International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) 2014

Motor Imagery EEG Signal Classification Scheme Based on Wavelet Domain Statistical Features

S.M. Imran, M.T.F. Talukdar, S. K. Sakib, N. S. Pathan, and S. A. Fattah[1]
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[1]fattah@eee.buet.ac.bd

Abstract—**Classification of electroencephalogram (EEG) data for different motor imagery (MI) tasks is a major concern in the brain-computer interface (BCI) applications. In this paper, an efficient feature extraction scheme is proposed based on the discrete wavelet transform (DWT) of the EEG signal. The EEG data of each channel is windowed into several frames and DWT is performed on each frame of data. Considering only the approximate DWT coefficients, a set of statistical features are extracted, namely wavelet domain energy, entropy, variance, and maximum. In order to reduce the dimension of the proposed feature vector, which is composed of average statistical feature values of all channels, principal component analysis (PCA) is employed. For the purpose of classification, k nearest neighbor (KNN) classifier is employed. Proposed classification scheme not only offers significant reduction in feature dimensionality but also provides satisfactory classification accuracy. For the purpose of performance analysis, publicly available MI dataset IVa of BCI Competition-III is used and a very satisfactory performance is obtained in classifying the MI data in two classes, namely right hand and right foot MI tasks.**

***Index Terms*—BCI, Classification, Discrete Wavelet Transform, EEG, Feature Extraction, KNN, Motor Imagery.**

III. INTRODUCTION

Electroencephalography (EEG) has become the most popular and widely studied non-invasive brain-computer interface (BCI) due to its excellent temporal resolution, usability and low setup costs. EEG-based BCIs use electrical brain activity in assisting, augmenting, and repairing human cognitive functions. The greatest challenge in BCI systems is to correctly and efficiently identify different EEG signals of different motor imagery (MI) tasks. Recently, MI EEG signal classification has received much attention by several researchers [1]-[4]. With the ease of availability of high quality low cost EEG leads, one major concern now-a-days is the huge amount of data to be handled in case of multi-lead EEG signal analysis. Different spatial filters can be used for EEG channel reduction, where a major drawback is the manual selection of regularized parameter. For example, sparse spatial filter optimization method, introduced in [1], can offer EEG channel reduction at the expense of manual parameter selection. Similarly, the regularized common spatial pattern (CSP) algorithm, reported in [2], involves two regularization parameters which are not selected optimally. Four different approaches for MI task classification using EEG signals are proposed in [3] based on CSP which itself is very sensitive to noise and often over-fits with small training sets. The regularization and aggregation techniques introduced in [4] solve the parameter determination problem by cross-validation method but only in a small sample setting. In [5], cross-correlation based statistical feature extraction method is developed using the least-square support vector machine (LS-SVM) classifier for MI task classification, where a prior knowledge of the classes is utilized. A probabilistic framework to search for optimal frequency band that can maximally discriminate the feature vectors of two classes is reported in [6], which provides low average classification accuracy and deserves further investigation. Works have also been done on time frequency domain to classify imagination activity, but either the methodology has been tested on different datasets or different features have been extracted for classification.

The objective of this paper is to develop a robust classification scheme based on the discrete wavelet transform (DWT) of the EEG signal corresponding to different MI tasks. The EEG data of each channel is preprocessed, windowed into several frames, and then the DWT operation is performed on each frame of data. A set of statistical features of the approximate DWT coefficients is extracted, namely wavelet domain energy, entropy, variance, and maximum. Principal component analysis (PCA) is used to reduce the feature dimension of the proposed feature vector consisting of the average statistical feature values of all channels. For the purpose of classification, K nearest neighborhood (KNN) classifier is used. Detail simulation results are presented considering a publicly available MI EEG dataset.

