Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting

Haiping Lu*, *Member, IEEE*, How-Lung Eng, *Member, IEEE*, Cuntai Guan, *Senior Member, IEEE*, Konstantinos N. Plataniotis, *Senior Member, IEEE*, and Anastasios N. Venetsanopoulos, *Fellow, IEEE*

***Abstract*—Common spatial pattern (CSP) is a popular algorithm for classifying electroencephalogram (EEG) signals in the context of brain–computer interfaces (BCIs). This paper presents a regularization and aggregation technique for CSP in a small-sample setting (SSS). Conventional CSP is based on a sample-based covariance-matrix estimation. Hence, its performance in EEG classification deteriorates if the number of training samples is small. To address this concern, a regularized CSP (R-CSP) algorithm is proposed, where the covariance-matrix estimation is regularized by two parameters to lower the estimation variance while reducing the estimation bias. To tackle the problem of regularization parameter determination, R-CSP with aggregation (R-CSP-A) is further proposed, where a number of R-CSPs are aggregated to give an ensemble-based solution. The proposed algorithm is evaluated on data set IVa of BCI Competition III against four other competing algorithms. Experiments show that R-CSP-A significantly outperforms the other methods in average classification performance in three sets of experiments across various testing scenarios, with particular superiority in SSS.**

***Index Terms*—Aggregation, brain–computer interface (BCI), common spatial pattern (CSP), electroencephalogram (EEG), generic learning, regularization, small sample.**

I. Introduction

NOWADAYS, electroencephalography (EEG) signal classification is receiving increasing attention in the biomedical engineering community [1]. EEG captures the electric field generated by the central nervous system. Due to its simplicity, inexpensiveness, and high temporal resolution, it is widely used in noninvasive brain–computer interfaces (BCI) [2], [3], where brain activity is translated into sequences of control commands that enable a subject, such as a disable person, to communicate to a device, such as a computer, without using the peripheral nervous system [2]. In noninvasive EEG-based BCI, the study of motor imagery is of particular interest. It is measurable as potential changes in EEG signals, the event-related desynchronization/synchronization (ERD/ERS) patterns. EEG has also been an important tool in epilepsy diagnosis [4] for seizure detection, classification, and localization.

EEG records brain activities as multichannel time series from multiple electrodes placed on the scalp of a subject. However, recorded multichannel EEG signals typically have very low signal-to-noise ratio (SNR) [2], and they are not directly usable in BCI applications. One of the most effective algorithms for EEG signal classification is the common spatial pattern (CSP) algorithm, which extracts spatial filters that encode the most discriminative information [5]–[8]. CSP was first introduced for the binary classification of EEG trials in [5]. It is designed to find spatial projections that maximize the power/variance ratios of the filtered signals for two classes. Its calculation is through a simultaneous diagonalization of the covariance matrices of two classes. Usually, only the first few most discriminative filters are needed for classification.

This paper focuses on EEG signal classification in a small-sample setting (SSS). There are two motivations for this problem. On one hand, this SSS problem often arises in practical EEG signal-classification problem, when there is only a small training set with limited number of trials available. It should be noted that although a large number of data points can be sampled from a trial with sufficiently high frequency, these data points are highly dependent. Generally, they are not representative enough for EEG signal classification and a large number of trials are still preferred for reliable classification performance. On the other hand, as the user usually has to perform a tedious calibration measurement before starting the BCI feedback applications, one important objective in BCI research is to reduce the number of training trials needed (and the time needed) for a specific task [9]. Since the conventional CSP algorithm is based on sample-based covariance-matrix estimation, the accuracy of the estimation will be affected significantly if there is only a small training set.

The problem due to SSS in classification is common in many other applications. Regularization was first introduced to tackle the small-sample problem for linear and quadratic discriminant analysis in the regularized discriminant analysis (RDA) [10]. It was pointed out in [10] that a small number of training

Manuscript received June 8, 2010; revised August 14, 2010; accepted September 6, 2010. Date of publication September 30, 2010; date of current version November 17, 2010. *Asterisk indicates corresponding author*.

