Proceedings of the 3rd International
IEEE EMBS Conference on Neural Engineering
Kohala Coast, Hawaii, USA, May 2-5, 2007

Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface

Quadrianto Novi, Cuntai Guan, Tran Huy Dat, and Ping Xue

***Abstract*— Brain-computer interface (BCI) is a system to translate humans thoughts into commands. For Electroencephalography (EEG) based BCI, motor imagery is considered as one of the most effective ways. Different imagery activities can be classified based on the changes in μ and/or β rhythms and their spatial distributions. However, the change in these rhythmic patterns varies from one subject to another. This causes an unavoidable time-consuming fine-tuning process in building a BCI for every subject. To address this issue, we propose a new method called Sub-band Common Spatial Pattern (SBCSP) to solve the problem. First, we decompose the EEG signals into sub-bands using a filter bank. Subsequently, we apply a discriminative analysis to extract SBCSP features. The SBCSP features are then fed into Linear Discriminant Analyzers (LDA) to obtain scores which reflect the classification capability of each frequency band. Finally, the scores are fused to make decision. We evaluate two fusion methods: Recursive Band Elimination (RBE) and Meta-Classifier (MC). We assess our approaches on a standard database from BCI Competition III. We also compare our method with two other approaches that address the same issue. The results show that our method outperforms the other two approaches and achieves similar result as compared to the best one in the literature which was obtained by a time-consuming fine-tuning process.**

I. INTRODUCTION

People suffering from Amyotrophic Lateral Sclerosis (ALS) lose their muscle movement degeneratively and at a later stage may become totally paralyzed. Nevertheless, for most ALS patients, their minds are un-affected. Brain-computer interface (BCI) addresses this concern by making it possible to translate human thoughts directly to the outside world [7]. Electroencephalography (EEG) has been chosen to capture brainwaves for BCI applications because of its simplicity, inexpensiveness and high temporal resolution.

In the imagination of limb movement, suppression of EEG signals happens in the specific region of the motor and somatosensory cortex due to loss of synchrony in μ and β bands, classically defined in the 12-16Hz and 18-24Hz respectively, is termed event-related de-synchronization (ERD) [4]. This brain rhythm will benefit ALS patients as it can be used as a control signal for assistive devices like wheelchair and neuroprosthesis. However in practise, frequency band which reflects ERD varies from subject to subject. This is one of the most challenging issues when designing a practical BCI.

In literature, the Common Spatial Pattern (CSP) method [6] has shown its efficacy in extracting topographic pattern of brain rhythm modulations, also known as the ERD. However, this spatial filter must only be applied to the informative frequency bands (μ and β bands), which is specific to each subject. This is related to the fact that neurophysiologically the discriminative band of ERD varies from one subject to another. In general, applying the CSP method to un-filtered or filtered EEG signals but with a poor frequency bands selection will result in a poor recognition accuracy. One way to find the "best" band is to do an exhaustive search and some manual tweaking for each subject. Although this method has been proven to be effective, it is a time consuming and meticulous process. Lack of standardization in doing this manual selection results in a wide range of performance across researchers working independently with the same CSP algorithm. From a practical point of view, a systematic and easy-to-implement way of revealing subject-specific spectral filter is important.

To overcome the limitation of CSP, the (Common Spatio-Spectral Pattern) CSSP [10] algorithm was proposed. In this algorithm simple filters (with one delay tap) are optimized together with the spatial filters. Recently, a further improvement to the CSSP was presented and called (Common Sparse Spectral Spatial Pattern) CSSSP [3]. This method allows simultaneous optimization of an arbitrary FIR filter within CSP analysis. However, due to inherent nature of optimization problem, the solution of filter coefficients will depend greatly on the initial points.

Here, we propose an alternative method based on Sub-band CSP (SBCSP) and score fusion. Instead of temporal FIR filtering, we decompose the EEG signals into sub-bands using a filter bank. The CSP is performed at each sub-band and subsequently a sub-band score is defined. The final decision is derived from fusion of this score from each sub-band. The usage of different frequency bands at the same time could be advantageous [2]. We propose two fusion methods. The first method is Recursive Band Elimination (RBE) where scores are ranked based on the margin maximization criteria. The second method is Meta-Classifier (MC) which employs a secondary classifier in order to compensate for errors from the Bayesian classifiers. The two methods are examined with a publicly available data set from BCI Competition III in 2005.

