Available online at www.sciencedirect.com
ScienceDirect
Measurement 41 (2008) 618–625
Measurement
www.elsevier.com/locate/measurement

EEG feature extraction based on wavelet packet decomposition for brain computer interface

Wu Ting [a,*], Yan Guo-zheng [a], Yang Bang-hua [a], Sun Hong [b]

[a] School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China
[b] School of Electronics and Information Engineering, Anhui Institute of Architecture and Industry, Hefei 230022, China

Received 12 March 2007; received in revised form 17 July 2007; accepted 31 July 2007
Available online 8 August 2007

Abstract

In the study of brain computer interfaces, a novel method was proposed in this paper for the feature extraction of electroencephalogram (EEG). It was based on wavelet packet decomposition (WPD). The energy of special sub-bands and corresponding coefficients of wavelet packet decomposition were selected as features which have maximal separability according to the Fisher distance criterion. The eigenvector was obtained for classification by combining the effective features from different channels; its performance was evaluated by separability and pattern recognition accuracy using the datasets of BCI 2003 Competition. The classification results have proved the effectiveness of the proposed method. This technology provides another useful way to EEG feature extraction in BCIs.

Keywords: Brain computer interface (BCI); Wavelet packet decomposition (WPD); Feature extraction; Energy of sub-band

1. Introduction

Brain computer interfaces (BCIs) are devices intended to help disabled people to communicate with a computer using the brains' electrical activity. The electrical activity can be measured by electroencephalogram (EEG) [1,2]. BCIs include two kinds, one is based on spontaneous EEG and the other is based on evoked EEG. The evoked EEG is produced by the neural stimulation of inside and outside. For example, P300 and SSVEP (steady-state visual evoked potential) belong to the evoked EEG [3,4]. The spontaneous EEG is produced by human specific thoughts, such as ERD/ERS (event-related desynchronization/event-related synchronization), SCP (slow cortical potention) and some EEG rhythm waves (α rhythm, β rhythm, γ rhythm) [5–7] . Most BCIs make use of spontaneous mental activities (e.g., imagining moving a finger, the hand, or the whole arm, etc.) to produce distinguishable electroencephalogram (EEG) signals [8,9]. The distinguishable EEG signals are then transformed into external actions. Over the past years a variety of evidences have evaluated the possibility to recognize a few mental tasks from EEG signals [10–12]. However, how to improve the recognition performance of EEG signals in signal processing is still a key problem. The recognition procedure mainly includes the feature extraction and the

* Corresponding author.
E-mail address: wu_ting@sjtu.edu.cn (W. Ting).

0263-2241/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2007.07.007

classification, in which the feature extraction plays an important role for the classification. This paper mainly focuses on the feature extraction.

At present, feature extraction methods for the motor imagery EEG mainly include the following methods: (1) *Fast Fourier transform* (FFT): In [13] and [14], the Fourier spectral features were computed with the Welch method using windowed Fourier transforms of signal segments. The main disadvantage of this method is that it only uses the frequency information and doesn't use time domain information. However, the research shows that the combination of frequency information and time domain information can improve the classification performance of EEG signal [15]. (2) *Autoregressive* (AR) *model*: From the AR spectrum, band power is calculated in several frequency bands and the power sum is used as independent variables [16]. In addition, the AR model coefficients or multivariate autoregressive (MVAR) model coefficients are used as features [17,18]. (3) *Time–frequency* (TF) *analysis*: Wang et al. use the TF analysis as a useful tool for oscillatory EEG components during motor imagery [19]. As we all know, oscillatory EEG components produced during motor imagery are both time and frequency related, therefore, the method obtained promising results. However, oscillatory EEG components may cause simultaneous shifts in slow cortical potentials. A combination of two correlated signals might be used to increase extracted information. The TF method only considers oscillatory EEG components. (4) Utilizing coefficients of wavelet transform, i.e., extracting coefficients of wavelet transform at the useful frequency band according to transcendent information [20]. However, the production mechanism of EEG is rather complex, thus, it is difficult to get accurate transcendent information and it is rather inflexible.

