Wavelet-based fractal features with active segment selection:
Application to single-trial EEG data

Wei-Yen Hsua, Chou-Ching Linb, Ming-Shaung Juc, Yung-Nien Suna,a

a Department of Computer Science & Information Engineering, Tainan, Taiwan, ROC
b Department of Neurology, University Hospital, Tainan, Taiwan, ROC
c Department of Mechanical Engineering, National Cheng Kung University, Tainan, Taiwan, ROC

Received 1 February 2006; received in revised form 29 January 2007; accepted 6 February 2007

Abstract

Feature extraction in brain–computer interface (BCI) work is one of the most important issues that significantly affect the success of brain signal classification. A new electroencephalogram (EEG) analysis system utilizing active segment selection and multiresolution fractal features is designed and tested for single-trial EEG classification. Applying event-related brain potential (ERP) data acquired from the sensorimotor cortices, the proposed system consists of three main procedures including active segment selection, feature extraction, and classification. The active segment selection is based on the continuous wavelet transform (CWT) and Student’s two-sample t-statistics, and is used to obtain the optimal active time segment in the time–frequency domain. We then utilize a modified fractal dimension to extract multiresolution fractal feature vectors from the discrete wavelet transform (DWT) data for movement classification. By using a simple linear classifier, we find significant improvements in the rate of correct classification over the conventional approaches in all of our single-trial experiments for real finger movement. These results can be extended to see the good adaptability of the proposed method to imaginary movement data acquired from the public databases.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Brain–computer interface; Event-related potential; Wavelet transform; Fractal dimension; Electroencephalogram; Single-trial classification

1. Introduction

The ultimate objective of brain–computer interface (BCI) technology is to provide human beings with an alternative communication channel that allows the direct transmission of messages from a brain to a computer by analyzing the brain’s mental activity (McFarland et al., 2003; Millán et al., 2004; Parra et al., 2002; Pfurtscheller et al., 1993, 2000b; Trejo et al., 2003; Wolpaw et al., 1991, 2002). The brain activity is recorded by means of multi-electrode EEG that is either invasive (Leuthardt et al., 2004; Levine et al., 2000) or non-invasive (Penny et al., 2000). Non-invasive recording is convenient and popular in BCI applications so it is adopted in this study. According to the definition suggested at the first international meeting for BCI technology, the term BCI is reserved for a system that must not depend on the brain’s normal output pathways of peripheral nerves and muscles (Wolpaw et al., 2000). BCI systems based on the single-trial analysis of electroencephalographic (EEG) signals associated with finger movements have grown rapidly in the last decade (Parra et al., 2002). The analysis has focused on discriminating between left and right finger movements and non-movement, using event-related brain potentials (ERP) in time–frequency EEG signal studies. It has been discovered that mu and beta rhythms reveal the variational characteristics of event-related synchronization (ERS) and desynchronization (ERD) over the sensorimotor cortices during finger movement tasks (Pfurtscheller and Lopes da Silva, 1999; Pineda et al., 2000). In this study, we propose a new system that classifies single-trial ERP data during finger movement tasks. The main procedures of the proposed system include active segment selection, feature extraction, and classification.

Left and right finger movements and non-movement can be discriminated by identifying active patterns of ERD and/or ERS located in the alpha and beta frequency bands. Since the ERD/ERS frequency bands are varied for different subjects, distinctive sensitive learning vector quantization has been proposed to determine the subject-specific frequency components (Pfurtscheller et al., 1998, 2000a). By using the spectral power at different frequency bands as the feature, a weighted distance
function is used to estimate the influence of different input features through supervised learning. Suitable weightings of features are then obtained for different frequency bands. The weighted features are then classified by using learning vector quantization, and error rates are obtained between 10% and 38%.

The discrete wavelet transform (DWT) is reportedly quite powerful in selecting features from the time–frequency domain (Demiralp et al., 1999). It is an efficient and structural approach in ERP representation but is sometimes difficult to adjust to ERP data with random lengths. The continuous wavelet transform (CWT) gives a highly redundant representation of EEG signals in the time-scale domain and is computationally very time consuming (Samar et al., 1999). However, it can be applied for the precise localization of ERP components in the time-scale domain (Bostanov, 2004a). Accordingly, we proposed to utilize the DWT in the training of active segment selection and use CWT and Student’s two-sample t-statistics for obtaining the optimal active segments from the time–frequency domain in the on-line detection.

Feature extraction in BCI work is one of the most important topics that significantly affect the success of classification. More specifically, the better the extracted feature, the higher the recognition rate we can expect. Consequently, an effective method of feature extraction is capable of obtaining excellent classification results even if the adopted classifier is unsophisticated. Several feature extraction methods, such as the band power evaluation and the autoregressive (AR) parameter model, are commonly used (Burke et al., 2005; Guger et al., 2001, 2003; Obermaier et al., 2001; Pardey et al., 1995; Penny et al., 2000; Pfurtscheller et al., 1997). Feature extraction methods based on band power estimation are usually obtained by computing their powers at a variety of alpha and/or beta bands, such as alpha bands (7–10, 10–13 Hz) and beta bands (16–20, 20–24, 24–30 Hz) (Guger et al., 2003; Obermaier et al., 2001). Frequency components in these bands are predominately involved in hand movement (Pfurtscheller et al., 1997). The estimated band powers are then computed with their logarithm values as descriptive parameters for every channel (Obermaier et al., 2001), or estimated by averaging over them (Guger et al., 2003). The parameters for various bands are then used for the subsequent analysis. On the other hand, AR models have also been popular in feature extraction for BCIs (Burke et al., 2005; Guger et al., 2001; Penny et al., 2000). The all-pole AR model lends itself well to modeling the EEG as filtered white noise with certain preferred energy bands. The EEG time series is fitted with an AR model. The AR model can be intuitively rephrased in the frequency domain as white noise source driving an all-pole spectral shaping network (Pardey et al., 1995). Thus, the band power and AR parameters are the two most popular features and selected as features for comparison in this study. Fractal geometry (Mandelbrot, 1982) provides a proper mathematical model to describe complex and irregular shapes that exist in nature by using fractal features. Fractal dimension is one of the most commonly used fractal features. Texture features based on fractal dimensions have been widely applied in image analysis for texture classification (Lee et al., 2003). In the last decade, feature extraction characterized by fractal dimension of EEG signals has been applied to various kinds of biomedical signal analyses such as routine detection of dementia (Henderson et al., 2000), seizure onset detection in epilepsy (Esteller et al., 1999; Gangadhar et al., 1995; Kirlangic et al., 2001) and EEG analyses of sleeping newborns (Accardo et al., 1998). However, not many fractal cases have been reported for BCI from EEG signals (Boostani and Moradi, 2004; Craig et al., 2005). In the proposed method, we first transfer the acquired EEG signals by using multiscale wavelet transform and then extract features from the multiresolution fractal dimensions. The extracted feature vectors are called multiresolution fractal feature vectors (MFFVs). Because different fractal sets that have strictly different appearances may share the same fractal dimension, single-scaled fractal dimension alone does not offer sufficient information to illustrate the signals (Mandelbrot, 1982). Hence, based on multi-resolution concepts, we provide a solution to describe the characteristic of fractal with different wavelet scales by MFFVs (Lee et al., 2003). In addition to multiple scale characteristics, it also contains important fractal information in the time-scale space. They are then used for movement discrimination. To confirm the system performance after applying active segment selection and feature extraction, linear discriminant analysis (LDA) is used for classifying the single-trial finger movement. The experiments show significant improvement in the rates of correct classification. Although the proposed EEG analysis system is designed for detection of real finger movement, it has also been tested on imaginary data acquired from publicly available datasets. The experiments show good results and great potential to adapt the current system to imaginary movement classification.

This paper is organized as follows. Section 2 describes the experimental design and the proposed signal analysis system. In Section 3, the experimental results on single-trial ERP data are presented. The discussion and conclusion are given in Sections 4 and 5, respectively.