*H. Lu is with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: hplu@ieee.org).

H.-L. Eng and C. Guan are with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: hleng@i2r.a-star.edu.sg; ctguan@i2r.a-star.edu.sg).

K. N. Plataniotis is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada (e-mail: kostas@comm.utoronto.ca).

A. N. Venetsanopoulos is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3 Canada (e-mail: tasvenet@gwemail.ryerson.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2010.2082540

samples tends to result in a biased estimation of eigenvalues. On the other hand, sample-based covariance estimates from these poorly posed problems are usually highly unreliable. Two regularization parameters were introduced by Friedman [10] to account for these undesirable effects. Recently, the regularization technique has been adopted to tackle small-sample problems in various applications such as face recognition [11]–[13] and gait recognition [14].

This paper studies the regularization of the CSP algorithm in SSS. A regularized CSP (R-CSP) algorithm is proposed to regularize the covariance-matrix estimation in CSP extraction. Two regularization parameters are adopted, as in [10]. The first regularization parameter controls the shrinkage of a subject-specific covariance matrix toward a more generic covariance matrix to lower the estimation variance. This is built upon the generic learning principle in [15]. The second regularization parameter controls the shrinkage of the sample-based covariance-matrix estimation toward a scaled identity matrix to account for the bias due to the limited number of samples. Furthermore, the problem of regularization parameter determination needs to be addressed for R-CSP. However, in SSS, the number of samples may not be large enough for determining regularization parameters by the commonly used cross-validation method [10]. Thus, the aggregation strategy introduced in [12] for tensor object recognition is adopted for regularization parameter determination in EEG signal classification through R-CSP, where a number of differently regularized CSPs are combined to give an ensemble-based solution. The experimental evaluation is performed on data set IVa from BCI Competition III. The proposed algorithm outperforms four other competing CSP-based algorithms across a wide range of testing scenarios with more advantage in SSS.

There are several other versions of regularized spatial filters in the literature. The adaptive spatial filter in [16] replaced the information used in training by *a priori* information for more robust performance by considering various artifacts in EEG signals. Heuristic parameter selection was used in [16]. The invariant CSP in [17] incorporated neurophysiological prior knowledge in covariance-matrix estimation to alleviate the nonstationarities in EEG signals, and the regularization parameter was determined by cross validation. A method of logistic regression with dual spectral regularization (LRDS) was introduced in [18] for EEG classification, where cross validation was employed for parameter selection too. A composite CSP was proposed in [8] to transfer information from other subjects to a subject of interest with fewer training samples in order to boost the performance in SSS. Ten values for the regularization parameter are tested, with the average results reported. Lately, the spatially regularized CSP (SRCSP) [19] is proposed to include spatial *a priori* in the learning process by penalizing spatially nonsmooth filters with a regularization term. Another recent work that involves regularization for EEG analysis is the regularized discriminative framework in [20].

The main contributions of this work are as follows.

1) The introduction of an R-CSP algorithm for EEG signal classification, which was first reported in a preliminary version in [21]. It should be noted that while there have been several approaches proposing variations of CSP through more robust covariance-matrix estimation [8], [16]–[18], [22], none has considered the effects of small training-sample set on the eigenvalues of the covariance matrices, as discussed previously. Thus, this paper complements existing works on CSP extensions by addressing this important problem frequently arising in practice. It also has a positive impact in data collection effort, processing efficiency, and memory/storage requirement in applications involving EEG signal classification, since now much fewer training samples are needed for the same level of performance.
2) The proposal of an aggregation solution for the problem of regularization parameter determination in R-CSP, where the commonly-used cross-validation scheme [17], [18] may not be applicable in SSS. This solution adopts the principles introduced in uncorrelated multilinear discriminant analysis [12], [23] for tensor object recognition to CSP extraction in EEG signal classification. This is a significant further progress from the preliminary publication in [21], where the regularization parameter determination is not solved and only a feasibility study was provided. In contrast, the composite CSP in [8] did not address the problem of regularization parameter determination.
3) A detailed study on EEG signal classification in SSS that considers 2–120 trials per condition, including extreme small number of trials (2–10) in contrast to the recent literature [3] that considers 10–100 trials per condition. This study consists of 1500 experiments in total in order to study the statistical significance of the obtained results. Another two sets of experiments are carried out for performance evaluation against four competing solutions. Based on the simulations, insightful observations have been made regarding the proposed algorithm. This is also a significant development from the earlier publication [21].