Due to the non-stationary property of EEG signals, traditional analysis methods such as the Fourier transform are not very suitable for this work. This paper discusses a feature extraction method based on the wavelet packet decomposition. It used the coefficients mean of wavelet transform (information in time-domain) and power at special subsets as the initial features whose separabilities were measured by the Fisher criteria, in which the features that had a higher separability were considered effective and were formed the final feature vector. This approach accorded with the result that the energies of EEG frequency range are different during subjects' having different imaging tasks, at the same time, some statistical information in time-domain had some changes [21]. The energy of special subbands and corresponding coefficients of wavelet packet decomposition were selected as features which have maximal separability according to the Fisher distance criterion. The performance and effectiveness of this method have been proved by classification results using the datasets of BCI 2003 competition.

2. Wavelet packet decomposition

Wavelet packet decomposition (WPD) is extended from the wavelet decomposition (WD). It includes multiple bases and different basis will result in different classification performance and cover the shortage of fixed time–frequency decomposition in DWT [22]. The wavelet decomposition splits the original signal into two subspaces, V and W, which are orthonormally? Complementary to each other, with V being the space that includes the low frequency information about the original signal and W includes the high frequency information. As shown in Fig. 1, the decomposition of the low frequency subspace V was repeated. WD only partitions the frequency axis finely toward low frequency, and WPD is a generalized version, which also decomposes the high frequency bands that are kept intact in wavelet decomposition. WPD leads to a complete wavelet packet tree, which is shown in Fig. 2, where $U_{j,n}$ is the nth (n is the frequency factor, $n = 0, 1, 2, \ldots, 2j - 1$) subspace of wavelet packet at the jth scale, and $U^n_{j,k}(t)$ is its corresponding orthonormal basis, where $U^n_{j,k}(t) = 2^{-j/2}u^n\ (2^{-j}t - k)$ (k is the shift factor) , it satisfies with (1) and (2)

$$u^n_{j,0}(t) = \sum_k h_0(k)u^i_{j-1,k} \quad (n \text{ is even}) \tag{1}$$

$$u^n_{j,0}(t) = \sum_k h_1(k)u^j_{j-1,k} \quad (n \text{ is odd}) \tag{2}$$

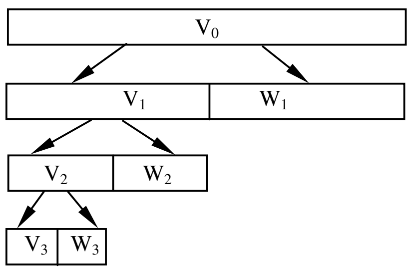

Fig. 1. The structures of wavelet decomposition.

<table>
<tr><td colspan="8">$U_0^0(V_0)$</td></tr>
<tr><td colspan="4">$U_1^0(V_1)$</td><td colspan="4">$U_1^1(W_1)$</td></tr>
<tr><td colspan="2">$U_2^0(V_2)$</td><td colspan="2">$U_2^1(W_2)$</td><td colspan="2">U_2^2</td><td colspan="2">U_2^3</td></tr>
<tr><td>$U_3^0(V_3)$</td><td>$U_3^1(W_3)$</td><td>U_3^2</td><td>U_3^3</td><td>U_3^4</td><td>U_3^5</td><td>U_3^6</td><td>U_3^7</td></tr>
<tr><td colspan="8">… …</td></tr>
</table>

Fig. 2. The structures of WPD.

where $j, k \in Z$, $n = 0, 1, 2, \ldots, 2^j - 1$, $h_0(k), h_1(k)$ is a couple of quadruple mirror filters (QMF) which is irrelevant to scales and satisfies with (3)

$$h_1(k) = (-1)^{1-k} h_0(1-k) \tag{3}$$

When the scale is just enough, sampling sequence of $f(t)(f(k\,\Delta\,t))$ is directly used as the coefficient of $U_0^0(d_0^0(k))$ in approximation. The coefficient of WPD at jth level and kth sample can be shown as (4) and (5) by the quadruple wavelet packet transformation

$$d_j^n(k) = \sum_m h_0(m-2k) d_{j-1}^{n/2}(m) \quad (n \text{ is even}) \tag{4}$$

$$d_j^n(k) = \sum_m h_1(m-2k) d_{j-1}^{(i-1)/2}(m) \quad (n \text{ is odd}) \tag{5}$$