2. Materials and methods

The proposed EEG signal analysis system for single trial finger movement classification is illustrated in Fig. 1. The flowchart describes the detailed procedures for both off-line training and on-line processing. During the off-line training, the data are trained by three main steps and an optional one to obtain the essential parameters for the corresponding steps in on-line processing. The three main steps of parameter training are for active segment selection, feature extraction and movement classification. The active segment selection is based on the CWT and Student’s two-sample t-statistics, and is used to obtain the location of optimal active segment in the time–frequency domain. Then we extract multi-resolution fractal feature vectors by using modified fractal dimension from the DWT data. Discrete wavelet transform (DWT) is much efficient in computation and is applied repeatedly for feature extraction for all the EEG data. If we directly use the multiscale DWT data as the features, the fea-
Fig. 1. Flowchart of BCI system. The system consists of three procedures including active segment selection, feature extraction, and classification.

The feature set becomes too large and complex and the resulting signal classification is inefficient and less reliable. Therefore, we proposed to use the modified fractal dimension method to extract the multi-resolution fractal feature vector from the DWT data. Finally, fractal features are used to train the parameters for a simple linear classifier. In addition, the optional procedure is designed for subband selection that selects appropriate subbands for each individual subject to reduce redundant information and increase classification accuracy. After the off-line training, we can utilize the obtained parameters for the on-line processing. The resulting location of selected active segment can be directly applied to the data in on-line processing. And then the same process of multiresolution modified fractal dimension is applied to extract feature vectors. The new feature vector is compact and effective for movement classification. At last a simple linear classifier with trained parameters is used to classify the states of finger movement. Other features, such as FFT and AR, are also extracted and compared in the experiments.

2.1. Data acquisition and mental task

EEG signals were recorded from five untrained subjects (four males and one female, two left-handed and three right-handed) in a shielded room using 13 silver/silver chloride electrodes. As illustrated in Fig. 2, they included 10 scalp EEG channels (C3, C5, FC3, C1, CP3, C4, C2, FC4, C6, and CP4), two EMG channels for monitoring left and right muscle activity, and one channel on the forehead to record possible EOG artifacts and eye blinks during the experiment (Jasper, 1958). All electrodes were referenced to the A1 lead at the left earlobe. Before being sampled at the rate of 256 Hz, the EEG data were filtered by an analog band-pass filter with cutoff frequencies at 0.5 and 100 Hz, and amplified by a multiple of 10,000. During the experiments, each subject was asked to perform four trials, which included left finger lifting, resting, right finger lifting, and resting in sequence, in each test. Each trial was 10 s in length; therefore, it took 40 s in a test. For each lifting trial, the first 4 s was quiet and then an acoustic stimulus was given as a cue to signify the beginning of left or right finger lifting. At the same time, each subject was asked to execute a finger lifting. We recorded 60 tests for each subject, and thus there were 240 trials for each subject. No trials were removed during the EEG data processing stage. Therefore, we used totally 1200 trials from five subjects in the finger movement experiments. Data segments representing three states were subsequently extracted from the experimental data set, that is, the above-mentioned left finger lifting, right finger lifting and finger resting. Data segments of the finger lifting were acquired from second $-2$ to second 2, where second 0 stands for the trigger of movement by detecting the peak EMG signal after linear envelope processing. Data segments of finger resting were randomly acquired from 4-s windows within the finger resting sections of each trial. Data segments of all states were then used in the subsequent data preprocessing.

2.2. Active segment selection in time–frequency domain

Prior to the feature extraction for final classification, each data segment comprises a 4-s window that is too redundant to...
and Kotchoubey, 2004b). Therefore, CWT is applied here for the precise localization of ERP components in the time–frequency domain. In this proposed method, active segment selection based on CWT and Student’s two-sample t-statistics is calculated and analyzed respectively:

The CWTs of the three states of the data segments that are randomly acquired from finger resting trials. The Laplacian filtering will be applied to the data sets for both off-line training and on-line processing. The filtered EEG signals are first used to obtain the system parameters based on the flowchart in Fig. 3(a). Fig. 4 shows the intermediate results obtained in the procedure of active segment selection from the training data set in the time–frequency domain. In Fig. 4, C3 and C4 represent the channels C3 and C4, respectively. LN, RN, and LR represent the cases of left finger lifting versus finger resting, right finger lifting versus finger resting, and left versus right finger lifting, respectively. Fig. 4 shows EEG data with 4 s segment ranging from sample 1 to 1024 at the sample rate of 256 Hz. The 4 s segment for finger lifting was acquired from the 10 s window of a finger lifting trial with its center standing for the trigger of movement, while the 4 s segment for finger resting was randomly acquired from finger resting trials.

2.2.2. Continuous wavelet transform (CWT)

The CWTs of the three states of the data segments that are posterior to Laplacian filtering in both channels C3 and C4 are analyzed respectively:

$$W_s^{C,n}(j, k) = \int_R f_s^{C,n}(x) \frac{1}{\sqrt{j}} \psi \left( \frac{x-k}{j} \right) \, dx,$$

where \((1/\sqrt{j})\psi((x-k)/j)\) are the dilated and translated versions of the wavelet function \(\psi(x)\) at scale \(j\) and shift \(k\), and \(W_s^{C,n}(j, k)\) represents the CWT of the data segment \(f_s^{C,n}(x)\), in which state \(s\) belongs to one of three distinct states, data segment \(C\) belongs to either channel C3 or C4 after the Laplacian filtering, and \(n\) represents each single trial. Here, the values of \(W_s^{C,n}(j, k)\) afford rapid computation and achieve high recognition accuracy. It is important to reduce the window by selecting only the active segment, which is most contributive to representing the motion states from the original 4-s segment. Obviously, a better selected active segment results in a higher degree of classification accuracy. Because CWT is capable of offering good time and frequency localization, it is often used to construct a multi-resolution representation of biological signal (Bostanov and Kotchoubey, 2004b). Therefore, CWT is applied here for precise localization of ERP components in the time–frequency domain. In this proposed method, active segment selection based on CWT and Student’s two-sample t-statistics is calculated and utilized to obtain the optimal active segment in time–frequency domain. A flowchart of active segment selection is shown in Fig. 3(a).

2.2.1. Data preprocessing

Non-EEG noise is significantly different from EEG signals in both topographical and frequency characteristics. The mu and beta rhythms of the EEG are those components with frequencies distributed between 8 and 32 Hz and located over the sensorimotor cortices. EOG signals are maximal at low frequencies distributed between 8 and 32 Hz and located over the mu and beta rhythms of the EEG are those components with frequencies...
Fig. 4. Procedure for active segment selection in the time–frequency domain. (a) Mean CWT. (b) Mean of the smoothed CWT data by short segmental power sum method. (c) Student’s two-sample t-statistics. (d) Student’s two-sample t-statistics without short segmental power sum method. Also, C3 and C4 represent the channels C3 and C4. LN, RN, and LR represent the cases of left finger lifting vs. finger resting, right finger lifting vs. finger resting, and left vs. right finger lifting.

can be represented with a 2D time-scale plot, which holds the special properties that keep the optimal scale separation of ERP components (Bostanov and Kotchoubey, 2004b). The reason we choose the Daubechies wavelet as the CWT function in Eq. (3) is mainly due to the special characteristic that Daubechies family wavelets are compactly supported with extreme phase and highest number of vanishing moments for a given support width. In addition to the property that the associated scaling filters are minimum-phase, Daubechies wavelet can be excellently applied on DWT as well as CWT and is suitably used for the detection of ERP components and salient oscillations (Daubechies, 1992; Ende et al., 1998). In Fig. 4(a), we can see the mean CWT which is obtained by directly averaging all the CWT data from one of the subjects in the training data set.

