], "one-against-one" [18], and directed acyclic graph SVM (DAGSVM) [19], while the other is by considering all data in one optimization formulation. Since it is pointed out that a much larger optimization problem is required for methods solving multiclass SVM in one step, this work only tried the one-against-all and one-against-one multiclass SVM and compared their performance.

## 3. Experiment and Result

### 3.1. *Signal Acquirement*

In this system, the surface EMG signals are amplified by 1000 times and filtered out the noise by an analog filter with a passband of 2~750Hz, and then converted to digital signal by an advantech card and transmitted to PC for recognition.

The digital signals form an EMG series, which is then segmented by a moving window. The recognition system extracts a feature vector from the segmented signals in the window and gives the result every recognition period $T_r$. Then the window moves forward by *s* steps, and we get $T_r = s / f_s$, where $f_s$ is the frequency of sampling. This procedure is shown in Fig.2. The signals in the window are *Data*(0)~*Data*(*w*-1) when *N*=1 and *Data*(s)~*Data*(w+s-1) when *N*=2, where *w* is the width of the window, *Data*(*x*) is the acquired signal at time $x/f_s$, and *N*=*k* corresponds to the *k-th* time of recognition. In our system, $f_s$=1024Hz, *w*=512, *s*=256, and $T_r$=0.25s. In other words, the width of moving window is 0.5s and the increment is 0.25 s.

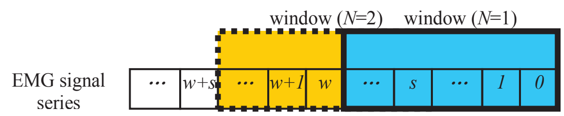

Fig.2. The process of EMG acquisition

In our experiment, four pairs of surface electrodes shown in Fig.3 were used to measure the EMG signal while one is used to provide a reference voltage, and the 7 motion patterns are stillness, wrist extension, wrist flexion, ulnar deviation, radial deviation, hand close, and hand open (mode1~mode7) as shown in Fig.4.

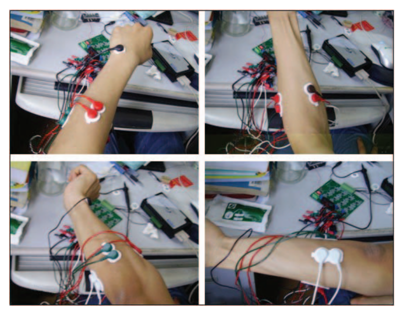
Fig.3. The electrode placement


Fig.4. Seven classes of motion to be recognized

Each 0.5 second we changed one pattern of wrist motion. The EMG signals were collected from 4 subjects (namely S1~S4), and for each subject, 80 samples for each pattern were recorded to train and test the system, so the total number of data to test the recognition system is 80*7*4=2240.

### 3.2. *Wavelet packet node energy and Feature selection*

The signals of each channel were firstly decomposed into a series of coefficients using a wavelet packet transform, of which the depth is 3. Then wavelet packet node energy was calculated for each node and all of them construct the original feature vector. The classification property of each node is very different from others, and feature selection is to select the ones that have better properties.

The first 60 samples collected from subject S1 were used to train the state trees and the results were shown in Fig.5. The *engaged* nodes are used to construct the feature vector.

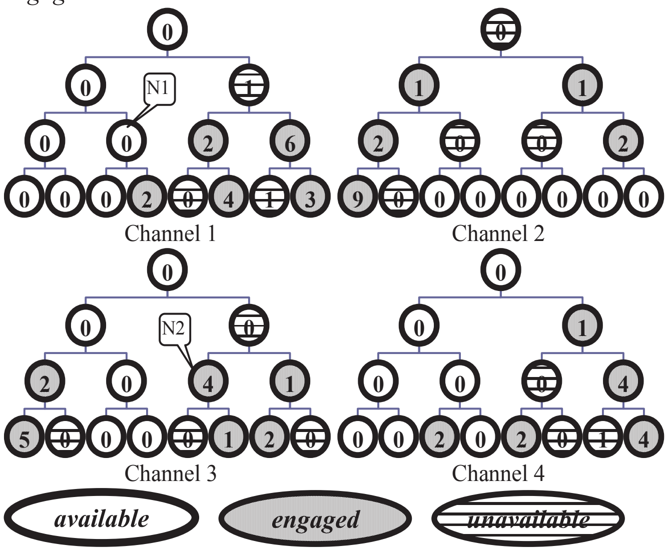


Fig.5. The State Tree of the EMG from 4 channels

To compare the sample distributions of two feature components, the following content takes the node N1 and N2 flagged in Fig.5 as examples to illuminate the feature selection and classification process. The node energies of 40*7 samples of node N1 and N2 are shown in Fig.6. From the result it is obviously the samples are more easily separated in the dimension of N2. Actually, node N2 is selected by the following class pairs {1, 5}, {2, 5}, {3, 5}, and N2 is retained after feature selection while N1 discarded.

