Incremental Common Spatial Pattern Algorithm for BCI

Qibin Zhao, Liqing Zhang, Andrzej Cichocki and Jie Li

***Abstract*— A major challenge in applying machine learning methods to Brain-Computer Interfaces (BCIs) is to overcome the on-line non-stationarity of the data blocks. An effective BCI system should be adaptive to and robust against the dynamic variations in brain signals. One solution to it is to adapt the model parameters of BCI system online. However, CSP is poor at adaptability since it is a batch type algorithm. To overcome this, in this paper, we propose the Incremental Common Spatial Pattern (ICSP) algorithm which performs the adaptive feature extraction on-line. This method allows us to perform the online adjustment of spatial filter. This procedure helps the BCI system robust to possible non-stationarity of the EEG data. We test our method to data from BCI motor imagery experiments, and the results demonstrate the good performance of adaptation of the proposed algorithm.**

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) are devices that translate the intention of a subject measured from brain signals directly into control commands, e.g. for computer operation or a neuroprosthesis[1], [2]. It provides an alternative means of communication and control for people with severe motor disabilities. In this paper, we focus on EEG based motor imagery BCIs. EEG based BCI measures specific components of EEG activity, extracts their features and translates these features into commands to control a cursor on the screen or devices (e.g., a robot arm, a wheelchair, etc.). Among various features extracted from EEG signal, common spatial pattern (CSP) is one of the most effective feature extraction methods in discriminating different classes of motor imagery of left and right hands [3], [4], [5]. To extract CSP features, an initial calibration process is needed to collect labelled training data for the estimation of the CSP transformation matrix, which is actually a spatial filter. In each new BCI experiment on different days, the user has to be re-trained over a number of sessions and features should be adjusted. Consider a situation when a transformation and classifier is trained on one block of recording and then applied to another block. A potential drawback is that the characteristic of the signal can be considerably affected by altered mental states with respect to, e.g., concentration or excitedness. A BCI system needs to adapt to such changes[6]. Therefore, the challenge is to adapt the transformation matrix and classifier which is optimized on the first block to the consecutive block.

It seems reasonable to assume that, over the course of a motor-imagery BCI experiment, the spatial patterns for relevant sources will change relatively little (we will assume that the positions of the sources in the motor cortex, and the spectral content of the signals they generate, are relatively constant). A straightforward approach is that we can collect data whenever new data are presented and then construct a provisional system by a batch learning over the collected data so far. However, it is obvious that such system only works under a condition of a large memory and high computation expenses, because the system would need to maintain a huge memory to store the data either previously learned, or newly presented. Moreover, the system has to discard the knowledge acquired in the past, even if the learning of 99.9% data has been used in the previous training, and repeat the learning from the beginning whenever one additional sample is presented. Obviously, one-pass incremental learning gives a solution to the above problem. In this learning scheme, a system must acquire knowledge with a single presentation of the training data and retaining the knowledge acquired in the past without keeping a large number of training samples. To solve this problem, in this paper we propose a new incremental common spatial pattern algorithm in which the feature extraction is performed on-line. In this way, the consistency of features is ensured. We tested our method on EEG data collected from BCI experiments, and the results demonstrated the validity of our algorithm.

Qibin Zhao, Liqing Zhang and Jie Li are with Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China (email: zhang-lq@cs.sjtu.edu.cn.)

Andrzej Cichocki is with Lab for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan

II. INCREMENTAL COMMON SPATIAL PATTERN (ICSP)

In this section, we show that the CSP feature extraction can be estimated and interpreted using the framework of Rayleigh coefficient maximization[6], [7]. Common Spatial Pattern (CSP) is a spatial filtering method widely used for motor imagery based BCIs, where the task is to classify two different state of brain activity, e.g., imagery of the movement of the left or the right hand. In this context, the event-related (de-)synchronization (ERD/ERS) of rhythmic brain activity is a widely used and well studied physiology. The EEG signal is commonly band-pass filtered around $\mu - band$ (7-15Hz) and $\beta-band$ (15-30Hz) rhythms and two covariance matrices Σ_l and Σ_r are calculated for the two classes.