Q. Novi is now with Statistical Machine Learning Program National ICT Australia and CSL, RSISE Australian National University, Canberra, Australia. novi.quadrianto@rsise.anu.edu.au

C. Guan is with Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 ctguan@i2r.a-star.edu.sg

T. H. Dat is with Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 hdtran@i2r.a-star.edu.sg

P. Xue is with School of Electrical and Electronic Engineering, Nanyang Technological University, Block S1, 50 Nanyang Avenue, Singapore 639798 epxue@ntu.edu.sg

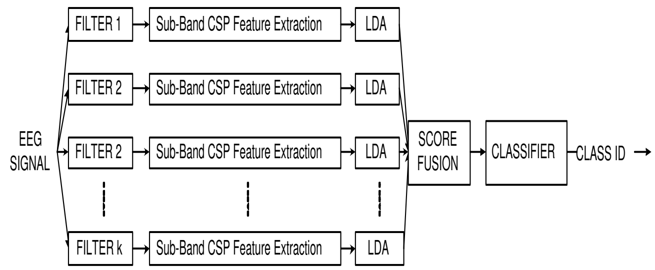


Fig. 1. System Flowchart

II. Sub-band Common Spatial Pattern (SBCSP) Method

A. Common Spatial Pattern (CSP)

The goal of Common Spatial Pattern (CSP) is to design spatial filters that lead to new time series whose variances are optimal for the discrimination of two classes of EEG [6]. Given a single trial, an N-channel spatial-temporal EEG signal E, where E is a N $\times$ T matrix and T denotes the number of samples in each channel. The normalized covariance matrix of the EEG can be obtained from

$$C = \frac{EE^T}{\mathrm{trace}(EE^T)} \tag{1}$$

In this way, the covariance matrix of each class, C_1 and C_2, can be computed by averaging over the trials. The composite covariance matrix and its eigenvalue decomposition is given by

$$C_c = C_1 + C_2 = F_c \psi F_c^T \tag{2}$$

where F_c is a matrix of normalized eigenvectors with corresponding matrix of eigenvalues, ψ. The whitening transformation,

$$P = \psi^{-1/2} F_c^T \tag{3}$$

equalizes the variances in the space spanned by the eigenvectors in F_c.

The CSP is extracted based on the simultaneous diagonalization of whitened covariance matrices

$$\hat{C}_1 = PC_1P^T \quad \text{and} \quad \hat{C}_2 = PC_2P^T. \tag{4}$$

The resulting decomposition maximizes the differentiation between two groups of data. This is done by calculating orthogonal matrix U and diagonal matrix λ,

$$\hat{C}_1 = U\lambda U^T \quad \text{and} \quad \hat{C}_2 = U(1-\lambda)U^T. \tag{5}$$

The CSP projection matrix will then be $W_{csp} = (U^T P)$.

B. Filtering

In this paper, the EEG signals are represented by their sub-band distributions. A Gabor filter is adopted as a bandpass filter, whose impulse response is defined by a harmonic function multiplied by a Gaussian function as follows

$$g(t, f_0, \sigma) = \exp\left(-\frac{t^2}{\sigma^2} + jf_0t\right) \tag{6}$$

with the bandwidth proportional to σ. In our method, a set of Gabor filters is convoluted with the input EEG signal resulting in a time-frequency representation.

C. Sub-band CSP

In our study, the extraction of brain rhythm topographic patterns by CSP is performed on each sub-band of EEG signal. Thus, the transformed signal at the k-th sub-band is in the form of

$$Z^{(k)} = W_{csp}^{(k)} E^{(k)}. \tag{7}$$

The signals $Z^{(k)}$ that maximize the difference in variances of two classes would correspond to largest eigenvalues of the simultaneous diagonalization result, i.e. the first r and the last r rows would contain the most discriminative information. Therefore, the SBCSP features are defined as:

$$f_p^{(k)} = \log\left(\frac{\mathrm{var}(Z_p^{(k)})}{\sum_{p=2r}\mathrm{var}(Z_p^{(k)})}\right) \text{ and } p = (1...2r). \tag{8}$$

The typical value of r is 1.

