Motor Imagery based BCI Classification via Sparse Representation of EEG Signals

Younghak Shin, Seungchan Lee, Minkyu Ahn, Sung Chan Jun and Heung-No Lee*
School of Information and Communications
Gwangju Institute of Science and Technology
Gwangju, Republic of Korea
{shinyh, seungchan, frerap, scjun, heungno}@gist.ac.kr

Abstract—**Electroencephalogram (EEG) based brain-computer interface (BCI) provides a new communication and control channel for people with severe motor disabilities. Motor imagery based sensorimotor rhythm (SMR) analysis is one of the widely used methods in the BCI field. However, these motor imagery signals are very noisy and strongly depends on subjects. Therefore, it is difficult to classify them and thus more powerful classification methods are needed. In this paper, we propose a new classification method based on sparse representation of EEG signals and ell-1 minimization. Using Mu and/or Beta rhythm as EEG features, we evaluate the performance of the proposed method with four data sets. Moreover, we make performance comparison with the linear discriminant analysis (LDA), another classification method. From the results, our proposed method shows the better classification accuracy.**

Keywords- ***Electroencephalogram (EEG), Brain-Computer Interface (BCI), Sensorimotor Rhythm (SMR), Sparse Representation, Compressed Sensing (CS), Common Spatial Pattern (CSP).***

I. INTRODUCTION

Many studies have shown over the past two decades that people can communicate by using the scalp recorded electroencephalogram (EEG) activity with no or very little voluntary movement. This is commonly called the EEG based brain-computer interface (BCI) systems. The BCI systems measure certain features of EEG activity, and use them to generate the control signal. Some systems use potentials evoked by stereotyped stimuli [1]. Other systems use EEG features in the frequency domain without specific sensory events [2], [3]. This motor imagery based BCI uses the sensorimotor rhythms (SMRs), including Mu (8-14Hz) and/or Beta (15-30Hz) rhythms, recorded from the scalp over the sensorimotor cortex. When subjects imagine the left or right hand movement, a distinct feature, such as the amplitude attenuation of Mu rhythm and an event related desynchronisation (ERD) in the EEG signaling activity, appears over the contralateral hand area at the sensorimotor cortex [4].

In the BCI system, translation is needed to transform the detected feature of a subject into a command. The translation is done through a classification algorithm. A classification algorithm uses distinctive features in identifying the class to which a test signal belongs. Widely used classification methods in the EEG based BCI field, have been adopted from the pattern recognition community, including the linear discriminant analysis (LDA), the support vector machine (SVM), and the k-nearest-neighbor (kNN) [5].

In this paper, we are interested in developing a new classification method based on the sparse representation. The idea of sparse representation has been used in the compressed sensing (CS) theory [6], [7] which is introduced in year 2006 in the Information Theory community, and draws a lot of attention recently as a new signal acquisition method. The CS theory asserts that many natural signals are sparse in a certain basis and thus can be sparsely represented. These sparsely representable signals can be compressed into *holistic* samples via simple linear projection operations as each sample contains a global view of the entire signal. These holistic samples are the compressed samples of the signal in the sense that the number of compressed samples required for good signal reconstruction is much less than the number of the Shannon-Nyquist rate samples. This is surprising and interesting, the reason for the buzz recently.

The signal recovery given the holistic samples in the CS theory is done via the so-called ell-1 minimization. The reason is that the ell-1 minimization provides a sparse solution and the solution is right if enough number of holistic samples have been obtained. This ell-1 minimization approach can be utilized as a sparse representation tool in this paper for the classification purpose. There are previous attempts in which researchers apply the ell-1 minimization to classification as well, including the face recognition application [8] and the EEG based driver's vigilance detection problem [9].

In this paper, a new classification method has been developed utilizing a sparse representation (SR) of EEG signals for the motor imagery based BCI application. We use a common spatial pattern (CSP) for preprocessing of EEG data and the SMRs as a feature of BCI system. To evaluate the performance of the proposed classification method, we use two classes, the left hand and the right hand, of motor imagery data sets collected from four subjects. We

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (Do-Yak Research Program, N0. 2010-0017944) and NRF grant (NO. 2010-0006135).

also evaluate the performance of the LDA method for comparison with the performance of the proposed method.

