A Study and Performance Analysis of Three Paradigms of Wavelet Coefficients Combinations in Three-class motor imagery based BCI

Ayad G. Baziyad, and Ridha Djemal
Department of Electrical Engineering, King Saud University
ayadgaafar@gmail.com, rdjemal@ksu.edu.sa

Abstract—Brain computer interface (BCI) provides an interface between a brain and a computer in order to enable people to control external devices without using muscles. In this work, authors report on the results of implementation of three algorithms using wavelet features collected with different kinds of features during imagining left hand, right hand, and foot movements. The features of event-related desynchronization (ERD/ERS) were extracted from alpha and beta frequency bands, and followed by one classifier among the three following ones: linear discriminant analysis (LDA), support vector machine (SVM) or K-nearest neighbor (KNN). The data were recorded from three subjects, provided by BCI-Competition III. The performance evaluation of the proposed algorithms was provided by Matlab simulation. The best combination was the wavelet coefficients and common spatial pattern algorithms, followed by the support vector machine classifier with an average classification accuracy of 75%, which is an interesting for motor imagery application.

Keywords—Brain-computer interface (BCI); motor imagery; Wavelet; EEG.

I. INTRODUCTION

Human relationships with external devices have been continuously improving according to the urgent need to find solutions especially for individuals who suffer from severe physical disabilities to improve their life. Over the last two decades, many studies have shown that the electric and magnetic fields generated during brain activities can be measured and analyzed to produce certain signals. Such signals, commonly called an electroencephalograph (EEG), can be recorded to be an input to an interface system using brain scalp [1].

A brain-computer interface (BCI) is a system that provides an alternative communication channel between a brain and a computer, whose task is to translate the EEG signals into appropriate control commands through the signal processing algorithm in order to operate a variety of applications such open a door, activate an AC, etc. This function involves the classification of different brain states with a high accuracy requirement which is the main goal in BCI research. Based on the properties of the EEG signals which are extremely difficult to understand, previous studies have proposed several methods to extract the most different features depending on data processing techniques to efficiently reduce and discriminate features set [2].

Previous studies have shown that the behavior of EEG signal can be described by using frequency analysis. This approach is called EEG spectrum which describes the characteristics of the waveforms within four frequency bands: delta (0-4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). The power spectral analysis provides a quantitative measurement of the frequency distribution of the EEG. Previous studies have also shown that amplitudes of alpha and beta rhythms are changed when people imagine movements [3]. This phenomenon is called Event related desynchronous/synchronous (ERD/ERS) [4]. This technique has been studied to discriminate multiclass motor imagery EEG signals. Several publications have appeared in recent years and showed more than 85% in 2-class BCI [5], [6] and [7].

The literature on three-class motor-imagery-based BCI shows a variety of approaches have been introduced by different methods of the feature extraction and classification. In [2] and [8], band power features have been extracted, and the LVQ NN and HMM were used as classifiers. In [9], the author studied improving the spelling rate of EEG-Based Virtual Keyboard presented a good classification accuracy, which band power was extracted by squaring and averaging the signals and the combination of three LDA was used as a classifier.

A combination of different features has been proposed in [10] based on wavelet packet decomposition combined with common spatial pattern. Some of previous studies use only three channels (C3, Cz and C4) such that information of motor imagery may be missed when using data sets recorded from many subjects.

In this paper, we report on the offline analysis of three-class motor imagery based BCI experiments. The aim of this paper is to recommend a suitable combination of wavelet features for synchronous three-class motor-imagery-based brain-computer interface experiments. Different measures are applied for analyzing, extracting and reducing characteristics of EEG data in time, frequency or spatial domain. The wavelet coefficients are extracted and combined with three kinds of different features each separately. The actual features is taken as inputs to a classifier (LDA, SVM or K-NN) to be classified into one of three classes (Left, Right or Foot). We chose autoregressive parameters (AR), common spatial pattern (CSP) and band power (BP) distributions of ERD/ERS. Classification Accuracy was used to evaluate performance of each approach.

The remainder of the paper is organized into three sections: Section II describes the EEG data set used in our work, and the feature extraction and classification methods. Experimental results are presented and discussed in Section III; Section IV concludes the paper.
II. METHODS

A. Dataset

The dataset (IIIa) of BCI Competition III were used in all our experiments. These data were provided by Graz University of Technology. The data were recorded from three right handed subjects whose ages between 22 and 34. The recording was made during a BCI experiment consists of (from 6 to 9) runs per subject during a single session. The recording was done with a 64-channel EEG amplifier from Neuroscan. Sixty channels were referenced to the left mastoid with ground at the right mastoid, as shown in Figure 1. The EEG was sampled with 250 Hz, band-pass filtered between 1 and 50, and notch filtered at 50 Hz to suppress line noise.