$$\Sigma_l = \sum_{j \in C_l} \frac{\mathbf{E}_j * \mathbf{E}_j^T}{trace(\mathbf{E}_j * \mathbf{E}_j^T)}, \quad \Sigma_r = \sum_{j \in C_r} \frac{\mathbf{E}_j * \mathbf{E}_j^T}{trace(\mathbf{E}_j * \mathbf{E}_j^T)}, \tag{1}$$

where $\mathbf{E}_j \in R^{m \times k}$ denotes an EEG data matrix of the jth trial, m is the number of selected channels, k is the number of samples in each trial, C_l and C_r refer to the two classes (left and right hand imagery) of the training data.

The CSP tries to find a spatial filter described by vector $\mathbf{w} \in R^d$ that maximizes the difference in the average band power of the filtered signal while keeping the sum constant.

$$\mathbf{w}\Sigma_l\mathbf{w}^T = \mathbf{D}, \qquad \mathbf{w}\Sigma_r\mathbf{w}^T = \mathbf{I} - \mathbf{D}, \tag{2}$$

978-1-4244-1821-3/08/$25.00

where $\mathbf{I}$ is an identity matrix, $\mathbf{D}$ is a diagonal matrix. We can then construct a matrix $\mathbf{W}$ called CSP transformation matrix, composed of the first m and the last m of $\mathbf{w}$ which correspond to the first m and the last m ordered eigenvalue.

A. Generalized Eigenvalue Decomposition

The CSP feature extraction can be explained in the framework of Rayleigh coefficient maximization. Maximizing Rayleigh coefficient is to solve the following optimization problem.

$$maxJ(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{R}_l\mathbf{w}}{\mathbf{w}^T\mathbf{R}_c\mathbf{w}}, \tag{3}$$

where $J(\mathbf{w})$ is Rayleigh coefficient, $\mathbf{R}_l$ and $\mathbf{R}_c$ are symmetric $m \times m$ matrices designed such that they measure the desired information and the undesired noise along the direction of $\mathbf{w}$, For signal processing applications, these are generally full covariance matrices of zero mean, stationary random signals. Taking the gradient of $J(\mathbf{w})$, we get

$$\nabla J \propto (\frac{1}{\mathbf{w}^T\mathbf{R}_c\mathbf{w}})[\mathbf{R}_l\mathbf{w} - (\frac{\mathbf{w}^T\mathbf{R}_l\mathbf{w}}{\mathbf{w}^T\mathbf{R}_c\mathbf{w}})\mathbf{R}_c\mathbf{w}]. \tag{4}$$

Equating the gradient of (4) to zero,

$$\nabla J = 0,$$

hence, the solution of (3) can be obtained by solving the following generalized eigenvalue decomposition (GED) problem,

$$\mathbf{R}_l\mathbf{w} = \lambda\mathbf{R}_c\mathbf{w}, \qquad \lambda = \frac{\mathbf{w}^T\mathbf{R}_l\mathbf{w}}{\mathbf{w}^T\mathbf{R}_c\mathbf{w}}, \tag{5}$$

where λ is a generalized eigenvalue, and $\mathbf{w}$ is a generalized eigenvector corresponding to λ. For real symmetric positive definite covariance matrices, the eigenvectors are all real and eigenvalues strictly positive.

Generalized eigenvectors accomplish simultaneous diagonalization of covariance matrices as $\mathbf{w}^T\mathbf{R}_l\mathbf{w} = \Lambda$, $\mathbf{w}^T\mathbf{R}_c\mathbf{w} = \mathbf{I}$ where $\mathbf{I}$ is an identity matrix. The generalized eigenvector works as a filter in the joint space of the two signals, minimizing the energy of one of the signals and maximizing the energy of the other at the same time. Thus GED can be used as a filter that can perform signal separation by suppressing the undesired component in a composite signal. This property can be used to quantify changes in the time series.

The Rayleigh coefficient in (3) is maximized when one covers as much as possible for the desired information simultaneously avoiding the undesired components.