2.2.3. Short segmental power sum method

However, the 2D time-scale plot for $W^C_{j,n}(j, k)$ generated from CWT is very noisy. It is difficult to accurately locate the active segment based on the extreme of spectral difference later in this method. Therefore, the short segmental power sum method is employed to smooth the power spectrum and reduce the noise effects in the time–frequency domain. Because the 4-s windows are too redundant for real time applications, we select the optimal active segments with a 1-s window for the subsequent processing. The length of the data segment with a 1-s window is a compromise between the cost of computation and the ERP component applications. Short segmental data which are processed by the short segmental power sum method for the three states of data segments in both channels C3 and C4, are obtained from the training data set:

$$P^C_{s,n}(j, k) = \sum_{t=-h}^{h} (W^C_{s,n}(j, k + t))^2,$$

where $h$ stands for the sample points of half a second. After smoothing the CWT data with the short segmental power sum method with Eq. (4), Fig. 4(b) illustrates the mean of smoothed CWT data.

2.2.4. Information accumulation in scale space

The values of $P^C_{s,n}(j, k)$ with different scales $j$ and time $k$ for the state $s$ can form a 2D time-scale plot, but this plot contains smoother characteristics compared to the original wavelet transformed signals. However, the best discriminant region does not exist at a single point but lies in a particular range of time and scale. We have considered the time range in the above-mentioned short segmental power sum method; we will now describe how the scale range is subsequently taken into account. In general, using a wider frequency range from the acquired EEG signals can achieve higher classification accuracy in comparison with a narrower one (Müller-Gerking et al., 1999). A wide frequency range containing all mu and beta rhythm components is adopted.
to include all the important signal spectrums for motion classification. In our experiments, the active frequency ranges for all subjects are almost located in the frequency band between 8 and 32 Hz. The frequency band 8–32 Hz is thus adopted in computing the Student’s two-sample t-statistics for obtaining the active segment of each subject. The frequency range is in fact subject-dependent and obtained from the CWT spectrum for each subject. It could be different from 8 to 32 Hz for some special subjects in some other experiments; however the subjects all have their band mainly in 8–32 Hz in the reported experiments. Therefore, we take the summation of the plot \( p_s^{C,n}(j,k) \) between the scale ranges \( 8 \leq j \leq 32 \) along the scale coordinate \( j \):

\[
\pi_s^{C,n}(k) = \sum_{j=8}^{32} p_s^{C,n}(j,k),
\]

(5)

where \( \pi_s^{C,n}(k) \) represents the sum of all the scales that contain mu and beta rhythm components for the state \( s \). Hence, this resulting function can suitably represent the brain ERP responses for the three states of finger movement.

2.2.5. Student’s two-sample t-statistics

Means and variances for the three states, i.e. left finger lifting, right finger lifting, and finger resting, of data after the process of information accumulation in scale space in both channels C3 and C4 are calculated from the training data set:

\[
\bar{\pi}_s^{C}(k) = \frac{1}{N_s} \sum_{n=1}^{N_s} \pi_s^{C,n}(k),
\]

(6)

where \( N_s \) denotes the number of trials in state \( s \).

\[\sigma^2(k) = \frac{1}{N_s - 1}\sum_{n=1}^{N_s} (\pi_s^{C,n}(k) - \bar{\pi}_s^{C}(k))^2,\]

(7)

Student’s two-sample t-statistics evaluated between any two of the three distinct states are subsequently represented as

\[
t_{s_1s_2}(k) = \frac{\bar{\pi}_{s_i}^{C}(k) - \bar{\pi}_{s_j}^{C}(k)}{\sqrt{((N_{s_1} - 1)\sigma_{s_i}^{2C}(k) + (N_{s_2} - 1)\sigma_{s_j}^{2C}(k))/(N_{s_1} + N_{s_2} - 2))((1/N_{s_1}) + (1/N_{s_2}))}},
\]

(8)

where \( s_1 \) and \( s_2 \) belong to two different states. The denominator in Eq. (8) indicates the pooled variance of the two states for channel C. The values of \( t_{s_1s_2}(k) \) with different time \( k \) for the two distinct states, \( s_1 \) and \( s_2 \), can form a 1D function with respect to time, but this contains different characteristics compared to the original wavelet transformed signal. Points that have local peak in profile \( t_{s_1s_2}^{C}(k) \) imply that they are with local maximal difference between the two states in the time-scale domain. In other words, it means that these two states of EEG signals are best discriminated at the particular time and scale range.

After applying information accumulation in scale space, we can then obtain the Student’s two-sample t-statistics as illustrated in Fig. 4(c). These t-statistics will then be used to select the optimal active segment. The active 1 s segment for each pair of finger movement states was selected with its center being the peak after the C3 and C4 channels were concatenated together. Besides, Fig. 4(d) shows the Student’s two-sample t-statistics without using the short segmental power sum method which is obviously too noisy to afford successful recognition.

2.2.6. Optimal active segment selection

Because channels C3 and C4 should be taken into account simultaneously for selecting the optimal active segment, each pair of channels \( t_{s_1s_2}^{C_3}(k) \) and \( t_{s_1s_2}^{C_4}(k) \) is concatenated together to obtain the value of \( t_{s_1s_2}^{C_{con}}(k) \). The optimal active segment \( A_{s_1s_2}^{C_{con}} \) is a 1-s window whose center is at the peak (global maximum) value of \( t_{s_1s_2}^{C_{con}}(k) \). Finally, the optimal active segments for each pair of states are obtained and used in the subsequent feature extraction and classification.

2.3. Feature extraction

Instead of directly classifying the native EEG data, feature extraction is performed on the 1-s window of the optimal active segment in the time–frequency domain. For each pair of distinct states selected from the three original states, there are in total three sets of optimal active segments used in the subsequent feature extraction. Feature extraction significantly affects the success of classification. Accordingly, the better the extracted feature, the more we can expect a high recognition rate. Although powerful classifiers, such as support vector machines (SVMs), may yield quite good recognition results using simple and/or redundant features, an effective feature extraction method is capable of producing excellent classification results even if an unsophisticated classifier, such as linear discriminant analysis (LDA), is used (Müller et al., 2003). In this study, EEG data are first filtered to the frequency range between 8 and 32 Hz with a Butterworth band-pass filter. Then, the DWT and fractal dimension computation can be carried out on the filtered EEG data. Because the filtered EEG data are within the frequency range of 8–32 Hz, we simply apply the regular Daubechies DWT to decompose the filtered EEG data into multi-resolution bands and feed the transformed data to obtain the FD feature vector. A flowchart of feature extraction in the study is shown in Fig. 3(b).

2.3.1. DWT

In general, multiresolution analysis decomposes a signal into numerous details at various resolutions, where each resolution represents a class of distinct physical characteristics within the signal. In other words, a signal is characterized with the formulation by decomposing it into subbands, and each subband can be treated individually based on its characteristics. A popular type of multiresolutional representation can be achieved by wavelet transform decomposition.

The optimal active segment \( A_{s_1s_2}^{C_{con},n} \) of each trial \( n \) in the training data for the two distinct states \( s_1 \) and \( s_2 \) can be represented
in terms of the DWT as,
\[
A_{s_{1/2}}^{\text{con}-n}(x) = \sum_{k=-\infty}^{\infty} (S_j)_{s_{1/2}}^{\text{con}-n}(k)2^{J/2}\phi(2^Jx-k) + \sum_{j=1}^{J} \sum_{k=-\infty}^{\infty} (D_j)_{s_{1/2}}^{\text{con}-n}(k)2^{J/2}\psi(2^Jx-k),
\]
where the expansion coefficients are determined by
\[
(S_j)_{s_{1/2}}^{\text{con}-n}(k) = \langle A_{s_{1/2}}^{\text{con}-n}(x), 2^{J/2}\phi(2^Jx-k) \rangle,
\]
\[
(D_j)_{s_{1/2}}^{\text{con}-n}(k) = \langle A_{s_{1/2}}^{\text{con}-n}(x), 2^{J/2}\psi(2^Jx-k) \rangle,
\]
where \((S_j)_{s_{1/2}}^{\text{con}-n}(k) \) and \((D_j)_{s_{1/2}}^{\text{con}-n}(k) \) represent the approximation and detail spaces of \(A_{s_{1/2}}^{\text{con}-n} \), respectively; \(2^{J/2}\phi(2^Jx-k) \) and \(2^{J/2}\psi(2^Jx-k) \) are the dilated and translated versions of the scaling function \(\phi(x) \) and the wavelet function \(\psi(x) \), respectively. The optimal active segment \(A_{s_{1/2}}^{\text{con}-n} \) is then decomposed into subbands \( (S_j)_{s_{1/2}}^{\text{con}-n}(k), \ldots, (S_1)_{s_{1/2}}^{\text{con}-n}(k), (D_j)_{s_{1/2}}^{\text{con}-n}(k), \ldots, (D_1)_{s_{1/2}}^{\text{con}-n}(k) \) by Eq. (10).