The rest of this paper is organized as follows. Section II presents the R-CSP algorithm for EEG signal classification. In Section III, the problem of regularization parameter determination is discussed, and an aggregation solution is formulated for R-CSP. Section IV provides an experimental study of the EEG signal-classification problem in SSS and an evaluation of the proposed algorithm. Finally, Section V draws the conclusion.

II. Regularized CSP for EEG Signal Classification

This section presents the R-CSP algorithm for classification of EEG signals. Regularized covariance-matrix estimation is used in R-CSP by employing the regularization technique introduced in [10] and the generic learning principle in [15]. The EEG classification scheme of R-CSP follows that in the conventional CSP algorithm [5].

A. Sample-Based Covariance Matrix in CSP

In CSP-based EEG signal classification, a matrix $\mathbf{E}$ of size $N \times T$ is used to represent a single N-channel EEG trial, with T samples in each channel for a single trial. The sample covariance matrix of a trial $\mathbf{E}$ is normalized with the total variance as [5] follows:

$$\mathbf{S} \quad \frac{\mathbf{E}\mathbf{E}^T}{\mathbf{E}\mathbf{E}^T} \tag{1}$$

where the superscript "T" denotes the transpose of a matrix, and $\cdot$ denotes the trace of a matrix. This paper considers only binary classification problems and the two classes are indexed by $c \quad \{\ ,\ \}$. For simplicity, it is assumed that M trials are available for training in each class for a subject of interest, indexed by m as $\mathbf{E}_{c,m}$, $m \quad ,\ldots,M$. Hence, each trial has a covariance matrix $\mathbf{S}_{c,m}$ and the average spatial covariance matrix is then calculated for each class as [5] follows:

$$\mathbf{S}_c \quad \frac{}{M}\sum_{m}^{M}\mathbf{S}_{c,m}, \qquad c \quad \{\ ,\ \}. \tag{2}$$

The discriminative spatial patterns in CSP are extracted based on this sample-based covariance-matrix estimation. When there are only a small number of training trials, such an estimation problem could be poorly posed [10] and the estimated parameters could be highly unreliable, giving rise to high variance. Moreover, the low SNR for EEG signals makes the estimation variance even higher.

B. Regularized Covariance-Matrix Estimation in SSS

Regularization technique has been proved to be effective in tackling the small-sample problem. It works by biasing the estimates away from their sample-based values toward more "physically plausible" values [10], which reduces the variance of the sample-based estimates while tending to increase bias. This bias-variance trade-off is commonly controlled by one or more regularization parameters [10].

As in [10], the proposed R-CSP calculates the regularized average spatial covariance matrix for each class as follows:

$$\boldsymbol{\Sigma}_c\ \beta,\gamma \qquad -\gamma\ \boldsymbol{\Omega}_c\ \beta \quad \frac{\gamma}{N} \quad \boldsymbol{\Omega}_c\ \beta\ \cdot \mathbf{I} \tag{3}$$

where β and γ are two regularization parameters ($\leq \beta,\gamma \leq$), and $\mathbf{I}$ is an $N \times N$ identity matrix. $\boldsymbol{\Omega}_c\ \beta$ comprises covariance matrices for the trials from the specific subject as well as generic trials from other subjects. It is defined as follows:

$$\boldsymbol{\Omega}_c\ \beta \quad \frac{-\beta\ \cdot \mathbf{S}_c \quad \beta \cdot \mathbf{S}_c}{-\beta\ \cdot M \quad \beta \cdot M} \tag{4}$$

where $\mathbf{S}_c$ is the sum of the sample covariance matrices for all M training trials in class c

$$\mathbf{S}_c \quad \sum_{m}^{M}\mathbf{S}_{c,m} \tag{5}$$

and $\mathbf{S}_c$ is the sum of the sample covariance matrices for M generic training trials $\{\mathbf{E}_{c,m}\}$ from other subjects in class c

$$\mathbf{S}_c \quad \sum_{m}^{M}\mathbf{S}_{c,m}. \tag{6}$$

In these definitions, $\mathbf{S}_{c,m}$ and $\mathbf{S}_{c,m}$ are the normalized sample covariance matrix defined in (1).

The term $\mathbf{S}_c$ introduced in (4) aims to reduce the variance in the covariance-matrix estimation, and it tends to produce more reliable results. This is built upon the idea of generic learning for the one-training-sample case in face recognition [15], and it also embodies the same principle as that in [8] and [17]. For the classification of EEG signals from a particular subject, the proposed training process constructs the regularization term $\mathbf{S}_c$ using corresponding EEG trials collected from some other subjects, i.e., generic trials from the population. When there are S subjects available in total, each with M trials for each class, $M \quad S- \quad \times M$.

The rationales of the regularization scheme in (4) follow those in [10]. The first regularization parameter β controls the degree of shrinkage of the training-sample covariance-matrix estimates to the pooled estimate, which is to reduce the variance of the estimates. The second regularization parameter γ controls the degree of shrinkage toward a multiple of the identity matrix, with the average eigenvalue of $\boldsymbol{\Omega}_c\ \beta$ as the multiplier. This second shrinkage has the effect of decreasing the larger eigenvalues while increasing the smaller ones. This is because the sample-based estimates in (1) tend to bias the eigenvalues in the opposite direction, especially in SSS [10]. Thus, γ is to counteract such bias due to the limited number of samples. From the aforementioned, the conventional CSP can be considered as a special case of R-CSP, i.e., when $\beta \quad \gamma$. In addition, the composite CSP introduced in [8] could be considered as a special case of R-CSP with γ .

The effects of the adopted regularization scheme are illustrated in Figs. 1 and 2 with some typical examples. In the figures, the first 20 largest eigenvalues of a typical average spatial covariance matrix are shown in descending order with magnitudes in log scale. Fig. 1 depicts the eigenvalue distribution without regularization and with a regularization $\beta \quad .\ (\gamma \quad)$ for five randomly selected training sets with M for the same class of a particular subject. It is observed that the variance of the eigenvalues are much higher when there is no regularization by β. Fig. 2 simply shows the eigenvalue distribution without regularization and with a regularization $\gamma \quad .\ (\beta \quad)$ for a fixed training set with M . It can be seen that the regularization by γ decreases the relative magnitudes of the larger eigenvalues over those of the smaller eigenvalues, which reduces the bias due the small number of training samples in turn.

C. R-CSP Feature Extraction and Classification

With the formulation of the regularized covariance-matrix estimation in SSS, feature extraction in R-CSP follows that in the classical CSP method [5]. The regularized composite spatial covariance is formed and factorized as follows:

$$\boldsymbol{\Sigma}\ \beta,\gamma \quad \boldsymbol{\Sigma}\ \beta,\gamma \quad \boldsymbol{\Sigma}\ \beta,\gamma \quad \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T \tag{7}$$

where $\mathbf{U}$ denotes the matrix of eigenvectors, and $\boldsymbol{\Lambda}$ denotes the diagonal matrix of corresponding eigenvalues. This paper adopts the convention that the eigenvalues are sorted in descending order. The full projection matrix is then formed as follows:

$$\mathbf{W} \quad \mathbf{B}^T\boldsymbol{\Lambda}^{-\ /\ }\mathbf{U}^T \tag{8}$$