The decomposition coefficient of jth level can be obtained by the$(j-1)$th level, finally we can get the coefficients of all levels through sequential analogy. After it is decomposed by j levels, the frequency ranges of all subspaces at the jth level (U_j^n) are $\left\{\left[0, \frac{f_s}{2^{j+1}}\right]; \left[\frac{f_s}{2^{j+1}}, \frac{2f_s}{2^{j+1}}\right]; \left[\frac{2f_s}{2^{j+1}}, \frac{3f_s}{2^{j+1}}\right]; \ldots; \left[\frac{(2^j-1)f_s}{2^{j+1}}, \frac{f_s}{2}\right]\right\}$, where f_s is the sampling frequency.

3. Formation of initial features

There are four kinds of feature representations in WPD [20–22],

(1) *Part decomposition coefficients*: They have a clear theoretical sense and clearly described the character of EEG signal, it is the fundamental of feature representation.
(2) *Statistical information of wavelet coefficients*: such as sub-band average, the number of zero crossing points, and etc. the computation is simple and the dimension of feature space is low.
(3) *Sub-band energies*: It can reduce the dimension of feature space, but not provide the information in time-domain.
(4) *Transformation modes of coefficients*: For example, the coefficients matrix can be processed by linear or nonlinear transformation, such as PCA (Principal components analysis), SVD (Singular value decomposition), ICA (Independent component analysis), projecting track etc.

At present, the feature extraction based on wavelet transform for spontaneous EEG directly withdraws the coefficients at the interested frequency-bands according to the prior knowledge. However, the production mechanism of EEG is very complicated and the accuracy of prior knowledge can not be acquired easily. Since the combining information of time domain and frequency domain can provide more completed features, and these features need to be as simple and as clear as possible, then we combine part coefficients average and sub-band energies as the optimal feature vector, and the selection rule is based on the Fisher distance criterion.

3.1. Average coefficients

The whole length of decomposition coefficients based on any wavelet packet basis is equal to the one of original discrete sequences, however, all components will have reconfiguration program, and the new sequence can centralize coefficients so that it is easy to extract essential features.

There are l EEG channels $(l = 1, 2, \ldots, i, \ldots C)$, the sampling of each channel is 2^N, and EEG frequency range is $0 \sim f_s/2$. Due to the frequency of useful EEG is lower than 50 Hz, it selected the sub-band means $(\mathrm{MEA}_{j,n})$ at jth level whose frequency range is 0–50 Hz as initial features

$$\mathrm{MEA}_{j,n} = \frac{2^N}{2^j} \sum_k d_j^n(k) \tag{6}$$

EEG signal in each channel is calculated according to (6), then the feature vector formed by all channels can be shown as $M = \{\mathrm{MEA}_{j,0}^1, \mathrm{MEA}_{j,1}^1, \ldots, \mathrm{MEA}_{j,0}^i, \ldots, \mathrm{MEA}_{j,0}^c, \mathrm{MEA}_{j,1}^c, \ldots\}$, simply written as $M = \{m1, m2, m3, \ldots\}$. In principle, the larger the decomposition level, the higher frequency resolution, but correspondingly the computation complexity also increased and the dimension of feature space is heightened, so the decomposition level (j) should be selected according to sampling frequency and reality.

3.2. Sub-band energy

From an energy point of view, WPD decomposes signal energy on different time–frequency plain, and the integration of square amplitude of WPD is proportional to signal power. Like the selection rule of sub-band mean, it selected the sub-band energy ($E_{j,n}$) at jth level whose frequency range is 0–50 Hz as initial features

$$E_{j,n} = \sum_k (d_j^n(k))^2 \tag{7}$$

EEG signal in each channel is calculated according to (7), then the feature vector formed by all channels can be shown as $N = \{E^1_{j,0}, E^1_{j,1}, \ldots, E^i_{j,0}, \ldots, E^c_{j,0}, E^c_{j,1} \ldots\}$, simply written as $N = \{n1, n2, n3, \ldots\}$.