Figure 2 shows the timing of the experiment. The task was the imagination of left hand movement, right hand movement, food movement, or tongue movement according to a random cue. At t=2s, a cross “+” is displayed at a blank screen; then at t=3s, an arrow is displayed to either left, right, downwards or upwards.

![Figure 1](image1.png)  The positions of electrodes of the dataset. Only 15 electrodes were used in our experiments (colored by yellow and orange); however, all the electrodes were used in CSP paradigm

![Figure 2](image2.png)  Timing Scheme

B. Feature Considerations

Although the power of alpha and beta frequency bands of training data set can be discriminated during motor imagery at the channels (C3, Cz and C4) [9], it is not balanced and can be changed depending on the subject. This problem can be solved by means of a surface Laplacian filter [11]. The signal from C3, Cz, and C4 are multiplied by 4 and subtracted by all surrounding four signals. Here, we use this filter except when extracting common spatial patterns because CSP requires a large number of channels to get better results [12]. Since our study is about the synchronous BCI. A 3.2 second time window with 0.4 offset was taken of the trial during motor imagery to extract the most features and avoid the high dimensionality which may lead to bad results.

C. Feature Extraction

Since the EEG signals are non-stationary signals, Fourier Transform is not suitable to extract features from such signals. Short Time Fourier Transform (STFT) can be used to analyze non-stationary signals, but it gives a constant resolution at all frequencies. The Wavelet Transform uses multi-resolution technique to analyze different frequencies with different resolutions. In addition, it can allow a smaller number of features for representing of the signal; this implies that it may be useful relatively for overcoming of the curse-dimensionality problem. Basically, wavelet transform analyzes the characteristics of signal in time and frequency domain by decomposing such signal into a number of functions by a single function to make shifting and detailing [13] and [14]. This function is called mother function and given by

\[ \psi(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R}, a > 0 \]  (1)

Where \( R \) is the wavelet space and ‘a’ and ‘b’ are the scaling parameter and shifting parameter, respectively. The wavelet transform is defined by

\[ F(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \]  (2)

We used Discrete wavelet transform (DWT) because it gives highly efficient wavelet representation [15]. In DWT, low pass and high pass filters are repeatedly applied to obtain the representation of digital signal as approximation and detail coefficients.

![Figure 3](image3.png)  Three-Level Discrete Wavelet Decomposition Transform

The equation that describes DWT decomposition is as follows:
Figure 4. Feature Extraction Methods with Classifiers: (a) AR Paradigm (b) Band Power Paradigm (c) Common Spatial Pattern Paradigm

\[ x(t) = \sum_{k=-\infty}^{\infty} c_{n,k} \theta(2^{-n} t - k) + \sum_{k=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} 2^{-j/2} \psi(z^{-j} t - k) \]  

Where \( n \) is the level, \( \theta \) is the function of scale, and \( c_{n,k} \) and \( d_{j,k} \) represent the approximation and detail coefficients, respectively.

In this work, we chose Daubechies wavelet function of order 8 (db8) because it has optimal properties of signals features which may be classified successfully. Since we work on mu and beta bands, we applied three-level wavelet decomposition. The so-called sub-bands coding algorithm of three-level wavelet decomposition is schematically shown in Figure 3.

For reducing high dimensionality, a set of statistical operations was applied; the variance, standard deviation, and energy of coefficients.

Six features were obtained for each channel. As previously mentioned, 15 channels were spatially filtered to produce 3 channels hold the most features of motor imagery movements. Subsequently, 18 coefficients were obtained as a result of wavelet decomposition. These coefficients were collected and combined with other features were extracted by AutoRegressive, Common Spatial Pattern or Band Power algorithm according to the algorithm used, as can be seen from Figure 4 in the following, a brief description of each algorithms.

1. AutoRegressive Combination

AutoRegressive coefficients were computed by a linear combination related to an input and an output of the AR filter. The AR model [16] can be defined by the following equation:

\[ x(n) = \sum_{i=1}^{p} a_p(i)x(n - 1) + \varepsilon(n) \]  

Where \( x(n) \) is the input signal at point \( n \), \( a_p(i) \) are the coefficients of AR, \( \varepsilon(n) \) is a zero mean white noise and \( p \) is the order of the autoregressive model. We set the order of AR model to 6 depending on recommendations in [17]. Burg’s method [18] was used to estimate accurate AR coefficients and obtain better resolution. This method minimizes forward and backward error to avoid the problem of spectral leakage.

Eighteen wavelet coefficients and six AR coefficients were collected, giving 24 features for a single motor imagery task. These features were taken as inputs to a classifier.