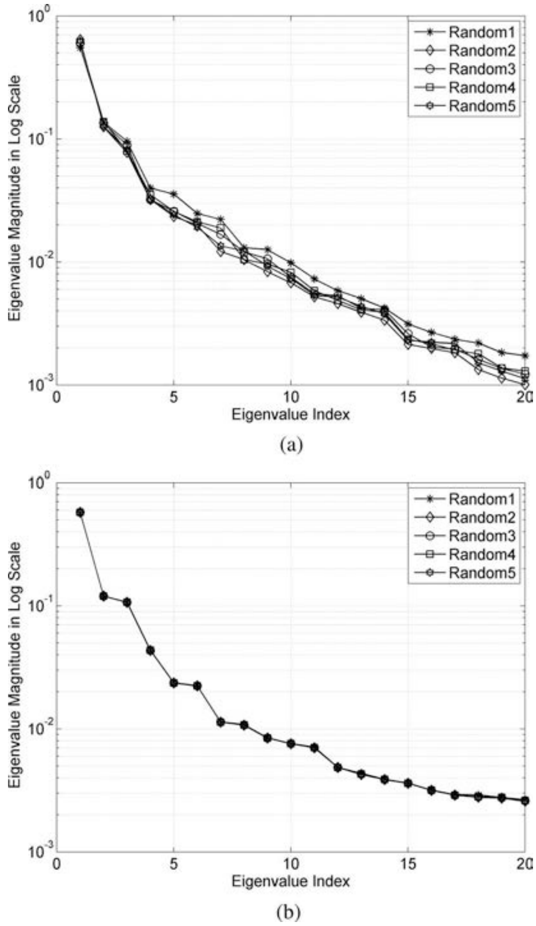


Fig. 1. Illustration of the effects of the regularization parameter β on five random training sets with M for the same class of a particular subject. (a) Eigenvalue distribution of a typical average covariance matrix without regularization by β. (b) Eigenvalue distribution of a typical average covariance matrix with regularization β . .

where $\mathbf{B}$ denotes the matrix of eigenvectors for the whitened spatial covariance matrix $\Lambda^{-\ /}\ \mathbf{U}^T\Sigma\ \ \beta,\gamma\ \mathbf{U}\Lambda^{-\ /}$.

To get the most discriminative patterns, the first and last α columns of $\mathbf{W}$ are retained to form an $N \times Q$ matrix $\mathbf{W}$ with $Q \quad \alpha$. In R-CSP feature extraction, an input trial $\mathbf{E}$ is first projected as follows:

$$\mathbf{X} \quad \mathbf{W}^T\mathbf{E}. \tag{9}$$

A Q-dimensional feature vector $\mathbf{y}$ is then constructed from the variance of the rows of $\mathbf{X}$ as follows:

$$y_q \qquad \left(\frac{\mathbf{x}_q}{\sum_q^Q \qquad \mathbf{x}_q}\right) \tag{10}$$

where y_q denotes the qth component of $\mathbf{y}$, $\mathbf{x}_q$ denotes the qth row of $\mathbf{X}$, and $\mathbf{x}_q$ denotes the variance of $\mathbf{x}_q$.