2.3.2. Modified fractal dimension

Fractal geometry provides a proper mathematical model to describe a complex shape that exists in nature with fractal features. However, fractal dimension (FD) is one of the most popular fractal features. Because FD is relatively insensitive to signal scaling and shows a strong correlation with human judgment of surface roughness, we select it as our feature extraction method (Pentland, 1984). A bounded set \( A \) in Euclidean \( n \)-space is self-similar, if \( A \) is the union of \( N_r \) non-overlapping copies of itself scaled down by a ratio \( r \) through all its coordinates. The fractal dimension of the bounded set \( A \) in \( R^n \) is a real number used to characterize the geometrical complexity of \( A \). The fractal dimension \( D_f \) of \( A \) is derived by the following relation (Mandelbrot, 1982):
\[
D_f = \lim_{r \to 0} \frac{\log(N_r)}{\log(1/r)}.
\]

Based on this equation, several approaches were presented to estimate the fractal dimension from signals or images (Gangepain and Roques-Carmes, 1986; Lee et al., 2003; Peleg et al., 1984; Pentland, 1984). Among them, the differential box counting (DBC) method that covers a wide dynamic range with a low computational complexity is popular and frequently used. \( N_r \) is estimated by DBC in the proposed method.

Assume that a signal with \( M \) samples has been scaled down to consist of \( s \) samples, where \( 1 < s \leq M/2 \) and \( s \) is an integer. Hence, the defined scale ratio \( r \) is \( s/M \). Let us consider the signal in two-dimensional space with \( x \) indicating position and \( y \) denoting amplitude. The \( x \) dimension is then partitioned into grids with length \( s \). On each grid, there is a column of rectangles each of which is of height \( h \) and size \( h \times s \). In addition, the difference between the maximum and minimum amplitudes of the signal is \( H \) in height.

However, the original DBC method is based on the difference between the minimum and maximum rectangle numbers, and it is easily disturbed by noise. Since the standard deviation of the amplitude represents the dispersion of the signal, the standard deviation is used to replace the difference in rectangle numbers in original DBC method. Moreover, DBC produces a stair-like function and may result in the underestimation of \( D_f \). In order to resolve these problems, we modify the calculation of \( D_f \), which we refer to as modified fractal dimension. The height \( h \) of a rectangle is calculated with \( h = Hs/M \) instead of \( [H/h] = [M/s] \). Then, the signal variation \( n_r(i) \) at the \( i \)th grid is computed as
\[
n_r(i) = 2\sigma_i / h + 1,
\]
where \( \sigma_i \) represents the standard deviation of the amplitude at the \( i \)th grid.

Summing up the signal variations from all grids, we obtain
\[
N_r = \sum_i n_r(i).
\]

This floating calculation can partially resolve the problems caused by a stair-like function. The contribution of all grids is calculated by Eq. (13). Then, the fractal dimension \( D_f \) can be estimated by the least-squares linear fitting of \( \log(N_r) \) versus \( \log(1/r) \).

2.3.3. Multiresolution fractal feature vectors

After the optimal active segment \( A_{s_{1/2}}^{\text{con}-n} \) is defined, the resulting subbands \( (S_j)_{s_{1/2}}^{\text{con}-n}(k), \ldots, (S_1)_{s_{1/2}}^{\text{con}-n}(k), (D_j)_{s_{1/2}}^{\text{con}-n}(k), \ldots, (D_1)_{s_{1/2}}^{\text{con}-n}(k) \) can then be used to extract the needed feature vectors for the subsequent classification process. The adopted feature vector is composed of the fractal dimensions calculated from the active segment and its corresponding subbands with the modified fractal dimension method described above. The \((2^J+1)\) resulting feature values are denoted as \( D_1, D_1^{(S_1)}, \ldots, D_1^{(D_1)} \), \( D_2, D_2^{(S_1)}, \ldots, D_2^{(D_1)} \), \( D_3, D_3^{(S_1)}, \ldots, D_3^{(D_1)} \), and \( D_4, D_4^{(S_1)}, \ldots, D_4^{(D_1)} \). The estimated fractal dimensions reflect the roughness (or complexity) of the evaluated data. Two types of multiresolution fractal feature vectors are investigated. One includes the fractal dimensions from all subbands, and the other covers only \((S_j)_{s_{1/2}}^{\text{con}-n}(k) \) and all detailed subbands. They are expressed as follows:

\[
\text{MFFV1} = \{D_1, D_1^{(S_1)}, \ldots, D_1^{(D_1)}\}
\]
\[
\text{MFFV2} = \{D_1, D_1^{(S_1)}, \ldots, D_1^{(S_1)}, D_1^{(D_1)}, \ldots, D_1^{(D_1)}\}.
\]

2.3.4. Genetic algorithm

The genetic algorithm (GA) (Whitley, 1994) is used as an optional procedure for subband selection. The GA is a specific class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, selection, crossover and mutation. The procedure of the GA used for subband selection is described in detail as follows:

1. Generate the initial population of \( N \) chromosomes from all subbands (\( N=20 \)). Each chromosome consists of a string
of 0 and 1, and its length $L$ is the number of subbands ($L = 2^J + 1$).

2. Evaluate the fitness of each chromosome in the population according to the correct recognition rate obtained by the LDA classifier.

3. Selection. The roulette wheel selection is applied to select parent chromosomes from a population according to their fitness. In other words, the better fitness, the bigger chance to be selected.

4. Crossover. With a crossover probability $p_c$, crossover the parents to form new offspring (here, $p_c = 0.3$).

5. Mutation. With a mutation probability $p_m$, mutate new offspring at each position in chromosome ($p_m = 0.1$).

6. If the end condition is satisfied, the GA stops and returns the best solution in current population; otherwise, it goes to step 2 to repeat.

Although the GA parameters, including the initial population ($N = 20$), the crossover probability ($p_c = 0.3$), and the mutation probability ($p_m = 0.1$), are chosen experimentally, the selection of GA parameters is not sensitive with respect to the classification results. The GA that reduces redundant information and increases classification accuracy is used in the study to select the appropriate subbands for each individual subject. When GA is applied, the new feature vectors are formed with near optimal subband selection.

2.4. Classification

Although many classification methods have been used to classify EEG signals, we use a simple classifier in this study. A more effective classifier may further improve the classification results. In the proposed system, each trial obtains a multiresolution fractal feature vector which is then used to classify the corresponding state using linear discriminant analysis (LDA). LDA classifies the two given classes based on the assumption that both classes are under normal distribution with their mean vectors and covariance matrices estimated from the training data set. The covariance matrices are usually assumed to be identical for the two classes. A test is classified to the state that has the larger discriminant value of LDA.