This paper is organized as follows: The EEG data acquisition is given in section II. In section III, proposed methods are introduced. Section IV shows the experimental result and discussion. Finally, conclusions are explained in section V.

II. DATA ACQUISITION

In this section, we aim to evaluate the proposed classification method by analyzing four different data sets obtained from four healthy subjects (A: 24 years old man, B: 23 years old woman, C: 17 years old girl, D: 17 years old boy). They are all novice subjects in the related BCI experiment. They have taken the same procedure of a BCI experiment in which there are two classes, left and right hand, of motor imagery movements. They use different number of EEG channels (A, B: 32 Channel, C, D: 12 Channel). The EEG signals are recorded off of the gold plated electrodes attached to the scalp (reference at earlobe) based on the international 10/20 standard, and we use the sampling rate of 256 samples/sec with a band pass filter of 1Hz to 100Hz and a notch filter of 60Hz.

Subjects are seated in a comfortable chair and asked to watch the monitor screen. Figure 1 shows the time procedure of one trial. At the beginning of each run, a 'Left Hand' or 'Right Hand' letter instruction randomly appears for four seconds on the center of the screen. Then, subjects imagine the left or right hand movement (e.g. clenching their fist repeatedly) corresponding to the instruction. After that, there is rest time for three seconds. One run consists of total 40 trials, 20 left and 20 right trials. Due to fatigue of subjects and difference in concentration time span, the size of each data set is different. To suppress the electrooculogram (EOG) artifacts, subjects were instructed not to blink or move their eyes in the instruction time. In the rest time, they can do it freely, but still cannot move their body.

We use a one second of signal samples (256 samples), collected at a certain point in time after the Cue has been presented, for the analysis in Section III. The best starting time (ST) for an individual subject varies; and thus, the best ST that maximizes the classification accuracy is selected for each individual.

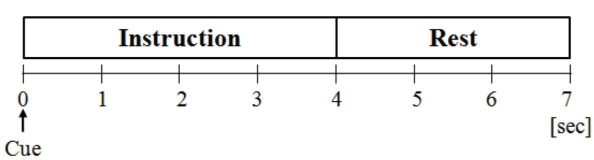

Figure 1. Motor imagery based BCI experimental time procedure of one trial

III. METHODS

We aim to develop a new sparse representation based classification (SRC) method in this section. The SRC method utilizes the common spatial pattern (CSP) method to preprocess the obtained signals and construct the dictionary. It then works by finding a sparse representation of the test signal in terms of the signals included in the dictionary. Here, we explain in the detailed procedure of the proposed method.

A. The Preprocessor

Due to the large dimensionality of EEG signals, we need a preprocessing method. In this paper, we use the common spatial pattern (CSP) method. CSP is a powerful signal processing technique suitable for EEG-based BCIs [10], [11]. CSP filters maximize the variance of the spatially filtered signal under one condition while minimizing it for the other condition.

Let $\mathbf{X} \in \mathbb{R}^{C \times T}$ be a segment of EEG signals where C is the number of EEG channels, and T is the number of sampled time points collected in all the trials for a single subject. In this study, we use 256 samples (one second). We have two classes of EEG training trials $\mathbf{X}^L \in \mathbb{R}^{C \times T}$ and $\mathbf{X}^R \in \mathbb{R}^{C \times T}$ each corresponding to the Left and Right hand motor imaginary movement. Using the CSP method, we can estimate the CSP filters $\mathbf{W} \in \mathbb{R}^{C \times C}$. We call each column vector $\mathbf{w}_i \in \mathbb{R}^C (i = 1,2,...,C)$ of $\mathbf{W}$ a spatial filter. Among them, we use n CSP filters from the front and another set from the back. Then, we can make the CSP filter matrix $\mathbf{W}_{\text{CSP}} \in \mathbb{R}^{C \times 2n}$, i.e., $\mathbf{W}_{\text{CSP}} := [\mathbf{w}_1, ..., \mathbf{w}_n, \mathbf{w}_{C-n+1}, \mathbf{w}_C]$. Given the two classes of EEG training trials $\mathbf{X}^L \in \mathbb{R}^{C \times T}$ and $\mathbf{X}^R \in \mathbb{R}^{C \times T}$, we can define the CSP filtered signals, i.e.,

$$
\begin{aligned}
\mathbf{X}^L_{\text{CSP}} &\in \mathbb{R}^{2n \times T} := \mathbf{W}^{\text{T}}_{\text{CSP}} \mathbf{X}^L \\
\mathbf{X}^R_{\text{CSP}} &\in \mathbb{R}^{2n \times T} := \mathbf{W}^{\text{T}}_{\text{CSP}} \mathbf{X}^R
\end{aligned}
\tag{1}
$$