D. Sub-band Score

We focus on the continuous outputs of classifiers in the subsequent analysis. Linear Discriminant Analysis (LDA) [9] finds projection matrix, W_{lda} that guarantees maximal separability by maximizing the ratio between-class variance, S_B to the within-class variance, S_W. The cost function at k-th sub-band would be:

$$G^{(k)} = \frac{W_{lda}^{(k)T} S_B^{(k)} W_{lda}^{(k)}}{W_{lda}^{(k)T} S_W^{(k)} W_{lda}^{(k)}}. \tag{9}$$

where $S_B^{(k)}$ and $S_W^{(k)}$ are defined as follow:

$$\begin{aligned} S_B^{(k)} &= (m_2^{(k)} - m_1^{(k)})(m_2^{(k)} - m_1^{(k)})^T \\ S_W^{(k)} &= \sum_{f_p^{(k)} \in c1} (f_p^{(k)} - m_1^{(k)})^2 + \sum_{f_p^{(k)} \in c2} (f_p^{(k)} - m_2^{(k)})^2. \end{aligned} \tag{10}$$

where c1 and c2 denote class 1 and class 2 respectively. The value of $m_1^{(k)}$ and $m_2^{(k)}$ are the empirical class means of SBCSP features computed from the training set. For a 2-class problem, the LDA will project the data to a one dimensional representation. Therefore, we define the score at k-th sub-band as

$$s_k = W_{lda}^{(k)T} f_p^{(k)}. \tag{11}$$

In this paper, we use this sub-band score as a feature.

III. Sub-band Score Fusion

A. Recursive Band Elimination (RBE)

Practically, EEG signals are corrupted by noise and hence, only some bands are useful. Therefore, the selection of bands would intuitively give more accurate classification. Feature selection methods are essentially divided into wrapper type and filter type. In general, wrapper methods which include classifiers as a black box perform better than filter methods. One of the state-of-the-art wrapper type feature selection

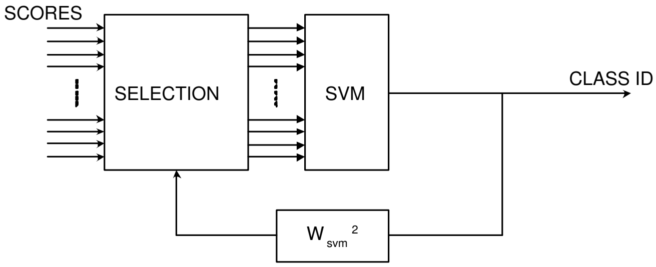


Fig. 2. Diagram of a Recursive Band Elimination Method

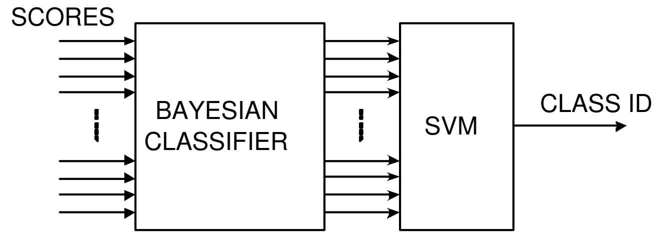


Fig. 3. Diagram of a Meta-Classifier Method

method is Support Vector Machine Recursive Feature Elimination (SVM RFE) [5] which has shown superior performance in a gene selection problem.

In this paper, we adopt SVM RFE as a bands selection method and refer this as Recursive Band Elimination (RBE). Let us consider the input vector X as a concatenation of the sub-bands scores defined in (11), i.e.

$$X = [s_1, s_2, ..., s_k]^T \tag{12}$$

SVM will try to separate data $X \subset \mathbf{R}^k$ from two classes by finding a weight vector $W_{svm} \in \mathbf{R}^k$ and an offset $b \in \mathbf{R}$ of a hyperplane

$$\begin{aligned} H : \mathbf{R}^k &\rightarrow \{1, 2\} \\ X &\mapsto \text{sign}(W_{svm}.X + b) \end{aligned} \tag{13}$$

The selection in the RBE method is done as follow: at each iteration, the algorithm removes one band with the smallest W_{svm}^2 until the set of "surviving" frequency bands is empty. This results in frequency bands which are ranked according to their predictive power. Only few best ranked bands are used for further processing and we shall call this number as an order of RBE. In this work, the RBE order is optimized empirically.

B. Meta-Classifier (MC)

Another way to implement fusion of scores features is by Bayesian classifiers. Assuming that the class conditional distributions of scores are equal normal distributions, i.e.

$$p(s_k|\omega_i) = (2\pi\sigma_i^{(k)2})^{-\frac{1}{2}} \exp\left(-\frac{(s_k - \mu_i^{(k)})^2}{2\sigma_i^{(k)2}}\right) \tag{14}$$

where $\mu_i^{(k)}$ and $\sigma_i^{(k)}$ are the mean and standard deviation of scores features, respectively, which were estimated from the training set. Typically for a Bayesian classifier, the log-likelihood ratio value is used. Thus, we define the output of Bayesian classifiers as a meta-score and is expressed as k-vectors $[X_1, X_2, X_3, ..., X_k]^T$ with

$$X_k = \log\left(\frac{p(s_k|\omega_1)}{p(s_k|\omega_2)}\right). \tag{15}$$

However, the above Bayesian solution of score feature will not be optimal if the covariance matrices for the classes are different [9].