3. Results

In the experimental study, the classification tests for both real and imaginary finger movement data are carried out using five-fold crossvalidation. More specifically, the data set for each subject is divided into five subsets, and the following procedure is repeated five times. Each time, one of the five subsets is used as the test set and the other four comprise a training set. The average recognition rate is immediately evaluated across all five-folds. In the first experiment, we evaluated the method in selecting the optimal active segment as a 1-s window from the original 4-s window. If the selected window contains sufficient movement information and can achieve discrimination results as good as (or even better than) the 4-s window, it can reduce computation and memory demands and make real-time (or on-line) EEG motion analysis feasible. In order to confirm the effectiveness of optimal active segment selection, data segments of the fixed 1-s window, the 4-s window, and the 1-s window with active segment selection are compared for their performances. The comparison of recognition rates using FFT spectra among the 4-s window and 1-s windows without and with active segment selection are listed in Table 1. The average recognition rate is immediately evaluated across all five-folds. In the first experiment, we evaluated the method in selecting the optimal active segment as a 1-s window from the original 4-s window. If the selected window contains sufficient movement information and can achieve discrimination results as good as (or even better than) the 4-s window, it can reduce computation and memory demands and make real-time (or on-line) EEG motion analysis feasible. In order to confirm the effectiveness of optimal active segment selection, data segments of the fixed 1-s window, the 4-s window, and the 1-s window with active segment selection are compared for their performances. The comparison of recognition rates using FFT spectra among the 4-s window and 1-s windows without and with active segment selection are listed in Table 1. The average recognition rates for FFT applied to the 4-s window are 74.7%, 77.0% and 64.7%, respectively. The average recognition rates for FFT applied to the 1-s window without active segment selection are 65.2%, 68.5% and 50.9%, respectively. After applying the selection of optimal active segments, the average recognition rates for FFT rise to 75.7%, 76.8% and 62.5%, respectively.

To demonstrate the effectiveness of the proposed active segment selection, we compare the resulting recognition rates by using the two proposed features MFFV1 and MFFV2 both with and without the active segment selection. In the experiments, the input features are extracted from a 1-s window of the selected active segment. However, for the case without active segment selection, the features are extracted from the 1-s segment centered at the trigger, which is at the center of every 4-s input data segment. The resulting recognition rates for MFFV1 are listed in Table 2(a). The averaged rates for MFFV1 without active segment selection are 70.7%, 75.7% and 52.8%, respectively. Correspondingly, the rates with active segment selection increase to 79.7%, 85.8% and 70.0%, respectively. In Table 2(b), the comparisons for MFFV2 are listed. The average rates for

<table>
<thead>
<tr>
<th>FFT</th>
<th>Subject</th>
<th>$S001$ (%)</th>
<th>$S002$ (%)</th>
<th>$S003$ (%)</th>
<th>$S004$ (%)</th>
<th>$S005$ (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 s window</td>
<td>Left vs. none</td>
<td>96.7</td>
<td>77.5</td>
<td>60.8</td>
<td>65.0</td>
<td>73.3</td>
<td>74.7</td>
</tr>
<tr>
<td></td>
<td>Right vs. none</td>
<td>94.2</td>
<td>87.5</td>
<td>78.3</td>
<td>67.5</td>
<td>57.5</td>
<td>77.0</td>
</tr>
<tr>
<td></td>
<td>Left vs. right</td>
<td>80.8</td>
<td>62.5</td>
<td>60.8</td>
<td>65.0</td>
<td>54.2</td>
<td>64.7</td>
</tr>
<tr>
<td>1 s window without active segment selection</td>
<td>Left vs. none</td>
<td>94.2</td>
<td>63.3</td>
<td>60.0</td>
<td>54.2</td>
<td>54.2</td>
<td>65.2</td>
</tr>
<tr>
<td></td>
<td>Right vs. none</td>
<td>94.2</td>
<td>71.7</td>
<td>70.8</td>
<td>56.7</td>
<td>49.2</td>
<td>68.5</td>
</tr>
<tr>
<td></td>
<td>Left vs. right</td>
<td>54.2</td>
<td>52.5</td>
<td>48.4</td>
<td>49.2</td>
<td>50.0</td>
<td>50.9</td>
</tr>
<tr>
<td>1 s window with active segment selection</td>
<td>Left vs. none</td>
<td>97.5</td>
<td>80.8</td>
<td>60.8</td>
<td>67.5</td>
<td>71.7</td>
<td>75.7</td>
</tr>
<tr>
<td></td>
<td>Right vs. none</td>
<td>94.2</td>
<td>89.2</td>
<td>77.5</td>
<td>67.5</td>
<td>55.8</td>
<td>76.8</td>
</tr>
<tr>
<td></td>
<td>Left vs. right</td>
<td>76.7</td>
<td>65.8</td>
<td>57.5</td>
<td>62.5</td>
<td>50.0</td>
<td>62.5</td>
</tr>
</tbody>
</table>
Left vs. none 97.5 71.7 70.8 56.7 55.0 70.3
Right vs. none 96.7 72.5 78.3 69.2 63.3 76.0
Left vs. right 55.0 57.5 49.2 54.2 57.5 54.7

With active segment selection

Without active segment selection
Left vs. none 99.2 86.7 73.3 72.5 72.5 80.7
Right vs. none 97.5 91.7 86.7 80.0 73.3 85.8
Left vs. right 83.3 79.2 65.0 65.0 57.5 70.0

With active segment selection

MFFV2 without active segment selection are 70.3%, 76.0% and 54.7%, respectively. Finally, the corresponding rates for MFFV2 with active segment selection increase to 80.7%, 85.3% and 69.8%, respectively.

Some experiments for evaluating the performance of the proposed multiresolution fractal feature vectors MFFV1 and MFFV2 are also given. Several commonly used features, including the fast Fourier transform (FFT) and the autoregressive (AR) model parameters are also used as the bases for comparison (Burke et al., 2005; Guger et al., 2001, 2003; Obermaier et al., 2001; Pardey et al., 1995; Penny et al., 2000; Pfurtscheller et al., 1997). The extracted FFT components are in the frequency range of 8–32 Hz, which contains all the mu and beta rhythm components of the EEG data. The AR model parameters of the same frequency range are obtained using Burg’s method for a 6th order AR model. It minimizes the forward and backward prediction errors while satisfying Levinson–Durbin recursion. The comparison of recognition rates among FFT, 6th order AR parameters, MFFV1, MFFV2 are listed in Table 3(a).

MFFV2 is formed by collecting fractal dimensions of the optimal active segment and all of its non-separated subbands. Since it contains all generated subbands, it is somewhat redundant for efficient computation. It also slightly reduces the classification accuracy in our experiments. However, the selection of subbands to form the multiresolution fractal feature vector depends on the specific applications and subjects. For obtaining more effective features, a GA (Whitley, 1994) that reduces redundant information and increases classification accuracy is used to select appropriate subbands for each individual subject in this study. The classification rates for MFFV2 after using the GA are listed in the bottom of Table 3(a).

Although the proposed method was designed for recognizing real finger movements, we also extend the studies to test its adaptability to imaginary finger movements. There are many factors, e.g. experimental design, subject training and detection strategy, concerned with an EEG signal analysis system dedicated for imaginary movement detection. To study the best factors for imaginary movement analysis is beyond the scope of our current studies. Therefore, we directly apply our method to some publicly available BCI data sets recorded by the Graz BCI group (Guger et al., 2001; Graz data set and description for the BCI, 2003 competition). These data are widely used in many studies and some BCI competitions (Haselsteiner and Pfurtscheller, 2000; Schlögl et al., 1997). The adopted data sets are recorded from three subjects in a cued experimental recording procedure. Each trial is 8–9 s in length. The first 2 s is quiet. With an acoustic stimulus signifies the beginning of a trial, a fixation cross is then displayed for 1 s. Then, an arrow (left or right) is displayed and the data considered as event related are recorded. At the same time, each subject is asked to move a bar in the direction of cue by imagining moving the left or right hand. The data consist of 280 trials for subject Im001 and 320 trials for subject Im002 and subject Im003. All signals are sampled at 128 Hz and filtered between 0.5 and 30 Hz. Two bipolar EEG channels were measured using two electrodes positioned 2.5 cm posterior and anterior to position C3 and C4 according to the international standard (10/20 system) electrode positioning nomenclature. All the procedures adopted in analyzing imaginary finger movements are the same as those used in analyzing the real ones. Comparison of recognition rates applied to these data sets among FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 after using GA are listed in Table 3(b).