4. Formation and selection of feature vector

In order to decrease the dimension of feature space, FFC (the Filter based on Fisher criterion) is independent of the classification algorithm. The Fisher criterion function,$F(\omega)$, can be written as

$$F(w) = (w^T S_B w)/(w^T S_I w) \tag{8}$$

Where S_B is the between-class scatter matrix, S_I is the within-class scatter matrix, and w is Fisher weight vector which can be obtained by maximizing the value of F. F can be used as a means of assessing the separability of two classes of data. The higher the value of F, the more separable the data is. There are two strategies based on the criterion. (1) *Strategy 1*: It evaluates each feature individually based on the value of F and then selects the d features with the highest values. (2) *Strategy 2*: It is called the forward sequential method that starts by choosing the best individual feature. Then the feature subset is built from the ground up, by repeatedly adding the next feature that works best with the previously chosen features.

The Fisher distance criterion (J) was used for measuring separability of features

$$J = \mathrm{tr}(S_w^{-1} S_b) \tag{9}$$

where S_b is the dispersion matrix between classes, S_w is the one in classes, when J is larger, the classification separability of this feature is better [23], then calculated the value of each component in $M(J_{m1}, J_{m2}, J_{m3,\ldots})$ and arranged them as $J^*_{m1} > J^*_{m2} > J^*_{m3}$, selected the front d coefficient means as features which is corresponding to $J^*_{m1} \sim J^*_{md}$, the mean features can be written as $M' = \{m'_1, m'_2, m'_3, \ldots m'_d\}$. The front $l$features from each component in$N$were selected according to the same Fisher criterion, finally obtained the energy features which were written as $N' = \{n'_1, n'_2, n'_3, \ldots n'_l\}$.

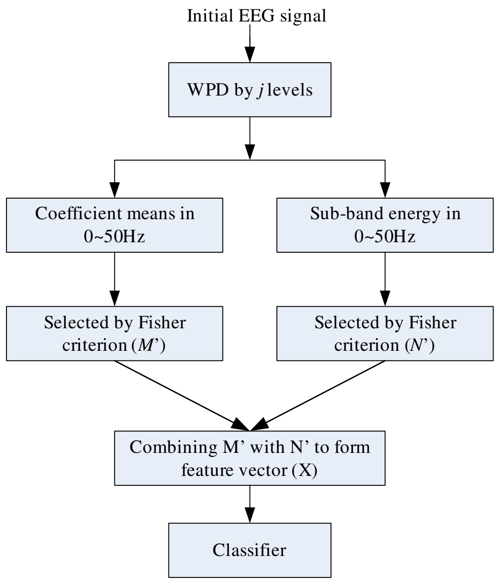

Fig. 3. Formation procedure of feature vector.

The coefficient mean and sub-band energy of wavelet packet represented different information, they can be used as feature vector, respectively or assembly. So it combined the two kinds of information as the final feature vector for classification which was written as $X = \{M', N'\}$. The formation procedure of feature vector is shown as Fig. 3.

5. Results of experiment

5.1. Datasets

The dataset is downloaded from the website of "BCI Competition 2003", the description on web includes experiment introduction, the structure and format of datasets. The goal of the "BCI Competition 2003" is to validate signal processing and classification methods for BCIs. For each dataset specific goals are given in the respective description. Technically speaking, each dataset consists of single-trials of spontaneous EEG activity, one part labeled (training data) and another part unlabeled (test data), and a performance measure. The goal is to infer labels for the test set from training data that maximizes the performance measure for the true (but to the participant unknown) test labels. We selected the dataset I(a) and dataset I(b), they belong to the self-regulation of SCPs. The following is a description of experiments.

The dataset I(a) were taken from six healthy subjects (three male and three female, 22–35 years old). The subjects were asked to move a cursor up and down (two mental activities) on a computer screen, while his slow cortical potentials (SCPs) were taken. Each trial lasted 6 s and consisted of three phases: a 1-s rest phase, a 1.5-s cue presentation phase, and a 3.5-s feedback phase, as shown in Fig. 4. The cue presentation is a visual target appearing either at the top or bottom. Data were recorded during the 3.5-s feedback at a sampling rate 256 Hz. The feedback is provided by a cursor whose vertical position indicated the current level of SCPs (Cz-Mastoids). Each subject completed 300 trials. The following six channels of EEG data were recorded (denotation follows the 10/20 system), as shown in Fig. 5:

Ch1: A1–Cz (A1 = left mastoid)
Ch2: A2–Cz (A2 = right mastoid)
Ch3: (2 cm anterior of C3)-Cz
Ch4: (2 cm posterior of C3)-Cz
Ch5: (2 cm anterior of C4)-Cz
Ch6: (2 cm posterior of C4)-Cz

The trials were separated into a training set (268 trials) and a test set (293 trials), both of which contained EEG data from only the feedback phase of each trial. The cue labels (class "cue 0" or "cue 1") for the training sets were used to tune the parameters of the classification algorithm, whose performance was subsequently assessed on the test set.