Let us define

$$\mathbf{R}_l = \Sigma_1, \Sigma_2, \Sigma_1 - \Sigma_2, \Sigma_2 - \Sigma_1; \quad \mathbf{R}_c = \Sigma_1 + \Sigma_2. \tag{6}$$

In Eq.(6), $\Sigma_i, i = 1, 2$ are covariance matrixes for EEG trials of class i. We can see that

$$\begin{aligned} maxJ_1(\mathbf{w}) \quad &\Leftrightarrow \quad max\frac{\mathbf{w}^T\Sigma_1\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} \\ &\Leftrightarrow \quad max(\frac{\mathbf{w}^T\Sigma_1\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} - \frac{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}) \\ &\Leftrightarrow \quad max(-\frac{\mathbf{w}^T\Sigma_2\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}) \\ &\Leftrightarrow \quad min(\frac{\mathbf{w}^T\Sigma_2\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}), \end{aligned} \tag{7}$$

and

$$\begin{aligned} maxJ_2(\mathbf{w}) \quad &\Leftrightarrow \quad max\frac{\mathbf{w}^T\Sigma_2\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} \\ &\Leftrightarrow \quad min(\frac{\mathbf{w}^T\Sigma_1\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}), \end{aligned} \tag{8}$$

where $J_1(\mathbf{w}) + J_2(\mathbf{w}) = 1$, hence, $maxJ_1(\mathbf{w})$ equivalent to $minJ_2(\mathbf{w})$ and $maxJ_2(\mathbf{w})$ equivalent to $minJ_1(\mathbf{w})$. Then, the $\mathbf{w}$ for maximizing the Rayleigh for Σ_1 is equivalent to minimizing the Rayleigh for Σ_2, it is the same objective as CSP. We also can see that,

$$\begin{aligned} & max\frac{\mathbf{w}^T(\Sigma_1-\Sigma_2)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} \\ \Leftrightarrow \quad & max(\frac{\mathbf{w}^T(\Sigma_1-\Sigma_2)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} + \frac{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}) \\ \Leftrightarrow \quad & max(\frac{\mathbf{w}^T\Sigma_1\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}), \end{aligned}$$

and

$$\begin{aligned} & max\frac{\mathbf{w}^T(\Sigma_2-\Sigma_1)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} \\ \Leftrightarrow \quad & max(\frac{\mathbf{w}^T(\Sigma_2-\Sigma_1)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}} + \frac{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}) \\ \Leftrightarrow \quad & max(\frac{\mathbf{w}^T\Sigma_2\mathbf{w}}{\mathbf{w}^T(\Sigma_1+\Sigma_2)\mathbf{w}}). \end{aligned}$$

Then for feature extraction, we can construct the optimal spatial pattern as $\mathbf{W} = [\mathbf{W}1, \mathbf{W}2]$, where $\mathbf{W}1$ is several largest vectors for (7) and $\mathbf{W}2$ is several largest vectors for (8).

B. Algorithm

According to the cost function of Rayleigh coefficient of the Eq.(3) and Eq.(7),(8), we get the solution of CSP by GED which solving the Eq.(5), where $\mathbf{R}_l = [\Sigma_1, \Sigma_2]$, $\mathbf{R}_c = \Sigma_1 + \Sigma_2$, $\mathbf{w}$ is the generalized eigenvector and λ is the generalized eigenvalue. This decomposition can always be found due to symmetric and positive definiteness of the $\mathbf{R}_l, \mathbf{R}_c$. For each generalized eigenvector $\mathbf{w}$, we can obtain the corresponding generalized eigenvalue which are always strictly positive. Therefore, from the generalized eigenvalue equation (5), if $\mathbf{R}_c = \mathbf{I}$, then it reduces to Rayleigh quotient and generalized eigenvalue problem will be simplified to PCA. Then, let's left multiply Eq.(5) using $\mathbf{R}_c^{-1}$ and rearrange it, we get

$$\mathbf{w} = \frac{\mathbf{w}^T\mathbf{R}_c\mathbf{w}}{\mathbf{w}^T\mathbf{R}_l\mathbf{w}}\mathbf{R}_c^{-1}\mathbf{R}_l\mathbf{w}. \tag{9}$$