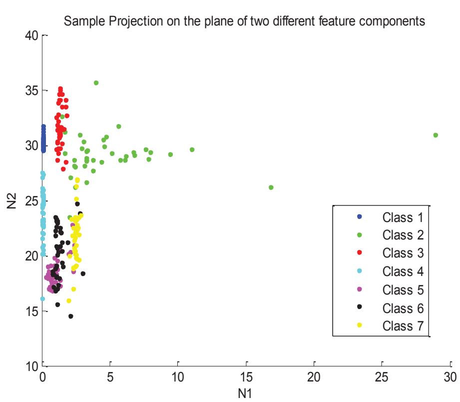


Fig.6. Comparison of node energy of the two nodes

### 3.3. Feature Projection

All values of the node energy of the selected nodes construct a feature vector and it is projected to a space with a lower dimension using NWFE. In this experiment, the original feature vector is projected to a 12-dimension space. The distribution of samples in the projected space is shown Fig.7, in which solid points represent the training samples while the circle ones represent test samples. The points with the same color correspond to the same pattern.

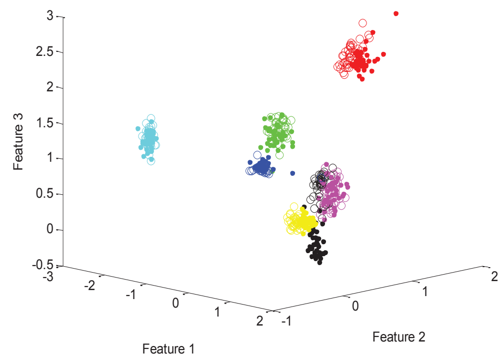


Fig.7. Distribution of the feature vector in the projected space

### 3.4. Classification with SVM and test result

After feature projection, the pattern of EMG signals was recognized by a multiclass support vector machine.

To separate the 7 motion patterns by SVM classifier, one-against-all and one-against-one methods are tried and compared in this experiment. A 4-fold cross-validation [20] was implemented to evaluate the performance of the recognition system. For each pattern, 80 samples were divided to 4 subsets, three of which were used to train the classifier and the other one were used to test. The error rate of each subset of samples can be calculated and shown in Table . Then the 4-fold cross-validation error [20] is regarded as an estimate of the true error rate:

$$e_{CV}^{one-agaisnt-all} = \frac{1}{4}\sum_{j=1}^{4} e_j = 1.964\% \qquad (17)$$

$$e_{CV}^{one-agaisnt-one} = \frac{1}{4}\sum_{j=1}^{4} e_j = 2.143\% \qquad (18)$$

Table 1 Error rate of samples

| Test sample | Subset1 | Subset2 | Subset3 | Subset4 |
|---|---|---|---|---|
| one-against-all | 1.429% | 0 | 0.714% | 5.714% |
| one-against-one | 2.857% | 0 | 0 | 5.714% |

### 3.5. Comparison with different methods of feature reduction

Some researchers use local discriminant basis (LDB) [21] to decide a basis of wavelet packet transform and reduce the dimension of the feature by PCA. This method gives a good result, while it is found that the performance could be further improvement by using our method. Fig.8 shows the distribution of the reduced vector using LDB for wavelet packet base selection and PCA for feature reduction. It is suggested that feature reduction by PCA be applied on each channel respectively, and the reduced feature vector be combined with the reduced vectors of all channels. In our experiment, the 3 principal components from each channel were selected and combined to form a 12-dimension feature vector. Obviously, the distribution shown in Fig.7 is better than Fig.8. The samples from S1 to S4 were identified by different classifiers respectively and the recognition accuracy is shown in the Table

2. From the results, it is shown that the combination of one-against-all SVM and NWFE is the most effective method for identifying the samples. In this experiment, NWFE performs better than PCA, because it considers the classification accuracy as the primary criterion while the aim of PCA is to provide the smallest mean-square error for the signal representation of the given data.

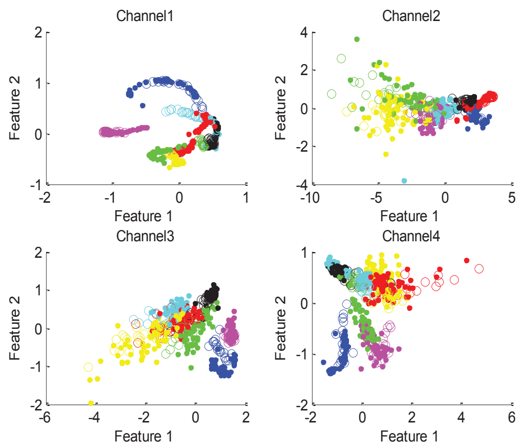


Fig.8. Distribution of the reduced vector using LDB+PCA

Table 2 Recognition accuracy comparation by different classifier

| SVM | One-against-all | | One-against-one | |
|---|---|---|---|---|
| Feature Reduction | NWFE | PCA | NWFE | PCA |
| Subject 1 | 98.036% | 97.500% | 97.857% | 97.500% |
| Subject 2 | 99.822% | 99.643% | 100% | 100% |
| Subject 3 | 99.286% | 98.036% | 99.822% | 98.750% |
| Subject 4 | 98.929% | 98.214% | 98.929% | 98.393% |

## 4. CONCLUSION

This paper has proposed a new real-time EMG pattern recognition method, which combined using WPT, NWFE and SVM. The main contribution of this work focus on developing a novel feature selection method to implement the dimensionality reduction of WPT energy features based on a depth recursive search algorithm. The FS method not only ensures nonsingular of the within-class matrix, but also retains as much information for classification as possible. By the proposed method, 7 motions patterns have been identified by using 4 channels of the EMG. The experimental results show that the whole recognition algorithm has good real-time performance and enough classification accuracy. For future research, the method should be tested with EMG input from amputees. Before real applications the robustness against shifts of the electrodes and the behavior in nonstatic contractions also are needed to be further investigate.