2. Band Power Combination

Similarly, the wavelet coefficients were obtained by the same previously mentioned method. The features combined with wavelet were extracted as follows. After spatial filtering, the samples of signals were bandpass filtered (8-32) and processed by a commonly used method for representing band power [17]. In this method, the samples were squared, averaged and then the logarithm of value was applied. Subsequently, three BP features were obtained for each trial. As a final feature extraction step, 21 features of wavelet and band power were introduced into the classifier.

3. Common Spatial Pattern Combination

In this approach, we used all channels without Laplacian filter. The same EEG signals used by AR and band power algorithms were also used here. The EEG signal was bandpass filtered into six sub-bands from 8 to 32 Hz (8-12, ..., 28-32); then, CSP [19]-[20] was applied. The CSP has high efficiency for detecting ERD effects, which maximizes the variance for one condition while minimizing the variance for the other. The spatial filters of CSP are constructed by
computing the projection matrix $W$ maximizing the ratio of variance between the two tasks, as in

$$W = M^TP$$  \hspace{1cm} (5)

Finally, the original EEG signal can be spatially filtered by means

$$Z = WX$$  \hspace{1cm} (6)

Different types of CSP have been used for multiclass BCI in the previous studies. Since we have three classes, one-versus-one was used to solve our problem, so that three CSP filters were produced. The most spatial patterns were derived from the first and the last columns of the spatial filters, where they have the largest variance of one class and the smallest variance of the other. Then, the actual CSP features were produced by normalizing the variances. The feature vector of each sub-band was determined as

$$f = [f_1 \ f_1 \ \ldots \ f_n]$$  \hspace{1cm} (7)

$$f_i = \log \left( \frac{\text{var}(zp)}{\sum_{j=1}^{n} \text{var}(zp)} \right)$$  \hspace{1cm} (8)

Where $zp$ represents the selected column of spatial pattern matrix.

D. Classification

Three kinds of classifier were used to classify the features. We chose Linear Discriminant Analysis (LDA) [21] and [21], Support Vector Machine (SVM) [22], and K-Nearest Neighbor (K-NN) [21]. For solving the three-class problem of each kind of classifier, three combinations of two classifiers were used. Each of them assigns the feature vector to class, and the final class was voted by the majority [23]. This is illustrated in Table I. This way is the most popular of combining classifier [24], [25] and [26].

<table>
<thead>
<tr>
<th>Features</th>
<th>Classifier</th>
<th>K3</th>
<th>K6</th>
<th>L1</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelet+AR</td>
<td>LDA</td>
<td>86.66</td>
<td>40</td>
<td>50</td>
<td>58.88</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>86.66</td>
<td>30</td>
<td>60</td>
<td>58.88</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>80</td>
<td>50</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Wavelet+BP</td>
<td>LDA</td>
<td>86.66</td>
<td>50</td>
<td>60</td>
<td>65.55</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>93.33</td>
<td>50</td>
<td>70</td>
<td>71.11</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>86.66</td>
<td>50</td>
<td>50</td>
<td>68.88</td>
</tr>
<tr>
<td>Wavelet+CS</td>
<td>LDA</td>
<td>93.33</td>
<td>70</td>
<td>60</td>
<td>74.44</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>86.66</td>
<td>70</td>
<td>70</td>
<td>75.55</td>
</tr>
<tr>
<td></td>
<td>KNN</td>
<td>86.66</td>
<td>50</td>
<td>50</td>
<td>68.88</td>
</tr>
</tbody>
</table>

Classification accuracy was calculated to evaluate the performance by the following equation:

$$\text{Accuracy} = \left( \frac{N_{\text{correct}}}{N_{\text{total}}} \right) \times 100\%$$  \hspace{1cm} (9)

Where $N_{\text{total}}$ is the number of overall samples to be classified, and $N_{\text{correct}}$ is the number of correct classified samples. The overall results are summarized in Table II. The results lists the classification results of all approaches proposed in our study. It has been found that the high average classification accuracy derived from three subjects was 75.5% with the combination of wavelet decomposition and common spatial pattern. We think that the main cause is the spatial filter of all electrodes. For subject K6 and L1, no significant difference existed between the results and they have an accuracy much lower than the third subject. That is probably because the third subject did more training in previous studies. The time segment used is of 3.6 to 6.0s, which extracted the most features and improved the results. The SVM produced better results than other classifiers; K-NN and LDA.

IV. CONCLUSION

In this work, an offline three-class motor imagery based BCI has been implemented. Three classification algorithms were used to extract optimal features as well as three different...
kinds of classifiers. The significant performance gain was achieved by using common spatial pattern algorithm with SVM classifier. In future work, we will record a new data from subjects and online training and test will be studied in three-class motor imagery based BCI using FPGA.

ACKNOWLEDGMENT
This work reported in this paper is supported by the National Plan for Science and Technology (NPST) at the King Saud University (Project Number: ELE1730).

REFERENCES