Another popular approach to evaluate the performance of feature extraction is by means of the receiver operating characteristics (ROC) curve. The ROC curve is a plot of the true positive rate against the false positive rate for every possible cutoff of a diagnostic test. The classification accuracy can be measured by the area under the ROC curve (AUC) that is used to describe the probability of random samples assigning to the correct class. The comparison of AUC among FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 after using GA is listed in Table 4.
### Table 3
Comparison of recognition rates among (a) FFT and (b) imaginary data FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 after using GA

<table>
<thead>
<tr>
<th>Subject</th>
<th>S001 (%)</th>
<th>S002 (%)</th>
<th>S003 (%)</th>
<th>S004 (%)</th>
<th>S005 (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) FFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>97.5</td>
<td>80.8</td>
<td>60.8</td>
<td>67.5</td>
<td>71.7</td>
<td>75.7</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>94.2</td>
<td>89.2</td>
<td>77.5</td>
<td>67.5</td>
<td>55.8</td>
<td>76.8</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>76.7</td>
<td>65.8</td>
<td>57.5</td>
<td>62.5</td>
<td>50.0</td>
<td>62.5</td>
</tr>
<tr>
<td>AR parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>95.8</td>
<td>77.5</td>
<td>68.3</td>
<td>58.3</td>
<td>60.0</td>
<td>72.0</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>90.8</td>
<td>85.0</td>
<td>87.5</td>
<td>64.2</td>
<td>60.8</td>
<td>77.7</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>83.3</td>
<td>61.7</td>
<td>61.7</td>
<td>63.3</td>
<td>56.7</td>
<td>65.3</td>
</tr>
<tr>
<td>MFFV1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>99.2</td>
<td>85.0</td>
<td>71.7</td>
<td>71.7</td>
<td>70.8</td>
<td>79.7</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>97.5</td>
<td>91.7</td>
<td>86.7</td>
<td>80.0</td>
<td>73.3</td>
<td>85.8</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>83.3</td>
<td>79.2</td>
<td>65.0</td>
<td>65.0</td>
<td>57.5</td>
<td>70.0</td>
</tr>
<tr>
<td>MFFV2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>99.2</td>
<td>86.7</td>
<td>73.3</td>
<td>71.7</td>
<td>72.5</td>
<td>80.7</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>96.7</td>
<td>91.7</td>
<td>87.5</td>
<td>76.7</td>
<td>74.2</td>
<td>85.4</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>80.8</td>
<td>79.2</td>
<td>61.7</td>
<td>70.0</td>
<td>57.5</td>
<td>69.8</td>
</tr>
<tr>
<td>MFFV2 after using GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>99.2</td>
<td>91.7</td>
<td>81.7</td>
<td>82.5</td>
<td>81.7</td>
<td>87.4</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>98.3</td>
<td>96.7</td>
<td>92.5</td>
<td>81.7</td>
<td>78.3</td>
<td>89.5</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>87.5</td>
<td>85.8</td>
<td>71.7</td>
<td>80.0</td>
<td>68.3</td>
<td>78.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Im001 (%)</th>
<th>Im002 (%)</th>
<th>Im003 (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b) Left vs. right</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FFT</td>
<td>84.9</td>
<td>80.4</td>
<td>83.3</td>
<td>82.9</td>
</tr>
<tr>
<td>AR parameters</td>
<td>87.4</td>
<td>76.8</td>
<td>85.9</td>
<td>83.4</td>
</tr>
<tr>
<td>MFFV1</td>
<td>91.1</td>
<td>87.5</td>
<td>90.8</td>
<td>89.8</td>
</tr>
<tr>
<td>MFFV2</td>
<td>90.3</td>
<td>83.9</td>
<td>91.4</td>
<td>88.5</td>
</tr>
<tr>
<td>MFFV2 after GA</td>
<td>93.6</td>
<td>91.2</td>
<td>92.5</td>
<td>92.4</td>
</tr>
</tbody>
</table>

### Table 4
Comparison of the area under the ROC curve (AUC) among FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 after using GA

<table>
<thead>
<tr>
<th>AUC</th>
<th>S001 (%)</th>
<th>S002 (%)</th>
<th>S003 (%)</th>
<th>S004 (%)</th>
<th>S005 (%)</th>
<th>Average (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>0.92</td>
<td>0.83</td>
<td>0.63</td>
<td>0.66</td>
<td>0.62</td>
<td>0.73</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>0.92</td>
<td>0.86</td>
<td>0.76</td>
<td>0.70</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.79</td>
<td>0.61</td>
<td>0.54</td>
<td>0.64</td>
<td>0.43</td>
<td>0.60</td>
</tr>
<tr>
<td>AR parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>0.92</td>
<td>0.80</td>
<td>0.70</td>
<td>0.59</td>
<td>0.62</td>
<td>0.73</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>0.91</td>
<td>0.82</td>
<td>0.84</td>
<td>0.68</td>
<td>0.64</td>
<td>0.78</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.82</td>
<td>0.55</td>
<td>0.65</td>
<td>0.55</td>
<td>0.42</td>
<td>0.60</td>
</tr>
<tr>
<td>MFFV1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>0.95</td>
<td>0.91</td>
<td>0.74</td>
<td>0.72</td>
<td>0.75</td>
<td>0.81</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>0.95</td>
<td>0.93</td>
<td>0.92</td>
<td>0.78</td>
<td>0.75</td>
<td>0.87</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.87</td>
<td>0.83</td>
<td>0.73</td>
<td>0.78</td>
<td>0.63</td>
<td>0.77</td>
</tr>
<tr>
<td>MFFV2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>0.95</td>
<td>0.90</td>
<td>0.75</td>
<td>0.76</td>
<td>0.77</td>
<td>0.83</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>0.95</td>
<td>0.93</td>
<td>0.91</td>
<td>0.78</td>
<td>0.72</td>
<td>0.86</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.86</td>
<td>0.83</td>
<td>0.68</td>
<td>0.80</td>
<td>0.61</td>
<td>0.76</td>
</tr>
<tr>
<td>MFFV2 after using GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left vs. none</td>
<td>0.95</td>
<td>0.92</td>
<td>0.78</td>
<td>0.76</td>
<td>0.79</td>
<td>0.84</td>
</tr>
<tr>
<td>Right vs. none</td>
<td>0.95</td>
<td>0.94</td>
<td>0.93</td>
<td>0.80</td>
<td>0.78</td>
<td>0.88</td>
</tr>
<tr>
<td>Left vs. right</td>
<td>0.89</td>
<td>0.84</td>
<td>0.75</td>
<td>0.83</td>
<td>0.68</td>
<td>0.80</td>
</tr>
</tbody>
</table>
4. Discussion

The experiments related to Table 1 compare recognition rates using FFT between the 4-s window and the 1-s windows without and with active segment selection. In fact, we are mostly concerned with the difference in performance between the 4-s window and the 1-s window with active segment selection. The differences in average recognition rates are 1.0%, −0.2% and −2.2% for the cases of left finger lifting versus finger resting, right finger lifting versus finger resting, and left versus right finger lifting, respectively. Evidently, there are no significant differences among the recognition rates for the three cases. Thus, the 1-s window with active segment selection is obviously a better alternative to the original 4-s segment. Moreover, the comparison between the cases of the 1-s window both with and without active segment selection in Table 1 also shows the effectiveness of active segment selection. For the case of left finger lifting versus finger resting, the recognition rates for subjects S001 and S003 increase only by 3.3% and 0.8%, while those for subjects S002, S004 and S005 increase substantially by 17.5%, 13.3% and 17.5% respectively. The average recognition rate increase of 10.5% causes the average rate for the case of left finger lifting versus finger resting to rise to 75.7%. We will attempt to explain the performance differences among the five subjects from the information accumulation data in scale space shown in Fig. 5. In Fig. 5, C3 and C4 represent the channels C3 and C4, respectively. LN, RN, and LR represent the cases of left finger lifting versus finger resting, right finger lifting versus finger resting, and left versus right finger lifting, respectively. For every subject, the top two graphs show the cases of left finger lifting versus finger resting, right finger lifting versus finger resting, and left versus right finger lifting, respectively.