IV. DATA ACQUISITION

In this paper, most widely used publicly available EEG dataset IVa of BCI Competition III is considered for the purpose of classification [7]. This dataset was recorded from five healthy subjects labeled aa, al, av, aw, and ay. In this dataset, EEG signals from 118 electrodes were collected based on the international 10/20 system. Fig. 1 depicts 118 electrode locations that were placed to collect the EEG data. There were 280 trials for each subject, namely 140 trials for each task per subject. During each trial, each subject was required to perform either of the two (right hand and right foot for IVa) MI tasks for 3.5 seconds. The down-sampled data at 100 Hz is generally

978-1-4799-4819-2/14/$31.00 ©2014 IEEE

used although the original sampling rate was 1000 Hz. Duration of the recording of the MI task is 3.5 seconds.
The down-sampled data at 100 Hz were preprocessed and band-pass filtered between .005 Hz and 200 Hz.

V. Proposed Method

A. Discrete Wavelet Based Feature Extraction

The discrete Wavelet transform (DWT) is a multi-resolution technique that offers localization both in time and frequency [8]. DWT exhibits good frequency resolution at low frequencies and good time resolution at high frequencies. Another important advantage of the DWT is its low computational cost and ease of implementation. Hence the DWT is chosen to extract features from the EEG signal.

The DWT coefficients of a signal $x(n)$ can be obtained as

$$C(a,b) = \sum_{n \in Z} x[n] \psi_{a,b}[n] \qquad (1)$$

Where a is the dilation or scale, b the translation, and $\psi_{a,b}[n]$ represents the discrete wavelet which is expressed as

$$\Psi_{a,b}(n) = \left(\frac{1}{\sqrt{a}}\right) \times \Psi\left(\frac{n-b}{a}\right) \qquad (2)$$

For dyadic wavelet transform, $a = 2^{-j}$, $b = k \times 2^{-j}$, $\psi_{a,b}[n] = 2^{j/2} \times \psi[2^j n - k]$ with $k \in Z, j \in N$. The DWT operation can be viewed as passing the signal $x[n]$ through two complimentary filters simultaneously, a low-pass filter with impulse response $g[n]$ and a high-pass filter with impulse response $h[n]$ followed by downsampling, which can be represented by the following two equations

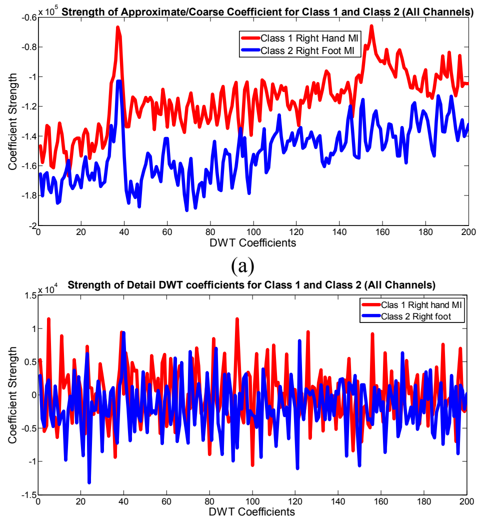
$$y_g[n] = \sum_{k=-\infty}^{k=\infty} x[k].\, g[2n-k] \qquad (3)$$

$$y_h[n] = \sum_{k=-\infty}^{k=\infty} x[k].\, h[2n-k] \qquad (4)$$

The DWT coefficients obtained from $y_h[n]$ and the low-pass filter $y_g[n]$ are termed as detail and approximate coefficients, respectively. The filtering operations in the DWT result in a change in the signal resolution, whereas the sub-sampling (down sampling/up sampling) causes change in the scale. Thus, DWT helps in analyzing the signal at different frequency bands with different resolutions. For the MI EEE signals, the low-frequency content is found to be significant part than the high-frequency content (Fig 1). As can be seen in the figure, the approximation coefficients of two MI classes seem to vary considerably than the detail coefficients of these classes. The detail coefficients show significant overlap between the two classes whereas the approximation coefficients show minimum overlapping. Hence, the approximate coefficients are taken in consideration in the proposed method.