Hence, we propose an additional classifier, SVM in our case, to be placed at the output of Bayesian decision functions. This meta-classifier is supposed to compensate for the errors produced by individual Bayesian classifiers by taking into account possible high level relations. This proposed method avoids the iterative scheme in the training phase which is present in the RBE method. In section IV, it is shown that the average performance between the MC and the RBE are of no significant difference.

IV. Experiments and Discussions

To evaluate our method, we test with dataset IVa from BCI competition III 2005 [1]. This dataset comprises of 118 electrode channels out of an EEG amplifier sampled at 100Hz. These data are collected from five subjects ('aa', 'al', 'av', 'aw', 'ay') performing left hand (L), right hand (R), and right foot (F) imagination. But only cues for right hand and right foot imagination are given for the competition. The visual cues at each trial last for 3.5 seconds and there are 280 trials for each subject.

The experiment results of exhaustive search and manual tweaking of CSP, the state-of-the-art CSSP and CSSSP are quoted from [8]. The parameters for manually optimized CSP used in [8] are based on the winner of BCI Competition 2005. We restricted ourselves to the data between 0.5 seconds and 2.5 seconds from the visual cue (i.e. 200 time points at each channel) according to [8]. For the filter bank system settings, we use 24 Gabor Filters each with bandwidth of 4 Hz. The RBE order used in this experiment is 10 and a linear kernel is used in the MC method. It is important to note that those parameters are applied for all subjects and empirically verified so that the choice of parameters is robust within a reasonable range.

The 10-fold cross validation error rate is shown in Table I. CSP with exhaustive search and manual tweaking will, in general, give an optimum performance. Statistically, there is no significant difference between CSP and our proposed methods, RBE and MC. The CSSP and CSSSP are also designed to remove manual spectral filter tuning. For subject 'aa', the RBE performs better than CSSP with $P < 0.05$ (t-test), which means we have 95% confidence of rejecting hypothesis that the two means have no difference. For subject 'ay', the MC and RBE perform better than CSSSP with $P < 0.01$ and $P < 0.05$, respectively. For other comparisons, there are statistically insignificant. It is worth mentioning that

TABLE I

COMPARISON OF PROPOSED METHODS (MC AND RBE) TO OPTIMIZED CSP, CSSP, AND CSSSP IN TERM OF % CLASSIFICATION ERROR

Sub	CSP[8]	CSSP[8]	CSSSP[8]	MC	RBE
aa	8.5±5.4	14.6±6.2	11.6±6.3	10.7±5.6	9.2±4.5
al	0.8±1.8	2.3±3.0	2.1±2.7	1.4±1.8	2.2±3.4
av	29.1±8.2	32.6±7.6	31.8±7.7	29.6±5.3	31.0±7.3
aw	3.1±2.8	3.5±3.3	6.5±4.3	4.3±4.0	4.2±3.3
ay	5.3±3.8	6.0±3.9	10.5±5.7	4.3±2.8	5.0±3.4
avg	**9.4**	**11.8**	**12.5**	**10.0**	**10.3**

the simplicity of our proposed methods is also one of the advantages.