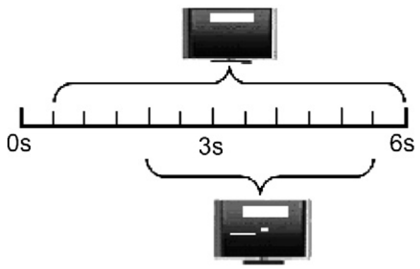

Fig. 4. The experiment paradigm.

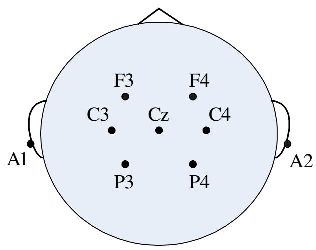

Fig. 5. Map of EEG electrodes for 6 channels.

The dataset I(b) were taken from an artificially respirated ALS patient. Besides the same six channels in dataset I(a), there is another EOG artifact channel to detect vertical eye movements. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Cortical positivity led to a downward movement of the cursor on the screen. Cortical negativity led to an upward movement of the cursor. Each trial lasted 8 s.

During each trial, the task was visually and audibly presented by a highlighted goal at the top (for negativity) or bottom (for positivity) of the screen from 0.5 s until 7.5 s of every trial. In addition, the task ("up" or "down") was vocalized at 0.5 s.

The visual feedback was presented from 2 s to 6.5 s. Only this 4.5 s interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5 s results in 1152 samples per channel for every trial.

These files contain 200 trials which were recorded on the same day and permuted randomly. The trained dataset each contained a data of 100 trials belonging to the corresponding class. The matrix dimensions are 100×8065 and 100×8065.

Every line of a matrix contains the data of one trial. The first column codes the class of the trial (0/1). The remaining columns contain the time samples of the 7 EEG/EOG channels. This starts with 1152 samples from channel 1 and ends with 1152 samples from channel 7.

5.2. Results of classification

Aiming at dataset I(a), Daubechies Wavelet (db4) was selected to decompose EEG signal on six levels. There are 64 sub-bands of wavelet packet at the 6th level whose corresponding frequency ranges are $[0,2]$, $[2,4], \ldots,$ $[126,128]$ Hz, among them 25 sub-bands belong to 0–50 Hz, so the dimensions of feature vector M and N both are 150 (6 channels, 25 sub-bands in each channel).

As shown in Fig. 6, the separability (J) of each component in M can be calculated according to the procedure of feature extraction, it has two peak points, to acquire the feature of two-dimensions, i.e.,$M' = \{m'_1, m'_2\}$. From Fig. 7, it selected the fea-

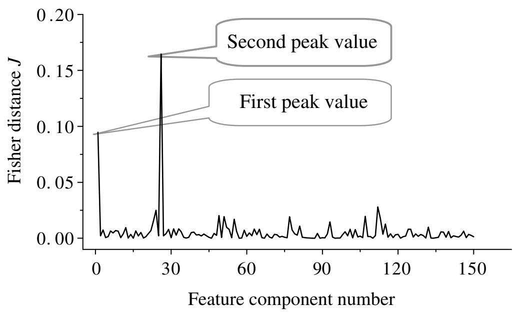

Fig. 6. The separability (J) of each component in M.