A large number of wavelet functions are available in the literature. The wavelet families are Daubechies, Symlets, Coiflet, Biorsplines, ReverseBior, and Discrete Meyer. Amongst them, since the Daubechies family shows good performance in EEG signal, we use the Haar wavelets of the Daubechies family for feature extraction.

The EEG data of a particular person is windowed into ten frames and the single stage DWT is performed on each frame of a channel of a trial. Instead of utilizing all the DWT coeffici-



(b)

Fig 1. DWT coefficients obtained for two different classes. (a) Approximate DWT coefficients and (b) detail DWT coefficients. different classes

nts obtained in a frame, different statistical features are investigated. However, following four statistical features are selected for feature extraction:

- Maximum value of wavelet coefficients
- Variance of wavelet coefficients
- Wavelet domain energy
- Wavelet domain entropy

First three statistical features are most widely used in different applications. However, the fourth one is introduced in the proposed feature as it offers better feature quality. The concept of entropy is very popular in information theory. Entropy of a discrete random variable X with $(N+1)$ number of possible values $\{x_0, x_1, x_2, \ldots, x_N\}$ is defined as

$$H(X) = E(I(X)), \qquad (5)$$

where $E(.)$ denotes the expectation operator and $I(X)$ represents the information content. For a particular value x_i of the random variable X, the information content can be expressed as

$$I(X) = -\log_2(p(x_i)), \qquad (6)$$

where $p(x_i)$ denotes the probability of occurrence of a particular value x_i. Using (2), the entropy in (1) can be rewritten as

$$H(X) = -\sum_{i=0}^{N} p(x_i) \times \log_2(p(x_i)). \qquad (7)$$

For the case of DWT coefficients obtained from a frame of EEG data, being normalized, the value of each coefficient varies within the range of 0 to 1. Considering M number of bins in the range of 0 to 1, the possibility of occurrence of a particular coefficient value is still random and varies depending on the pattern of wavelet coefficients as well as the given EEG data. The probability of occurrence of a particular coefficient value x_i would be $p(x_i)=n_i /N$, where n_i is the number of occurrence of coefficients in a bin in proximity of x_i value among the N number of values, i.e. $\sum_i n_i=N$.

The variation of these features at ten different windows for all the channels is shown for two classes in Fig. 2. These parameters have been calculated by taking the sum of features of all channels for all the trials of subject aa for both the classes. As can be seen, in Fig. 2 (a), feature 1(maximum) of two classes is different at different frames of the event. At the early frames, class 1 max values are greater than class 2 max values. At some of the frames, the difference is almost nil, so these turn out to be redundant values in the feature matrix. For Fig. 2 (b), feature 2 (variance) of class 2 is always greater or equal to class 1 except for window 5, where class 1 task values seem to be dominating class 2 task values. For Fig. 2 (c) and (d) (energy and entropy), class 1 task values are always lower than class 2 task values which seem to be a discriminating information for class seperability. Note that in Fig. 2 (c), the energy in both the classes are increasing. This happens since the intensity of motor imagination increases with time, the subject being at the most relaxed state initially and at the most excited state just prior to end of the experiment. Since the waveform of Fig. 1 already shows that the two classes have good difference in their mean values, we take the mean of each statistical feature of ten windows to form the feature set for each class.