Fig. 5. Student’s two-sample t-statistics with short segmental power sum method. (a) For Subject S001. (b) For Subject S002. (c) For Subject S003. (d) For Subject S004. (e) For Subject S005.
lifting versus finger resting, the middle two show the case of right finger lifting versus finger resting, and the bottom two show the case of left versus right finger lifting. From the top two graphs in Fig. 5, subjects S001 and S003 have selected active segments close to the trigger region, which is at the center of the 4-s segment, while the other three cases have their selected segments considerably further away from the trigger region. Accordingly, the use of active segment selection is more effective for subjects S002, S004 and S005 than for subjects S001 and S003 in the case of left finger lifting versus finger resting. In the cases of right finger lifting versus finger resting, similar results are obtained and illustrated in the middle two graphs of Fig. 5. Subjects S001 and S003 have the active segments closer to the trigger and the resulting rates are slightly improved, while those of S002, S004 and S005 are further away and the corresponding rates are more significantly improved in Table 1. The average recognition rate increases by 8.3% for the case of right finger lifting versus finger resting, reaching 76.8%. Finally, the case of left versus right finger lifting is shown in the bottom two rows of Fig. 5. The selected active segment for subject S005 is very close to the trigger region, while those for subjects S001, S002, S003 and S004 are further away. Thus, after applying active segment selection, the recognition rate for subject S005 is not significantly improved, while the rate for subjects S001, S002, S003 and S004 increases substantially by 22.5%, 13.3%, 13.3% and 9.1%, respectively. Due to the average recognition rate for all subjects improving by 11.6%, the overall average recognition rate for left versus right finger lifting reaches 62.5%. From these experiments, active segment selection is proven an effective and stable procedure to improve the recognition rates for four out of the five subjects after one-time training for selecting the active segments. Only the subject S001 had an original average recognition rate too good to be significantly improved.

Table 2(a) and (b) show the cases adopting MFFV1 and MFFV2 as the features for the recognition tasks with and without applying the active segment selection. In comparison to the results using FFT with active segment selection in Table 1, the average recognition rates, by using MFFV1 for the three finger lifting experiments, increase by 4%, 9% and 7.5%, respectively. Similarly, using MFFV2 results in average recognition rates increasing by 5%, 8.5% and 7.3%, respectively. For the cases without active segment selection, the recognition rates are improved by 5.5%, 7.2% and 1.9% for MFFV1, and by 5.1%, 7.5% and 3.8% for MFFV2. Therefore, these tables clearly show that the proposed features can significantly improve the overall performance in movement recognition. In fact, if we compare the cases without segment selection in Table 1 with the cases with segment selection in Table 2(a), the overall performances with MFFV1 for the three different recognition cases increase by 14.5%, 17.3% and 19.1%, correspondingly. And in Table 2(b), the MFFV2 improves the performances by 15.5%, 16.8% and 18.9%, correspondingly. After using the proposed features together with optimal active segment selection, motion recognition is greatly improved for all subjects. In particular, subjects S001 and S002 become very reliable in achieving acceptable recognition rates (mostly better than 80%) for almost all recognition tasks. The improved ratios for subject S002 are all above 20% in all three recognition tasks. The other three subjects also obtain large improvements (more than 20% on average). However, most recognition rates are still under 80% since the original recognition rates were quite low.

Table 3(a) lists the comparison of recognition rates among FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 features selected by the GA. All features are used together with the optimal active segmentation selection in this table. In other words, the listed numbers demonstrate only the deviations of performance by applying different types of features. For the case of left finger lifting versus finger resting, the best average recognition rate (80.7%) is achieved using MFFV2 as the features, while the classification features using 6th order AR parameters obtain the worst average recognition rate (72.0%). The difference between the best and worst average recognition rates is 8.7%. Moreover, using MFFV1 as the classification features obtains the second best average recognition rate (79.7%), which is only 1.0% less than the best one. For the case of right finger lifting versus finger resting, the best average recognition rate (85.8%) is achieved using MFFV1 as the classification features, while the classification features using the frequency components of FFT obtain the worst average recognition rate (76.8%). The difference between the best and worst cases is 9.0%. On the other hand, MFFV2 obtains the second best average recognition rate (85.4%), which is slightly worse than MFFV1 (0.4% less). For the case of left versus right finger lifting, the best average recognition rate (70.0%) is achieved using MFFV1, while using FFT yields the worst average recognition rate (62.5%). The difference is 7.5%. Meanwhile, MFFV2 obtains the second best average recognition rate (69.8%), which is slightly worse than MFFV1 (0.2% less). Furthermore, from the four kinds of features we can find that all subjects have higher recognition rates for the case of right finger lifting versus finger resting than those for the case of left finger lifting versus finger resting, except for subject S001. This can be explained by the fact that subject S001 is left handed. In summary, both proposed fractal features MFFV1 and MFFV2 are good for motion classification. The difference between their average recognition rates is less than 1.0% for all cases. However, the computation cost for MFFV2 is more than that for MFFV1. This is because only separated subbands are evaluated for MFFV1, while fractal dimensions of all subbands need to be estimated for MFFV2. If we take the computation cost into account, MFFV1 is likely the better feature for motion classification among the four.

The fractal feature vector MFFV2 is formed by collecting the fractal dimensions of the optimal active segment and all its subbands. However, the difference of average recognition rates between MFFV2 and MFFV1, which is only the subset of MFFV2, is almost negligible. However, it is quite possible to achieve higher recognition results if we select more representative features from MFFV2. As the selected subbands depend in practice on each specific subject, a GA is used in this study to select the appropriate bands. In addition, the same data and five-fold crossvalidation procedure are used to train the parameters of both the GA and LDA. The last row of Table 3(a) lists the recognition rates after applying the GA to select features from MFFV2. For the case of left finger lifting versus resting,
For subject S004 while all the others have only slight contributions (or even degradations) to the classification. Although GA can significantly improve the recognition rates in some cases for most subjects, the influence may decay after a long duration of operations. This is because EEG signals are non-stationary and the characteristics may vary after some time. Because the number of selected fractal features is only seven, retraining with the GA from the updated data set is not particularly time consuming in our studies. However, the strategy for overcoming the inherently non-stationary nature of EEG signals is obviously a very interesting and challenging issue for future study.

Table 3(b) lists the recognition rates for the imaginary data acquired from public available database. Although the proposed method was designed for the recognition of real finger movement, we also test it to see its adaptability to imaginary data. The adopted procedures for active segment selection and feature extraction are the same as the ones in the above mentioned real movement recognition. The detailed processes and experimental design to adjust the proposed method for the imaginary cases are beyond the current scope of this paper. In these experiments, the recognition rates of FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 features selected by the GA are resulted from 280 trials for subject Im001 and 320 trials for both subject Im002 and subject Im003. As the provided data is only for the case of left versus right finger lifting, the experiments are carried out accordingly. In the resulting table, the recognition rates for subjects Im001 and Im003 are both better than 90% for both cases of MFFV1 and MFFV2. The recognition rate for subject Im002 is slightly worse than the other two subjects, however is still better than the rates by both FFT and AR methods. The average rate of the three subjects by using MFFV1 is 89.8% and by MFFV2 is 88.5%. They are significantly better than the one by FFT (82.9%) and the one by AR method (83.4). As the computation cost for MFFV2 is more than the one for MFFV1, MFFV2 is better in both accuracy and efficiency aspects. Similarly, if we use GA to select the most appropriate bands for each individual subject, the average recognition rate can be further improved by 3.9% as shown in the last row of Table 3(b). The improvement for subjects Im002 is especially obvious, the recognition rate increases by 7.3%. Although the real and imaginary data are slightly different in their formats, they are all aiming for movement classification with EEG signals. In this study, all the procedures adopted in imaginary finger movements are similar to those used in the real ones. More importantly, we adopted the same features for both real and imaginary movement classification. The experimental results reveal that the proposed feature extraction is more effective in comparing with the other methods. Because the proposed method provided very encouraging results in both overt and imaginary movements, it will help people who are newly hand injured and cannot have normal hand function, and it also has good potential to help people whose BCI purely depends on imaginary operations.