Finally, R-CSP classification in this paper employs the Fisher's discriminant analysis (FDA) followed by the simple nearest-neighbor classifier (NNC). FDA solves for a projection $\mathbf{v}$ to maximize the ratio of the between-class scatter to the within-class scatter

$$\mathbf{v} \qquad {}_{\mathbf{v}} \frac{\mathbf{v}^T\Psi_B\mathbf{v}}{\mathbf{v}^T\Psi_W\mathbf{v}} \tag{11}$$

where Ψ_B and Ψ_W are the between-class scatter matrix and the within-class scatter matrix [24] for the features $\mathbf{y}$ in (10), respectively. This problem can be solved as a generalized eigenvalue problem [25] and the discriminant feature vector $\mathbf{z}_m$ is obtained as follows:

$$\mathbf{z} \quad \mathbf{v}^T \quad \mathbf{y}. \tag{12}$$

In NNC classification, the nearest neighbor is found as μ^* $_{\mu}\ \|\mathbf{z}-\mathbf{z}_\mu\|$, where $\mathbf{z}_\mu$ is the feature vector for the μth training trial, μ , , . . . , M, and $\|\cdot\|$ is the Euclidean norm for vectors. The class label of the μ^*th training sample c_{μ^*} is then assigned to the test trial $\mathbf{E}$.

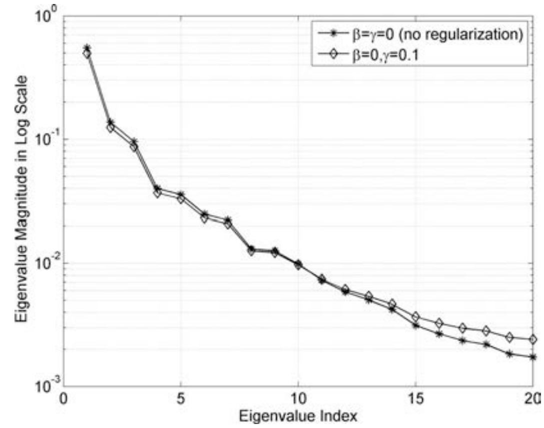


Fig. 2. Illustration of the effects of regularization parameter γ on a fixed training set: the eigenvalue distribution of a typical average spatial covariance matrix with and without regularization by γ.

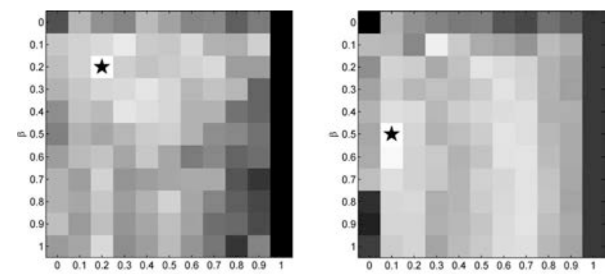


Fig. 3. Two examples showing the variation of classification accuracy (coded as the gray levels in the displayed checkerboard) for 121 pairs of regularization parameters β and γ. The pair resulting in the highest classification accuracy is marked with a black star.

Fig. 3 gives two examples on the variation of classification accuracy for 121 pairs of regularization parameters β and γ. The classification accuracy is coded as the gray levels (white for the highest and black for the lowest) in the displayed checkerboard. A black star is used to mark the highest classification accuracy in each example. The effectiveness of both β and γ is observed in the figure. At the same time, it could be seen that the classification accuracy could be sensitive to parameter settings, and determining the optimal pair of regularization parameters is a challenging problem.

III. Aggregation of R-CSPs

As pointed out at the end of Section II, there is one important problem remaining for the proposed R-CSP algorithm, i.e., the problem of regularization parameter determination, which is a model selection problem [10]. This problem is important since it is unlikely to know what values are good for the two regularization parameters in advance, as illustrated in Fig. 3. In earlier work [21], 121 regularization parameter combinations were tested and the best result for each case was reported, which is a close-set optimization scheme. Consequently, the evaluation is not a fair one. This section proposes an aggregation solution to this problem.

Traditionally, the problem of parameter determination is solved through cross-validation (and the overall assessment, such as generalization error estimation is performed by nested cross validation) [10], [17], [18], which is a sample-based estimation method. Typically, one round of cross validation partitions a sample set of data into complementary subsets. Analysis is performed on one subset (the training set), and the other subset (the validation set) is used to validate the analysis. Usually, several rounds of cross validation are needed using different partitions to reduce variability.