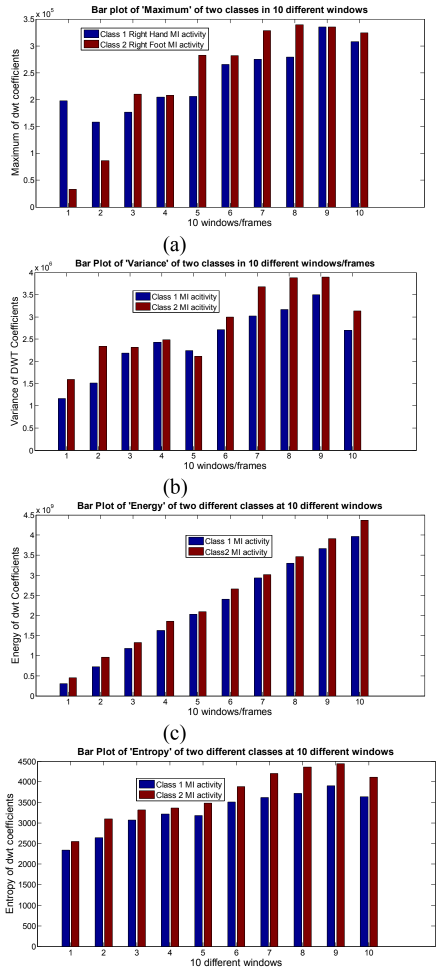
It is believed that these features reduce feature size as well as it preserves all important information of the original signal patterns. As a result, for each subject, a feature matrix is constructed based on four features from each approximate sequence.

B. Dimensionality Reduction Using PCA

In the next step, principal component analysis is performed on our feature matrix as one step for dimensionality reduction. PCA is a very well known and efficient orthogonal linear transformation [9]. It reduces dimension of the feature space and the correlation among the feature vectors by projecting the original feature space into a smaller subspace through a transformation. PCA transforms the original p-dimensional feature vector into the L-dimensional linear subspace spanned by the leading eigenvectors of the covariance matrix of feature vector in each cluster ($L<p$). PCA is theoretically the optimum transform for given data in the least square sense. For a data matrix $\mathbf{X}^{T}$ with zero empirical mean, PCA transformation is given by

$$\mathbf{Y}^{T}=\mathbf{X}^{T}\mathbf{W}=\mathbf{V}\mathbf{\Sigma}^{T} \tag{8}$$

where the matrix $\mathbf{\Sigma}$ is a $m\times n$ diagonal matrix with nonnegative real numbers on the diagonal and $\mathbf{W\Sigma V}^{T}$ is the singular value decomposition of $\mathbf{X}$. For the proposed algorithm,



(a) (b) (c) (d)

Fig 2. Variation of proposed features at different windows of a trial (for two classes) (a) Maximum, (b) variance, (c) energy, and (d) entropy .

implementation of PCA on the derived feature space could efficiently reduce the feature dimension without losing much information. Hence PCA is employed to reduce the dimension of the proposed feature space.

C. Classification

The k-nearest neighborhood algorithm (KNN) is one of the most reliable but simple method of classifying objects based on closest training examples in the feature space. KNN is a type of instance-based learning or lazy learning where the function is only approximated locally and all computations are deferred until the classification. In this paper, for the classification of the EEG data into two classes based on the dwt features, the KNN classifier is employed.

D. Training and Testing

We take leave-1-out trial for testing and the remaining trials for training. We perform the operation 280 times and update the accuracy of the classifier performance accordingly for each subject. It should be noted that our dataset comprises of the complete dataset (both training and testing as set exclusively

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT METHODS FOR BCI COMPETITION III DATASET IVA

Method	Classification Accuracy rate (%)					
	aa	**Al**	**av**	**Aw**	**ay**	***Average***
Proposed	80.67	80.21	76.45	79.35	74.64	78.26
CSP [10]	66.96	89.29	52.25	47.77	52.38	61.79
CSSP [10]	79.46	92.86	52.55	91.52	51.59	73.60
BSSFO [10]	79.46	94.64	57.65	91.96	53.57	75.46
FBCSP [10]	69.64	80.36	47.96	55.36	48.41	60.35
DCSP [10]	69.64	82.14	54.08	50.89	48.41	61.03
SSRCSP [1]	70.54	96.43	53.57	71.88	75.39	73.56
TRCSP [1]	71.43	96.43	63.27	71.88	86.9	77.98
WTRCSP [1]	69.64	98.21	54.59	71.88	85.32	75.93

by the dataset website) for each subject.