Eq.(9) is the basis of our iterative algorithm which can solve the CSP by online adaptive solution. Let the weight vector at iteration $(n-1), \mathbf{w}(n-1)$ be the estimate of the principal weight vector. Then, the estimate of the new weight vector at iteration n according to (9) is

$$\mathbf{w}(n) = \frac{\mathbf{w}^T(n-1)\mathbf{R}_c(n)\mathbf{w}(n-1)}{\mathbf{w}^T(n-1)\mathbf{R}_l(n)\mathbf{w}(n-1)}\mathbf{R}_c^{-1}(n)\mathbf{R}_l(n)\mathbf{w}(n-1). \tag{10}$$

It is obvious that (10) tracks the generalized eigenvalue equation at every time step just like the RLS update rule that tracks the Wiener solution at every time step [8]. By using Sherman-Morrison-Woodbury matrix inversion lemma [9] and making further simplifications, we get,

$$\Sigma^{-1}(n) = \Sigma^{-1}(n-1) - \frac{\Sigma^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^T(n)\Sigma^{-1}(n-1)}{1+\mathbf{x}^T(n)\Sigma^{-1}(n-1)\mathbf{x}(n)}. \tag{11}$$

$$\Sigma(n) = \frac{n-1}{n}\Sigma(n-1) + \frac{1}{n}\mathbf{x}\mathbf{x}^T. \tag{12}$$

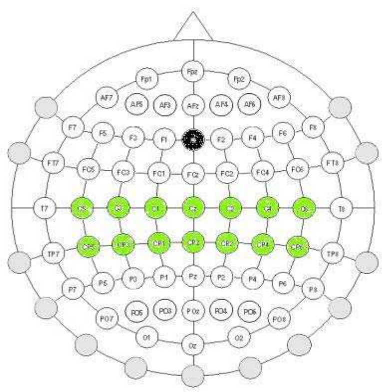

Fig. 1. Electrode placement: Electrodes are placed in arrays over the primary motor cortex, black electrode denotes the ground electrode and the green electrodes were used for the analysis.

C. Extraction of Minor Components

Equation (10) gives us the first spatial pattern. For the minor components, we can use the standard deflation procedure. Consider the following pair of matrices, $\hat{\mathbf{R}}_l = \left[\mathbf{I} - \frac{\mathbf{R}_l\mathbf{w}_1\mathbf{w}_1^T}{\mathbf{w}^T\mathbf{R}_l\mathbf{w}}\right]\mathbf{R}_l, \hat{\mathbf{R}}_c = \mathbf{R}_c$, where $\mathbf{w}_1$ is the best estimate of the first eigenvector using (10). For this pair of matrices, $\hat{\mathbf{R}}_l\mathbf{w}_1 = 0$ and $\hat{\mathbf{R}}_l\mathbf{w}_i = \lambda_i\mathbf{R}_c\mathbf{w}_i, i > 1$. The time index n is implicit and is omitted for convenience.

$$\begin{aligned}
\hat{\mathbf{R}}_l &= \left[\mathbf{I} - \frac{\mathbf{R}_l\mathbf{w}_1\mathbf{w}_1^T}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1}\right]\mathbf{R}_l \\
&= \mathbf{R}_l - \frac{\mathbf{R}_l\mathbf{w}_1\mathbf{w}_1^T\mathbf{R}_l}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1} \\
&= \mathbf{R}_l - 2\frac{\mathbf{R}_l\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1}\mathbf{w}_1^T\mathbf{R}_l + \frac{\mathbf{R}_l\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1}\mathbf{w}_1^T\mathbf{R}_l \\
&= E\left[\left(\mathbf{x}_l - \frac{\mathbf{R}_l\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1}\mathbf{y}_l\right)\left(\mathbf{x}_l^T - \frac{\mathbf{w}_1^T\mathbf{R}_l}{\mathbf{w}_1^T\mathbf{R}_l\mathbf{w}_1}\mathbf{y}_l\right)\right] \\
&= E[\hat{\mathbf{x}}_l\hat{\mathbf{x}}_l{}^T],
\end{aligned} \tag{13}$$