The cross-validation method has been employed in our study to determine the regularization parameters of R-CSP for EEG signal classification in SSS. However, the R-CSP determined this way could perform worse than the conventional CSP algorithm in some cases. The main cause is that in SSS, there may be insufficient number of samples for the construction of the training and validation subsets to get reliable estimates of the regularization parameters. For example, in the case that only three samples are available for each class per subject, only one sample can be used for the training, validation, and testing, respectively. When only two samples per class from a subject are available for training and testing, there is no data for validation except the testing data, therefore, cross validation can not be performed. Therefore, the cross-validation scheme may not always be applicable to parameter determination of R-CSP for EEG signal classification in SSS.

Based on the aforementioned study, this paper adopts the technique of aggregation for regularization parameter determination developed in face-recognition and gait-recognition applications [12], resulting in R-CSP with aggregation (R-CSP-A). In R-CSP-A, instead of using a single pair of regularization parameters from the interval $[0, 1]$, a fixed set of regularization parameter pairs is used and the results from differently regularized CSPs are combined to form an aggregated solution. This approach embodies the principle of ensemble-based learning. Since different regularization parameter pair will result in different discriminative features, such diversity is good for ensemble-based learning, based on the generalization theory explaining the success of boosting [26]–[28]. As for the combination scheme, there are various ways including the feature-level fusion [29], the matching score-level fusion [30], [31], and more advanced ensemble-based learning, such as boosting [26], [32], [33]. In this paper, the simple sum rule for matching score fusion is employed, as in [12].

Input: A set of M EEG trials $\mathbf{E}_{(c,m)_{(s)}}$ for each class of S subjects, where $c = \{1, 2\}$, $m = 1, 2, ..., M$, and $s = 1, ..., S$. A test trial $\mathbf{E}$ for subject s^*, A pairs of β and γ, the number of most discriminative columns from the full projection matrix $Q = 2\alpha$. Subject s^* is considered as the subject of interest and other subjects with $s \neq s^*$ are considered as the generic data.

Output: The class label for $\mathbf{E}$.

R-CSP-A algorithm:

Step 1. Feature extraction

- Obtain $\mathbf{S}_{(c,m)}$ for all subjects $s = 1, ..., S$ according to (1).
- Form $\mathbf{S}_c$ for subject s^* according to (5) and form $\hat{\mathbf{S}}_c$ from other subjects $s \neq s^*$ according to (6).
- For $a = 1 : A$
  - Follow (4), (3), (7), and (8) to get the full projection matrix.
  - Retain the first and last α columns of the full projection matrix to get $\hat{\mathbf{W}}_{(a)}$.
  - Follow (9) and (10) to obtain the feature vector $\hat{\mathbf{y}}_{(a)}$.

Step 2. Aggregation at the matching score level for classification

- For $a = 1 : A$
  - Apply (11) and (12) on $\hat{\mathbf{y}}_{(a)}$ to get $\mathbf{z}_{(a)}$.
  - For $c = 1 : 2$
    - $*$ Obtain the nearest-neighbor distance $d(\mathbf{E}, c, a)$.
  - Normalize $d(\mathbf{E}, c, a)$ to $[0, 1]$ to get $\tilde{d}(\mathbf{E}, c, a)$.
- Obtain the aggregated distance $d(\mathbf{E}, c)$.
- **Output** $c^* = \arg\min_c d(\mathbf{E}, c)$ as the class label for the test sample $\mathbf{E}$.

Fig. 4. Pseudo-code implementation of the R-CSP-A algorithm for EEG signal classification in SSS.

Fig. 4 provides the pseudocode implementation of R-CSP-A for EEG signal classification in SSS, where $s = 1, \ldots, S$ is used to index the S subjects, each with M trials for each class. In feature extraction, the input trials $\mathbf{E}_{(c,m)_{(s)}}$, $c = \{1, 2\}$, $m = 1, 2, \ldots, M$, and $s = 1, \ldots, S$ are fed into A differently regularized CSP feature extractors with parameters β_a and γ_a to obtain a set of A different feature vectors $\hat{\mathbf{y}}_{(a)}$.