In addition to correct classification rate, the performances are also measured by the area under the ROC curve (AUC) to comprise both the sensitivity and specificity tests. Table 4 lists the comparison of AUC among using FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 features selected by
the GA. From the first four rows of Table 4, we observe that the best average AUC is the feature using either MFFV1 or MFFV2 for the three previously mentioned cases. The difference between the best and second average AUC is quite small. The largest difference between them among the three cases is only 0.02. In addition, the worst average AUC for the three cases is using FFT and/or 6th order AR parameters as features. The differences between the best and the worst average AUC over the three cases are 0.10, 0.10, and 0.17, respectively. It clearly reflects that the proposed feature vectors are superior to the conventional features in both recognition rates and average AUC. Moreover, after applying the GA to select features from MFFV2, the average AUC further increases by (0.11, 0.11, and 0.20) with respect to the worst average AUC using FFT and/or 6th order AR parameters. In summary, both proposed fractal features MFFV1 and MFFV2 are good features for motion classification. Since the computation cost for MFFV2 is undoubtedly more intensive, MFFV1 is likely the better feature among the four if the computation cost is taken into account.

To verify if it is significantly different under the condition “without and with (W/Wout) ASS”, the ANOVA and multiple comparison tests are performed for each pair of finger movement state on features FFT, MFFV1, and MFFV2. The \( p \)-values for the three feature types under three comparison state pairs are all less than 0.05. The \( p \)-values indicate that the active segment selection is capable of significantly increasing the classification accuracy. The tests on multiple comparison tests of means also indicate that the classification results (W/Wout ASS) are significantly different for all features under different pairs of finger movement states. The proposed active segment selection is not only effective in improving recognition accuracy but reduces computation to 1 s-epoch that is more practical for on-line applications. On the other hand, to confirm whether it is significantly different among features, the ANOVA and multiple comparison tests are also estimated for each pair of finger movement states among features FFT, 6th order AR parameters, MFFV1, MFFV2, and MFFV2 after using GA. The \( p \)-values are evaluated under the condition “among features” and in the cases “between each pair of features”. The \( p \)-values under the condition “among features” are all less than 0.001. This implies that significant difference exists among the five selected features. The more detailed comparisons of \( p \)-values between each pair of features and multiple comparison tests of means are then performed. The results indicate that the proposed features (MFFV1, MFFV2, and MFFV2 after using GA) achieve better classification accuracy, which is significantly different from the traditional ones (FFT and AR parameters) for finger movement experiments.

In the proposed method, the active segment selection helps to overcome the traditional constraint, long duration of ERD and ERS, and to facilitate on-line applications. Generally, the duration of ERD and ERS band is about 6–8 s and is too long to achieve on-line processing. We adopted 1 s window as the signal length for data analysis and feature selection. The experimental results comparing the classification accuracy between 4 s windows and 1 s window with active segment selection were given in Table 1 which strongly supported our idea with significant improvements. The active segment selection restricts the processing time to 1 s and enhances the applicability of on-line processing. In addition to the active segment selection, we also proposed to extracted the multi-scale fractal feature from discrete wavelet transform data, the new feature sets can be obtained in less than 0.1 s for a 1 s window and achieve more reliable recognition rate in comparison with the conventional ones. Therefore, we did bring out a new method to extract effective motion features in a short-time (1 s) window that can afford EEG motion recognition in almost real-time applications.

For better recognition of EEG signals, we then proposed the multi-resolution fractal features, which are obtained with fractal dimension (FD) from DWT data. To speed up on-line processing, DWT is applied and yields more economical computation. We adopted the Daubechies wavelets for both CWT and DWT. The DWT provides a similar representation and faster computation and the resulting FD features are reliable for on-line classification. The present multi-resolution fractal features combine the good characteristics of DWT and FD that we can obtain the multi-resolution features with self-similarity of fractal geometry from various subbands decomposed by DWT. The experimental results comparing the proposed fractal features (MFFV1, MFFV2, and MFFV2 after using GA) with the other popular features (FFT and AR parameters) were shown in Table 3. It indicates that the fractal features obtain better recognition rates than the FFT features and AR parameters for both real and imaginary EEG motion classification. The proposed method to extract multi-resolution fractal features from DWT data for the recognition of EEG motion signal is a new idea that helps the BCI user to analyze the EEG signal from a new point of view. The effective features that combine the multiscale nature of wavelet transform with the character of fractal descriptor may find better usages to other EEG recognition tasks in similar applications.

The proposed active segment selection was applied to the new and conventional features, experimental results showed significant improvements on all features. The other experiment showed the recognition performance of five different features after applying active segment selection, and the proposed fractal features also achieved significantly better improvement. These observations implied that the combination of active segment selection and multiscale fractal features is a good feature extraction tool in identifying both the real and imaginary finger movements.

After feature extraction in these experiments, LDA is employed as the only classifier for motion classification because LDA is simple and fast. Its simplicity also makes it less possible to confuse the effects of either active segment selection or feature extraction, which are the main themes of the proposed method. If more effective (or sophisticated) classifiers, for example AdaBoost or Support Vector Machines are employed, the recognition rates may be further improved.

5. Conclusion

In this paper, we have proposed an effective EEG analysis system for the single-trial classification of motion evoked by voluntary, self-paced left versus right finger lifting. In our system, new feature extraction schemes together with optimal
active segment selection were proposed and applied to the EEG signals acquired from the sensorimotor cortices. As seen from the experimental results, the proposed active segment selection scheme can effectively obtain the optimal active segment by taking the time–frequency domain information into account to significantly increase the classification accuracy. Furthermore, the proposed fractal features, MFFV1 and MFFV2, are also very useful in improving motion classification. They both obtain much better classification results than the other two conventional features, FFT and 6th order AR parameters. The difference between the average recognition rates of these two fractal features is small for most cases. However, as the computation cost for MFFV1 is less expensive, we consider that MFFV1 is likely the better feature among the four selected for motion classification in our experiments. After a GA is applied to pick the dominant features from the original feature set of each subject, we can obtain excellent classification accuracy in all single-trial finger movement experiments by using only simple LDA as the classifier. These selected features usually work for a long period of time but need to be updated after several experiments or a change in environment. However, our system is generally well-suited for real-time analysis because of its high recognition rate and low computation cost. The selected active segment is only a 1-s window and the number of selected fractal features is small. However, the operations of active segment selection and the application of the GA are only needed in the training phase, while we only extract feature vectors and classify them during the online analysis. The proposed active segment selection and fractal features are indeed a good combination for effective single-trial EEG motion classification. The proposed system has also been tested with imaginary finger movement data and achieves good experimental results. Therefore the proposed system renders very good potential for BCI applications.

Acknowledgement

This research was partially supported by grants from the National Science Council, (NSC91-2213-E-006-047, NSC92-2218-E-006-061), Taiwan, ROC, and is gratefully acknowledged.

References


Daubechies I. Ten lectures on wavelets. CBMS-NSF Lecture Notes nr.61, SIAM, 1992.


The document contains a list of references cited in a scientific paper. The references cover various topics related to brain-computer interfaces and EEG-based communication. The references include works on linear spatial integration, texture analysis, EEG-based communication, and various applications of EEG in communication and control. The authors of the paper are W.-Y. Hsu et al., and the paper is published in the Journal of Neuroscience Methods.