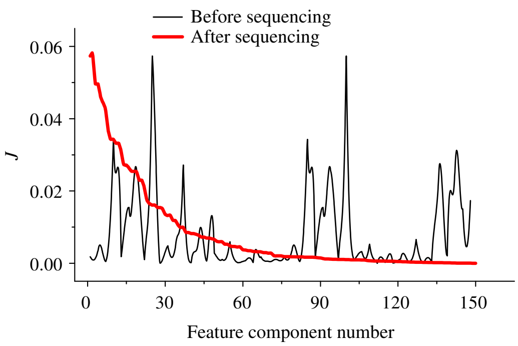

Fig. 7. The separability (J) of each component in N.

ture of 15 dimensions in N, i.e., $N' = \{n'_1, n'_2, n'_3, \ldots n'_{15}\}$, so X is 17 dimensions feature vector, i.e., $X = \{M', N'\}$. The distributions of m'_1 and n'_1 on training data are shown as Figs. 8 and 9, there are some differences in feature components of different classes.

As for the classifier, at present, classification methods for spontaneous EEG mainly include such as linear classifier Kalman filter, ANN (Artificial Neural Network), and etc [24–26]. Linear classifier is simple and easily realized; it needs less computation time and storage capacity. However, since

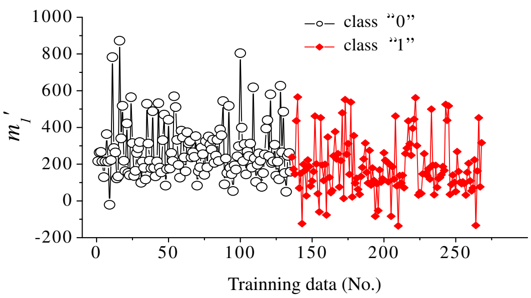

Fig. 8. Distribution of $M1'$ on training data.

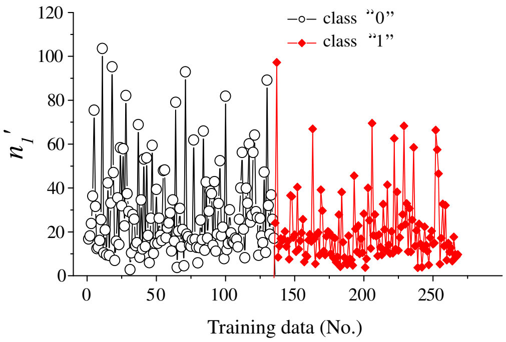

Fig. 9. Distribution of $N1'$ on training data.

EEG is nonlinear, and the feature vector is non-linearly separable, this kind of classifier does not have a high precision of recognition. Kalman filter is an algorithm based on probability theory, it needs transcendent knowledge which is difficult to be obtained since EEG is a very complicated physiological signal. ANN is a comparatively good mechanical learning method. It can solve many non-linear problems and has been successfully applied to a wide variety of engineering problems. In this paper, we use the Probabilistic Neural Network (PNN) as our classifier. By virtue of easy training and a solid statistical foundation in Bayesian estimation theory, the PNN has become an effective tool for solving many classification problems [27] The neural network structure can be illustrated in Fig. 10. Four layers, one input layer, one output layer, one hidden layer and one added layer, are designed. The input layer transfers $\vec{x}$ to the network without any computation. After the hidden layer receiving $\vec{x}$, the output and input of the jth neuron at the ith mode is defined by (1)

$$\phi_{ij}(\vec{x}) = \frac{1}{(2\pi)^{d/2}\sigma^d} \exp\left[-\frac{(\vec{x} - \vec{x_{ij}})(\vec{x} - \vec{x_{ij}})^{\mathrm{T}}}{\sigma^2}\right] \quad (10)$$

There are 268 groups of training samples, 135 groups belong to class "0", 133 groups belong to class "1", and 293 groups of testing samples. The number of neurons in hidden layer of each class is 10, The feature vector X was sent to PNN (probability neural network) for recognition, the accuracy of test data is 90.8%, it increased 2.1% comparing with the best result (88.7%) in BCI 2003 competition.

Aiming at dataset I(b), the steps of signal processing is similar with dataset I(a), the result of classification are shown in Table 1.