Where $\mathbf{y}_l = \mathbf{w}_1^T\mathbf{x}_l(n), \hat{\mathbf{x}}_l = \mathbf{x}_l - \frac{\mathbf{R}_1\mathbf{w}_1}{\mathbf{w}_1^T\mathbf{R}_1\mathbf{w}_1}\mathbf{y}_l(n)$. Using the above relation for doing on-line deflation, the update rule for the second eigenvector is given by (10). Note that the deflation given by (13) does not increase the complexity of the algorithm because all the terms in (13) are pre-computed. Other minor components can be obtained in the same manner.

III. Experiment results

The EEG singals were recorded with a portable g.tec amplifier using a cap with 64 integrated scalp electrodes located at standard positions of the international 10-20 system. The 14 channels of EEG positioned on the head over the primary motor cortex (see Fig.1) were recorded with a common reference on the left and right mastoid. The ground electrode was located on the forehead (Fz). The sampling rate is set to be 256Hz. The EEG signals were band-pass filtered between 2 and 30 Hz, and 50 Hz notch filter was applied to remove AC artifacts. All electrodes are standard Ag/AgCI and electrode impedances are checked throughout the experiment and kept below 10kΩ.

The procedure of off-line experiment are depicted as below: The subjects sat in a reclining chair looking at the center of a monitor placed about 100 cm before them. The training process consisted of the repetitive epoches of triggered movement imagery trials. Each trial started with the full black screen, at 3 s an visual cue was displayed at the center of the monitor for 4 s, representing the mental task to perform. Depending on the symbol (left arrow, right arrow) presented, the subject was instructed to perform different task: imaging a movement of left hand, right hand. The trail was ended after 4 s imagination and a blank screen was shown until the beginning of the next trial. The mental tasks represented by visual cue were chosen randomly to avoid adaptation.

The algorithm is tested with EEG time series based on motor imagery of left hand and right hand. For comparison purpose, we use batch based CSP and ICSP to train the spatial pattern individually on the whole off-line data.

Fig.2 shows the results of CSP and ICSP. The upper row of each figure shows spatial pattern of first component and second component for left, and the lower row shows the same results for right. As we can see, in this situation, the CSP and ICSP get the same results only slightly different in the minor component. Based on this, it can be proved that the

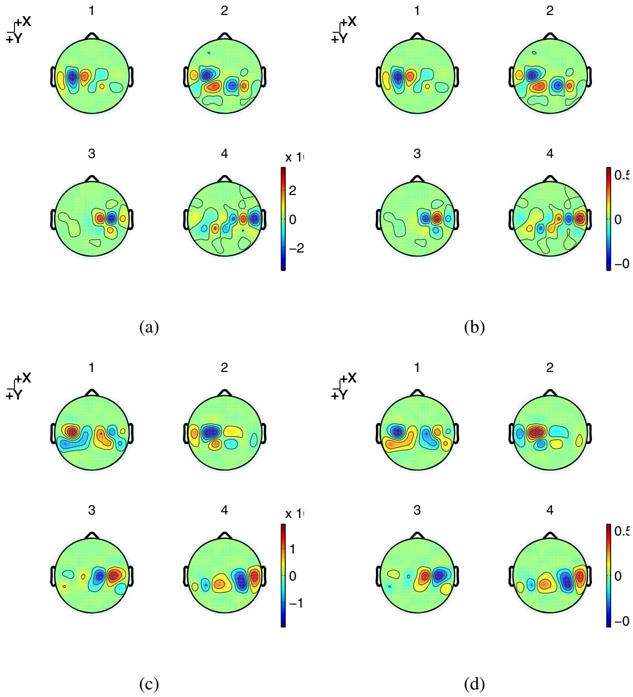

Fig. 2. Four CSP patterns with first two for the left and the other two for the right. (a)(b) CSP and ICSP spatial filter for subject A. (c)(d)CSP and ICSP spatial filter for subject B