Next, we compute FFT of each set of $T = 256$ samples and keep the magnitude of the T FFT coefficients. Because the sampling rate is 256 samples/sec, the frequency resolution is 1 Hz. We then keep the FFT coefficients of the Mu band (8~14Hz) and/or the Beta (15~30Hz) band and throw away the rest. Thus, there are maximum $N_f = 24$ columns in each matrix. They are:

$$
\begin{aligned}
\mathbf{X}^L_{\text{CSP}}(f) &\in \mathbb{R}^{2n \times N_f} \\
\mathbf{X}^R_{\text{CSP}}(f) &\in \mathbb{R}^{2n \times N_f}
\end{aligned}
\tag{2}
$$

Further reduction on the set of frequency terms N_f used in our classification can be made depending on individual subjects.

B. Dictionary and Linear Sparse Representation Model

In this section, we aim to introduce the sparse representation of the test signal. Let N_t be the number of total training signals for each class $i = L, R$. We define the dictionary matrix $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, ..., \mathbf{a}_{i,N_t}]$ for $i = L, R$ where each column vector $\mathbf{a} \in \mathbb{R}^{m \times 1}$ having dimension $m = 2n \times N_f$ is obtained by concatenating the $2n$ rows of $\mathbf{X}^L_{\text{CSP}}(f)$ and taking the transpose. Let's call this vectorization. The same procedure is repeated for the right part, $\mathbf{X}_{\mathrm{CSP}}^{R}(f)$. By combining the two matrices, we form the complete dictionary, $\mathbf{A} := [\mathbf{A}_L; \mathbf{A}_R]$. Thus, the dimension of $\mathbf{A}$ is $m \times 2N_t$.

We apply the same procedure done to obtain the columns of the dictionary to the test signal. That is, the test signal is transformed to a vector $\mathbf{y} \in \mathbb{R}^{m\times 1}$ through the CSP filtering, FFT, and vectorization. Thus, the dimesion of $\mathbf{y}$ is the same as the dimension of the columns of the dictionary $\mathbf{A}$. Then, this test signal $\mathbf{y}$ can be sparsely represented as a linear combination of some columns of $\mathbf{A}$:

$$\mathbf{y} = \sum_{i=L,R} x_{i,1}\mathbf{a}_{i,1} + x_{i,2}\mathbf{a}_{i,2} + \cdots + x_{i,n_t}\mathbf{a}_{i,N_t} \tag{3}$$

where $x_{i,j} \in \mathbb{R}, j = 1,2,...,N_t$ are scalar coefficients. Then, we can represent this as a matrix algebraic form:

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{4}$$

where $\mathbf{x} = [x_{L,1}, x_{i,2}, ..., x_{i,N_t} x_{R,1}, x_{i,2}, ..., x_{i,N_t}]^T \in \mathbb{R}^{2\cdot N_t}$. For example, we expect that the test signal $\mathbf{y}$ of class L can be represented as the training signals of class L .

$$\mathbf{y}_L = \mathbf{A}\mathbf{x}_L \in \mathbb{R}^{m\times 1} \tag{5}$$

where $\mathbf{x}_L = [\mathbf{a}_{L,1}, \mathbf{a}_{L,2}, ..., \mathbf{a}_{L,N_t}, 0, ..., 0]^T \in \mathbb{R}^{2N_t}$ is a coefficient vector whose elements are zero except some elements associated with test signals of class L . Sparse representation of the test signal $\mathbf{y}$ can be made when the number of non-zero coefficients of $\mathbf{x}$ is much smaller than N_t .