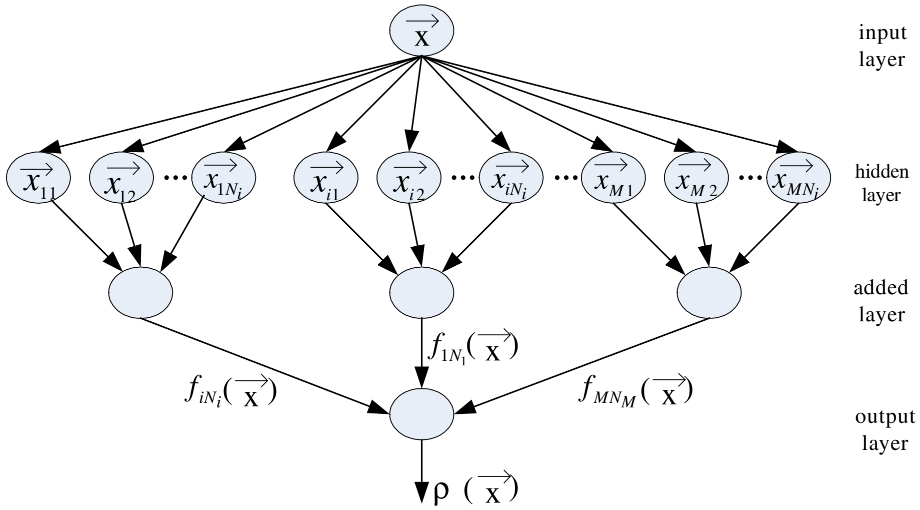

Fig. 10. The structure of probabilistic neural network (PNN).

Table 1
The result of classification

Dataset (SCP)	The best result of competition	Result of the proposed method
Ia	88.7%	90.8%
Ib	54.4%	59.1%

5.3. Discussion

In the present study, we have demonstrated that the WPD is an efficient way for the feature extraction of motor imagery EEG signal. The WPD is an excellent signal analysis tool, especially for non-stationary signals. Due to the non-stationary property of EEG signals, the WPD is very appropriate to analysis EEG signal.

Compared with the WPD, the FFT and the AR is used to analyze stationary signals and thus can not obtain good performance when they are used to analyze non-stationary signals, such as EEG. Our experimental results show that the WPD is superior to the AR model. In addition, the TF analysis is used to analysis EEG signals during motor imagery in [19]. In the TF method, raw EEG signals are filtered by Laplacian method and then the filtered EEG signals are decomposed into some band bins with band-pass filters. The corresponding envelopes of EEG in each band are down-sampled to form features. The TF method obtains a promising result due to the following reasons: 1) Spatial filters are used as a means of accentuating localized activity and reducing diffused activity, which is favorable to the feature extraction. 2) The combination of time domain information and frequency information can provide better classification performance than using any one of them. The WPD is a good time-frequency analysis tool and in the following our research, we will continuously use the WPD as an analysis tool. In addition, we will adopt some advantages of other methods, such as the Laplacian filter technique used in TF method.

6. Conclusion

In this paper, the WPD feature extraction method for classifying EEG signals during the motor imagery tasks is investigated. WPD yields a redundant representation of the signal and its over completed structure provides the flexibility for features representation to achieve better accuracy. We adopted the coefficients mean of wavelet transform (information in time-domain) and power at special subsets (information in frequency-domain) as the initial features whose separabilities were measured by Fisher criteria, the features that had a higher separability were considered effective and were formed the final feature vector. Comparing with other approaches of feature extraction, it supplied more information and put forward a selection rule of Fisher distance criterion, and the performance and effectiveness have been proved by classification results using the datasets of BCI 2003 competition, In addition, the whole program for signal processing has been realized by C++, it is benefit for its application in a real BCI system. So it supplied an effective path for EEG feature extraction in brain computer interface.

References

[1] C. Guger, C. Neuper, D. Walterspacher, T. Strein, G. Pfurtscheller, Rapid proto-typing of an EEG-based brain–computer interface (BCI), IEEE Trans. Neural Syst. Rehab. Eng. 9 (1) (2001) 49–58.

[2] J. Kalcher, D. Flotzinger, C. Neuper, S. Golly, G. Pfurscheller, Graz brain-computer interface II:towards commnunication between humans and computers based on online classification of three different EEG patterns, Med. Biol. Eng. Comput. 34 (1996) 382–388.

[3] J.D. Bayliss, D.H. Ballard, Single trial P3 epoch recognition in a virtual environment, Neurocomputing 32 (33) (2002) 637–642.

[4] M. Middendorf, G. McMillan, G. Calhoun, Brain-computer interfaces based on the steady-state visual-evoked response, IEEE Trans. Rehab. Eng. 8 (2) (2003) 683–689.