VI. SIMULATION AND RESULTS

Sample values are first imported into EEGLAB GUI for preprocessing. A reference electrode TP10 is chosen post hoc during data import. Without a reference electrode, it is possible that minimum level of 40 dB noise is accumulated with our channel information [11]. Typical recording references in the EEG recording are TP10, Cz or channel on the nose tip. Since there is no 'best' common reference site, we choose TP10 as our reference electrode. EEGLAB was also used to separate frames of class 1 and class 2 samples from the jumbled dataset. After completely preprocessed data, we divide each channel data into ten frames and find the approximate coefficients of single stage dwt of each frame, and go on to extract the maximum, variance, energy, and entropy sets from each approximate coefficient sequences. In that way, we need not need prior knowledge of the class while calculating the approximate sequence of the single stage DWT. After extracting the feature set, feature reduction is achieved by means of principal component analysis. The original feature matrix is then projected to the new optimum domain. Finally this matrix is used to classify the two classes by means of kNN. In order to examine the accuracy of the proposed methodology, we compare our approach with other recently reported techniques. We used both the training and testing sets of the dataset IVa (as set by BCI III) containing right hand and right foot motor imagery tasks to make a complete dataset and then perform feature extraction. Table I reports the comparison results of the classification accuracy rates for the proposed method and the six algorithms for dataset IVa [7]. This table presents the separate classification performance for the five subjects as well as the overall mean accuracy. The comparison works are reported in Yong et al. [1] and Suk et al. [6]. From Table I, it is noted that our proposed methodology provides competitive average classification accuracy compared to all the reported algorithms. The results indicate that our methodology improves the classification performance significantly even after obviation of prior knowledge of classes.

VII. CONCLUSION

High classification accuracy is desirable in MI-based BCI systems. The proposed MI task classification scheme is based on single stage discrete wavelet transform, feature extraction, PCA and KNN classifier. The classification accuracy is obtained from the five individual subjects' data from the BCI competition III database. Instead of using all wavelet coefficients from all frames, use of wavelet domain statistical features not only reduces the computational burden but also provide satisfactory discriminant features. As a result, the proposed scheme offers a very satisfactory classification performance for all subjects in comparison to some of the existing methods. Moreover, use of PCA provides significant control on feature dimension reduction as well as performance enhancement. Because of the quality of the extracted features, even simple KNN classifier can provide satisfactory classification performance.

REFERENCES

[1] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in *Proc. IEEE Int. Conf. ICASSP* Mar 30-Apr 4, 2008.

[2] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. IEEE 31st Annu. Int. Conf. EMBS*, Sep. 2-6, 2009.

[3] F. Lotte, and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355-362, Feb. 2011.

[4] H. Lu, H. L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.* vol. 57, no. 12, pp. 2936-3945, Dec. 2010.

[5] S. Siuly and Y. Li, "Improving the separability of Motor Imagery EEG signals using a cross correlation-based least square support vector machine for brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 526–538, Jul. 2012.

[6] H. Suk and S.-W. Lee, "A Novel Bayesian Framework for Discriminative Feature Extraction in Brain Computer Interfaces", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, Feb. 2013.

[7] BCI Competition III [Online]. Available: http://www.bbci.de/ competition/iii B.

[8] K. Englehart, B. Hudgins, A. Philip, "A wavelet based continuous classification scheme for multi-function myoelectric control", *IEEE Trans. On Biomed. Eng.*, vol. 48, no.3, pp. 302-311, 2001.

[9] I.T. Jolloffe, "Principal component analysis," *Springer Verlag*, Berlin, 1986.

[10] B. Blankertz, K.R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw *et al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Rehabil. Eng.*, vol. 14, no. 2, pp. 153-159, Jun. 2006.

[11] A. Delorme, T. Fernsler, H. Serby, and S. Makeig, "EEGLAB Tutorial," vol. 601b, pp.25, April 12, 2006. Copyright University of San Diego California.