C. Sparse Representation by ell-1 Minimization

We have the number of total training signals $2N_t$ which is larger than the number of FFT coefficients ($m = 2n \times N_f$) . The more trials, the larger the dictionary, and thus the better the sparse representation result will be. The column size is larger than the row size of the dictionary $\mathbf{A}$. Then, the linear equation (5) is under-determined ($m < 2N_t$) . Recent studies in the Compressed Sensing theory have shown that the ell-1 norm minimization, given below, can solve this under-determined system well in polynomial time [12]:

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x} \tag{6}$$

Unlike the conventional ell-2 norm minimization, the ell-1 norm minimization gives a sparse representation result. There are many ell-1 minimization algorithms. In this paper, we use one of the standard linear programming methods [13], the SolveBP function implements the basis pursuit algorithm available in the SparseLab, which is a free MATLAB software package [14].

D. Sparse Respresentation based Classification

After solving the ell-1 minimization problem, the nonzero elements of $\mathbf{x}$ must be corresponding to the column of class i . Because the EEG signal is very noisy, the nonzero elements may appear in the indices corresponding to the column of another class. To make use of the sparse representation result, the coefficient vector $\mathbf{x}$, in a classification problem, we introduce the characteristic function δ [8]. For each class i , we define its characteristic function $\delta_i : \mathbb{R}^{2N_t} \to \mathbb{R}^{2N_t}$ which selects the coefficients associated with class i . For $\mathbf{x} \in \mathbb{R}^{2N_t}$, $\delta_i(\mathbf{x}) \in \mathbb{R}^{2N_t}$ is a new vector which is obtained by nulling all the elements of $\mathbf{x}$ that are associated with the other class. Then we can obtain the residuals $r_i(\mathbf{y}) := \|\mathbf{y} - \mathbf{A}\delta_i(\mathbf{x})\|_2$ for L and R. Then, the classification rule is given by:

$$\text{class}(\mathbf{y}) = \arg\min_i r_i(\mathbf{y}) \tag{7}$$

That is, we determine the class i that has the minimum residuals.

E. The Classification Algorithm

1. Input: Training signals $\mathbf{A} \in \mathbb{R}^{m\times 2N_t}$ for i classes, a test signal $\mathbf{y} \in \mathbb{R}^{m\times 1}$.
2. Normalize the columns of $\mathbf{A}$ and $\mathbf{y}$.
3. Solve the convex optimization problem :
   $$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$
4. Compute the residuals $r_i(\mathbf{y}) := \|\mathbf{y} - \mathbf{A}\delta_i(\mathbf{x})\|_2$ for class i
5. Output: $\text{class}(\mathbf{y}) = \arg\min_i r_i(\mathbf{y})$

IV. RESULTS AND DISCUSSION

We have analyzed four data sets, which have different number of trials. For each data set, we compare performance of the proposed sparse representation based classification (SRC) with the linear discriminant analysis (LDA) classification method. To make a fair comparison, in the LDA classification, we also use the CSP filtering, the FFT, and the Mu(8~14Hz) and/or the Beta(15~30Hz) rhythms. In this paper, for all the subjects, we use the same 8~15Hz frequency amplitudes as the feature in both classification methods.

We use the statistical k -fold cross-validation method to evaluate the average performance of the classifiers [15]. The data set is divided into k subsets. Each time, one of the k subsets is used as the test set and the union of the other $k-1$ subsets is the training set. Repeat this same procedure k times with different subset selections. Then the average performance over all subset selections is computed. TABLE I shows the classification accuracy of subject A and B (32 channels) using the k -fold cross-validation method. When the number of subsets k increases, most signals are used as the training signals. The classification accuracy is calculated from the following equation:

$$\text{Accuracy}(\%) = \frac{\text{correct test trials}}{\text{total test trials}} \times 100 \qquad (8)$$

TABLE I. CLASSIFICATION ACCURACY OF SUBJECT A AND B

Subject	# of training signals	# of test signals	Cross-validation k	Accuracy (%)	
				LDA	*SRC*
A (total 200 signals)	150	50	4	63.50	71.25
	160	40	5	67.75	75.50
	180	20	10	68.50	77.75
	190	10	20	69.25	79.50
	199	1	200	68.75	79.00
B (total 100 signals)	50	50	2	67.50	71.50
	80	20	5	68.00	80.00
	90	10	10	69.50	82.50
	95	5	20	72.00	82.00
	99	1	100	72.50	82.00

We compare the classification accuracies for LDA and SRC while the number of training signals increases. For both subject A and B, the accuracy of the SRC is superior to that of the LDA method. TABLE II indicates the classification accuracies of subject C and D (12 channels). For subject C, SRC is better than LDA. For subject D, when the number of training signals is small LDA is better. But, when the number of training signals increases, SRC is better than LDA.