[5] T. Hinterberger, J. Mellinger, N. Birhaumer, The thought translation device: structure of a multimodal brain-computer communication system, Presented at Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering, 2003.

[6] G. Dornhege, B. Benjamin, C. Gabriel, Speeding up classification of multi-channel brain-computer spatial patterns for slow cortical potentials, Presented at Proceedings of the 1st international IEEE EMBS Conference on Neural Engineering, 2003.

[7] J.R. Wolpaw, D.J. McFarland, T.M. Vaughan, The wadsworth center brain–computer interface (BCI) research and development program, IEEE Trans. Neural Syst. Rehab. Eng. 11 (2) (2003) 204–207.

[8] T.N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, Support vector channel selection in BCI. IEEE Trans, Neural Syst.Rehab. Eng. 5 (6) (2004) 1003–1010.

[9] G. Pfurtscheller, C. Neuper, A. Schlogl, K. Lugger, Separability of EEG signals re-corded during right and left motor imagery using adaptive autoregressive parameters, IEEE Trans. Rehab. Eng. 6 (3) (1998) 316–325.

[10] J. Millán, Adaptive brain interfaces, Commun. ACM. 46 (2003) 74–80.

[11] J. Millán, J. Mouriño, Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project, IEEE Trans. Neural Syst. Rehab. Eng. 11 (2003) 159–161.

[12] J.R. Wolpaw, D.J. McFarland, G.W. Neat, C.A. Formeris, An EEG-based brain-computer interface for cursor control, Electroenceph. Clin. Neurophysiol. 78 (3) (1991) 252–259.

[13] M. Varsta, J. Heikkonen, J. Millan, Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP committee. in: 15th International Conference on Pattern Recognition, 2002.

[14] P. Mark, K. Aleksandar, Feature Extraction in development of brain-computer interface: a case study. in: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003.

[15] D.M. Brett, W. Justin, H.S. Seung, BCI competition 2003-Data Set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1052–1056.

[16] D.J. McFarland, L.A. Miner, T.M. Vaughan, J.R. Wolpaw, 2000 Mu and beta rhythm topographies during motor imagery and actual movements, Brain Topogr. 12 (2004) 177–186.

[17] D.P. Burke, S.P. Kelly, P.d. Chazal, R.B. Reilly, C.A. Finucane, parametric feature extraction and classification strategy for brain–computer interfacing, IEEE Trans. Neural Syst. Rehab. Eng. 13 (2004) 12–17.

[18] C.W. Anderson, E.A. Stolz, S. Shamsunder, Multivariate autoregressive models for classification of spontaneous electroencephalographic signals, IEEE Trans. Biomed. Eng. 45 (2004) 277–286.

[19] T. Wang, J. Deng, B. He, Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns, Clin. Neurophysiol. 115 (2000) 2744–2753.

[20] M. Fatourechi, S.G. Mason, G.E. Birch, A wavelet-based approach for the extraction of event related potentials from EEG, ICASSP 2004.

[21] A. Kübler, N. Ghanayim, T. Hinterberger, The thought translation device (TTD) for completely paralyzed patients, IEEE Trans. Rehab. Eng. 8 (2) (2000) 190–193.

[22] J.Z. Xue, H. Zhang, C.X. Zheng, Wavelet packet transform for feature extraction of EEG during mental tasks, Presented at Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an, 2003.

[23] D. Mohamed, A.A. Ahmed, A new algorithm for EEG feature selection using mutual information, Presented at IEEE, 2001.

[24] K. Matthias, M. Peter, BCI competition 2003-Data Set IIb: support vector machines for the P300 speller paradigm, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1073–1076.

[25] D. Mohamed, A.A. Ahmed, A new algorithm for EEG feature classification using mutual information, Presented at IEEE, 2001.

[26] B.O. Petersa, G. Pfurtschellerb, H. Flyvbjerg, Mining multi-channel EEG for its information content: an ANN-based method for a brain-computer interface, Neural Networks 11 (1998) 1429–1433.

[27] K.Z. Mao, K.C. Tan, W. Ser, Probabilistic neural-network structure determination for pattern classification, IEEE Trans. Neural Networks 39 (2003) 1010–1017.