## 5. ACKNOWLEDGMENT

This work has been supported by project of Zhejiang Education Department under Grant Y201121841, the Natural Science Foundation of China under Grants 61202203, and in part by Zhejiang Provincial Natural Science Foundation of China under Grant LY12F01023 and the Fundamental Research Funds for the Central Universities (Huazhong University of Science and Technology) under Grant 2011JC027.

## REFERENCES

[1] M. A. Oskoei and H. Hu, “Myoelectric control systems – A survey,” *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275 – 294, 2007.

[2] D. Farina, M. F. Lucas, and C. Doncarli, “Optimized Wavelets for Blind Separation of Nonstationary Surface Myoelectric Signals”, *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, Jan. 2008.

[3] C. J. De Luca and R. Merletti, “Surface myoelectric signal cross-talk among muscles of the leg,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 69, no. 6, pp. 568–575, 1988.

[4] M. Zecca, S. Micera, M. C. Carrozza, P. Dario, “Control of multifunctional prosthetic hands by processing the electromyographic signal,” *Critical Reviews$^{TM}$ in Biomedical Engineering*, vol. 30, no. 4-6, pp. 459–485, 2002.

[5] G. Wang, Z. Wang, and X. Hu, et al., “Classification of surface EMG signals using harmonic wavelet packet transform,” *Physiological measurement*, vol. 27, no.12, pp.1255-1267, 2006.

[6] Z. Liu and Z. Luo, “Hand motion pattern classifier based on EMG using wavelet packet transform and LVQ neural networks,” in *Proc. IEEE Int. Symp. IT Med. Educ.* (ITME), xiamen, China, Dec. 2008, pp. 28–32.

[7] J.U. Chu, I. Moon, Y.J. Lee, et al., “A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control,” *IEEE/ASME Trans Mechatronics*, vol.12, no.3, pp.282-290, 2007.

[8] G. G. Yen, K.-C. Lin, “Wavelet Packet Feature Extraction for Vibration Monitoring,” *IEEE Transactions on Industrial Electronics*, vol. 47, no. 3, pp. 650-667, 2000.

[9] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification. *New York: Wiley*, 2001.

[10] K. Englehart, B. Hudgins, P. Parker, and M. Stevenson, “Time-frequency representation for classification of the transient myoelectric signal,” in *Proc. 20th IEEE EMBS Annu. Int. Conf.*, Hong Kong, China, 1998, pp. 2627–2630.

[11] B.-C. Kuo, D. A. Landgrebe, “Nonparametric Weighted Feature Extraction for Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096-1105, 2004.

[12] J. U. Chu, I. Moon, and M. S.Mun, “A real-time EMGpattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2232–2239, 2006.

[13] D. Zhang, A. Xiong, X. Zhao and J. Han, “PCA and LDA for EMG-based control of bionic mechanical hand, ” in *Proceeding of the 2012 IEEE International Conference on Information and Automation*, Shenyang, China, June, 2012. pp: 960-965.

[14] S. G. Mallat, “A Theory for Multiresolution Signal Decomposition: The Wavelet Representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, 1989.

[15] K. Fukunaga, “Introduction to Statistical Pattern Recognition,” *New York: Academic*, 1992.

[16] D. B. Percival, A. T. Walden, “Wavelet Methods for Time Series Analysis”, *New York: Cambridge University Press*, 2000.

[17] L. Bottou, C. Cortes, J. Denker, et al. “Comparison of Classifier Methods: A Case Study in Handwritten Digit Recognition”, in *Proc. Of the International Conference on Pattern Recognition*. pp. 77~87, 1994.

[18] S. Knerr, L. Personnaz, G. Dreyfus, “Single-Layer Learning Reviseited: A Stepwise Procedure for Building and Training a Neural Network,” Neurocomputing: Algorithms, Architectures and Applications. J. Fogelman, Ed. *New York: Springer-Verlag*, 1990.

[19] J. Platt, N. Cristianini, J. Shawe-Taylor, “Large Margin DAG’s for Multiclass Classification”. In Advances in Neural Information Processing Systems. *Cambridge, MA: MIT Press*, vol. 12, no. 1, pp. 547-553, 2000.

[20] A. R. Webb. Statistical Pattern Recognition. *Oxford University Press, New York*, 1999.

[21] K. Englehart, B. Hudgins, and P. A. Parker, “A wavelet-based continuous classification scheme for multifunction myoelectric control,” IEEE Trans. Biomed. Eng., vol. 48, no. 3, pp. 302-311, March 2001.