TABLE II. CLASSIFICATION ACCURACY OF SUBJECT C AND D

Subject	# of training signals	# of test signals	Cross-validation k	Accuracy (%)	
				LDA	*SRC*
C (total 60 signals)	40	20	3	89.17	91.67
	50	10	6	89.17	91.67
	55	5	12	89.17	90.84
	58	2	30	88.33	91.67
	59	1	60	88.33	91.67
D (total 60 signals)	40	20	3	80.83	77.50
	50	10	6	85.83	81.67
	55	5	12	86.67	78.33
	58	2	30	84.17	85.00
	59	1	60	85.00	86.67

In the cases of subjects C and D, the difference in accuracy is not noticeable for the LDA and the SRC. This is perhaps due to the fact that the data size (the total number of signals) is not large enough for the SRC to stand out since for the SRC, the larger the dictionary the better the result is. In addition, the accuracy of LDA is already quite high and there is a little room for improvement.

When the number of training signals is large enough, say 59 training signals in the case of subject C, the proposed SRC method shows high accuracy than LDA does. For comparison of complexity, we have checked the run times in MATLAB of the two algorithms: SRC took 37sec while LDA took 34sec. Thus, we note that the computation cost of SRC is a little larger than that of LDA.

V. CONCLUSIONS

In this paper, we apply the idea of sparse representation as a new classification method to the motor imagery based BCI. In sparse representation, a well constructed dictionary matrix is important. We use the CSP filtering and the FFT to produce the columns of the dictionary matrix. We have shown that a good classification result can be obtained from the proposed sparse representation based classification method developed in this paper, which is novel to the best of author's knowledge. We compare the result of the proposed method with that of the LDA based classification. From the comparison, we can see that the proposed classification method shows an accuracy better than the LDA method does, especially when the number of training signals is large.

REFERENCES

[1] L. A. Farewell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*,vol. 70, pp. 510-523, 1988.

[2] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interface-a new communication device for handicapped persons," *J. Microcomput. Appl.,* vol. 16, pp. 293-299, 1993.

[3] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol.,* vol. 78, pp. 252-259, 1991.

[4] G. Pfurtscheller, Ch. Neuper, D. Flotzinger, M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.*,vol. 103, pp. 642-651, 1997.

[5] A. Schlogl, F. Lee, H. Bishof, G. Pfurtscheller and G. "Characterization of four-class motor imagery data for the BCI-competition 2005"", *J. Neural Eng.*, vol. 2, pp. 14 2005.

[6] D.Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, pp. 1289-1306, 2006.

[7] E. J. Candes, M. B. Wakin., "An Introduction To Compressive Sampling," *IEEE Signal Proc. Magazine*, vol. 25, No. 2.,21 March 2008, pp 21-30.

[8] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, Yi Ma, "Robust Face Recognition via Sparse Representation" IEEE Trans. *Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210~227, February 2009.

[9] Hongbin Yu, Hongtao Lu, Tian Ouyang, Hongjun Liu, Bao-Liang Lu, "Vigilance Detection Based on Sparse Representation of EEG," *32nd international conference- IEEE/EMBS,* Aug.31-Sep.4 Buenos Aires, 2010, pp2439-2442.

[10] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[11] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," *IEEE Signal Proc. Magazine*, 25(1):41-56, January 2008.

[12] E. Cande`s, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," *Comm. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207-1223, 2006.

[13] S. Chen, D. Donoho, and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Rev*., vol. 43, no. 1, pp. 129-159, 2001.

[14] David Donoho, Victoria Stodden and Ysskov Tsaig, SparseLab, http://sparselab.stanford.edu/

[15] Larry Wasserman, "All of Statistics: A Concise Course in Statistical Inference", Springer, 2010, pp. 63-64.