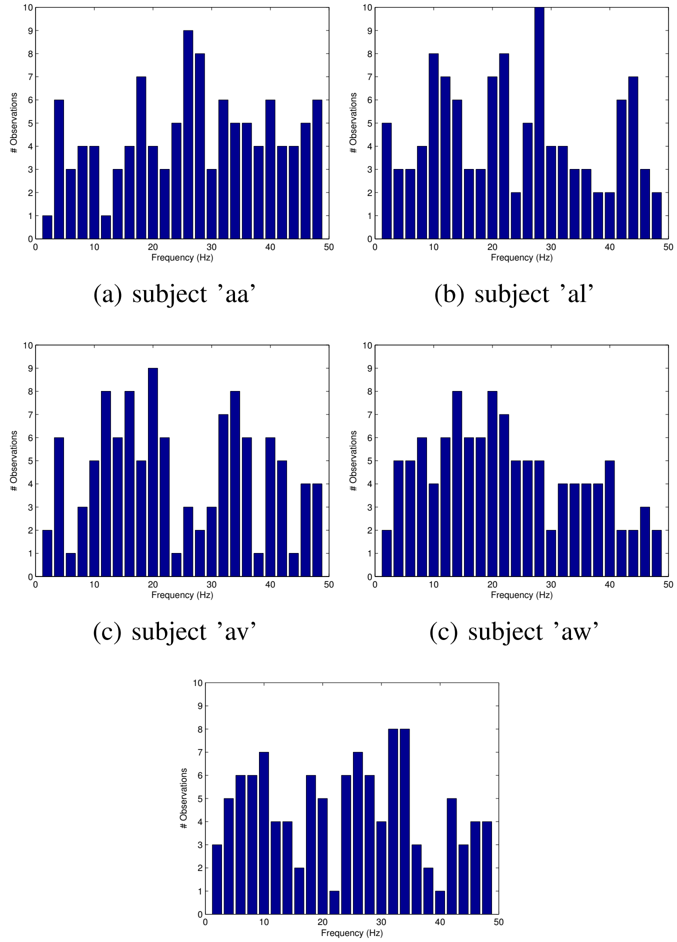
Furthermore, the results of ranking the importance of each band in the RBE method would provide insights to the underlying cortical activity pattern. From the histogram in Fig. (4), it can be seen that there exists a significant variety of the discriminative bands among different subjects. This variety makes it necessary for traditional approaches to be tuned in a time-consuming manner so as to achieve the optimal performance. Our approach overcomes such a fine-tuning process and can easily achieve results close to the optimum.

V. CONCLUSIONS

In this paper, we have developed a new framework to overcome the problem of a time-consuming fine-tuning process in building a BCI for each subject. This problem is prevalent in BCI research and it causes more trouble when researchers try to benchmark and compare different methods. As such fine-tuned system is hardly repeatable by other research groups. We propose a Sub-band CSP (SBCSP) framework, which provides an alternative solution to address this issue. With the benchmark using standard database from BCI Competition III, SBCSP achieved similar result as the best method in the literature, which was obtained by a time-consuming fine-tuning process. We also compared our method with two competing approaches in the literature, namely CSSP and CSSSP, our approach outperforms both of them. The major contribution of our work is that, it not only achieves the best accuracy, but gives a robust and consistent solution.

ACKNOWLEDGMENT

The authors would like to thank Fraunhofer FIRST, Intelligent Data Analysis Group and Campus Benjamin Franklin of the Charité - University Medicine Berlin, Department of Neurology, Neurophysics Group for providing the data used in BCI Competition III.



(a) subject 'aa' (b) subject 'al' (c) subject 'av' (c) subject 'aw' (c) subject 'ay'

Fig. 4. Histogram of Frequency Bands of the RBE method. The y-axis describes number of times a particular band is selected during 10-fold cross validation.

REFERENCES

[1] BCI Competition III, `http://ida.first.fraunhofer.de/projects/bci/competition_iii`, 2005.

[2] G. Blanchard and B. Blankertz, *BCI competition 2003-dataset IIa: spatial patterns of self-controlled brain rhythm modulations*, IEEE Trans. Biomedical Engineering, vol.51, pp. 1062-1066, 2004.

[3] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio and K.-R. Müller, *Optimizing spatio-temporal filters for improving BCI*, In Advances in Neural Info. Proc. Sys. vol. 18, 2006.

[4] G. Pfurtscheller and F. Lopes da Silva, *Event-related EEG/MEG synchronization and desynchronization: basic principles*, Clin. Neurophys., vol. 110, pp. 1842-1857, 1999.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, *Gene selection for cancer classification using support vector machines*, Journal of Machine Learning Research, vol. 3, pp. 1439-1461, 2003.

[6] J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, *Designing optimal spatial filters for single-trial EEG classification in a movement task*, Clin. Neurophys., vol. 110, pp. 787-798, 1999.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller and T. M. Vaughan, *Brain-computer interfaces for communication and control*, Clin. Neurophys., vol. 113, pp. 767-791, 2002.

[8] L. Song and J. Epps, *Classifying EEG for brain-computer interfaces: learning optimal filters for dynamical system features*, 23rd International Conference on Machine Learning (ICML), 2006.

[9] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, New York: Wiley, 2nd ed., 2001.

[10] S. Lemm, B. Blankertz, G. Curio and K.-R. Müller, *Spatio-spectral filters for improving the classification of single trial EEG*, IEEE Trans. Biomedical Engineering, vol. 52, pp. 1541-1548, 2005.