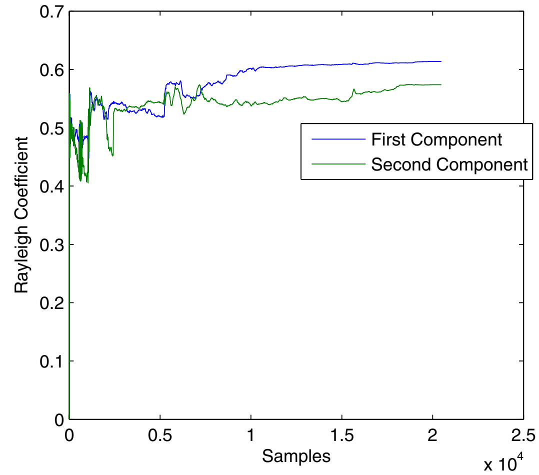

Fig. 3. Rayleigh tracks for first component and second component using ICSP method

ICSP can get the same spatial pattern with the online adaptive mode. Fig.3 illustrates the procedure of ICSP running on the EEG data. As the incremental learning proceeds, after every new sample comes, the spatial components will be updated based on iterative algorithm. It shows the Rayleigh tracks of online iteration. As can be seen that the Rayleigh converged after half data samples. If we learn the ICSP on the whole data, both CSP and ICSP have the same Rayleigh and same spatial components. Therefore, we can use ICSP to update the spatial components on-line which will be suitable for the non-stationary EEG signals.

IV. Conclusion

We have proposed a novel formula of incremental CSP method which trains the common spatial patterns on-line. The rule extracts the first CSP component and the minor components are estimated using an on-line deflation procedure. The proposed formula can be applied for adapting a common spatial pattern which is trained on a block of recording data and being applied to sequence of blocks which possibly has different distribution with on-line adjusting the spatial patterns. The proposed algorithm uses the incremental learning method for adaptive computation of CSP which has lower computational cost compared to re-training the whole data, and make it more suitable for development of on-line BCI system. Detailed analysis with useful application of on-line BCI and comparison with fixed CSP pattern will be provided in a later paper.

Acknowledgment

The work was supported by the National High-Tech Research Program of China (Grant No. 2006AA01Z125) and the National Basic Research Program of China (Grant No. 2005CB724301).

References

[1] J. R. Wolpaw, N. Birbaumer, *et al.*, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, March 2002.

[2] G.Pfurtscheller, Ch.Neuper, *et al.*, "Eeg-based discrimination between imagination of right and left hand movement," *Electroencephalography and clinical Neurophysiology*, vol. 103, pp. 342–651, 1997.

[3] H.Ramoser, J.Müller-Gerking, and G.Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 40, pp. 441–446, Dec 2000.

[4] G.Dornhege, B.Blankertz, and G.Curio, "Speeding up classification of muli-channel brain-computer interface:common spatial patterns for slow cortical potentials," in *Proceeding of the 1st International IEEE EMBS Conference on Neural Engineering.Capri 2003*, 2003, pp. 591–594.

[5] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao, and F. Yang, "BCI competition 2003-data set IV: An algorithm based on CSSD and FDA for classifying single-trial EEG," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1081–1086, 2004.

[6] Y. Li and C. Guan, "A Semi-supervised SVM Learning Algorithm for Joint Feature Extraction and Classification in Brain Computer Interfaces," *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pp. 2570–2573, 2006.

[7] R. Tomioka, J. Hill, B. Blankertz, and K. Aihara, "Adapting Spatial Filtering Methods for Nonstationary BCIs," *transformation*, vol. 10, p. 1.

[8] Y. Rao and J. Principe, "An RLS type algorithm for generalized eigendecomposition," *Neural Networks for Signal Processing XI, 2001. Proceedings of the 2001 IEEE Signal Processing Society Workshop*, pp. 263–272, 2001.

[9] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.