In classification, FDA is applied to $\hat{\mathbf{y}}_{(a)}$ to get $\mathbf{z}_{(a)}$ for NNC. For each a, the nearest-neighbor distance of the test trial $\mathbf{E}$ to each candidate class c is calculated as [12] follows:

$$d(\mathbf{E}, c, a) = \min_{\mu, c_\mu = c} \| \mathbf{z}_{(a)} - \mathbf{z}_{\mu(a)} \|. \quad (13)$$

The range of $d(\mathbf{E}, c, a)$ is then matched to the interval $[0, 1]$ as [32] follows:

$$\tilde{d}(\mathbf{E}, c, a) = \frac{d(\mathbf{E}, c, a) - \min_c d(\mathbf{E}, c, a)}{\max_c d(\mathbf{E}, c, a) - \min_c d(\mathbf{E}, c, a)}. \quad (14)$$

Finally, the aggregated nearest-neighbor distance is obtained employing the simple sum rule as [12] follows:

$$d(\mathbf{E}, c) = \sum_{a=1}^{A} \tilde{d}(\mathbf{E}, c, a). \quad (15)$$

and the test sample $\mathbf{E}$ is assigned the label: c^* $_c d$ $\mathbf{E}, c$. Since only two classes are considered in this paper, the aforementioned aggregation process is equivalent to a simple majority voting in this case. Nonetheless, the aggregation formulation here is applicable in future work for more than two classes.

In addition, it should be noted that there are other ensemble-based extensions of CSP [34]–[36]. In [35], EEG signals are decomposed into subbands, where CSP is applied to extract features from each subband, and then the subband scores are fused to give the final classification result. In the mixtures of CSP approach [36], multiple CSP feature extractors are constructed by bootstrap sampling of the training set to improve the classification performance. These two algorithms apply the ensemble-based learning principle on the training data, while the proposed R-CSP-A applies the ensemble-based learning in the feature-extraction process with fixed training data. These two approaches employ the same principle at different stages of processing, therefore, they could be combined to work together for even better classification performance. However, this is out of the scope of this paper, therefore, it is left for future research.

IV. Experimental Study

This section presents a large number of experiments carried out in support of the following objectives.

1) Investigate how the performance of EEG signal classification is affected by the number of training samples.
2) Evaluate the performance of R-CSP-A against the conventional CSP algorithm as well as other competing CSP-based algorithms on EEG signal classification.

A. Experimental Data and Design

Experiments are carried out on data set IVa of BCI Competition III [34], [37]. In each capturing session of this data set, visual cues were presented to a subject for 3.5 s with the indication of one of the three motor imageries that the subject should perform: left hand, right hand, and right foot. For the subject to relax, the cue presentation was separated by intervals with random length ranging from 1.75 to 2.25 s. Only the right hand and right foot motor imageries of five healthy subjects ("aa," "al," "av," "aw," and "ay") are provided for public use. The EEG signals were recorded with 118 electrodes located at the positions of the extended international 10/20 system. There are 140 trials for each class per subject, i.e., a total of 280 trials for each subject. All EEG signals were down sampled to 100 Hz and bandpass filtered. Thus, N and T . In addition, the first and last three columns of $\mathbf{W}$ are used for classification, i.e., Q (α), as recommended in [2] and [3]. For a subject, whose EEG signals are to be classified, the training process of R-CSP employs the corresponding EEG trials collected for other four subjects in the regularization term $\mathbf{S}_c$, e.g., the generic training trials for "ay" consist of all the trials from "aa," "al," "av," and "aw". Therefore, M . For the aggregation, the research in ensemble-based learning [32] indicates that high diversity of the learners to be combined is preferred. Thus, based on the study in [21] and the experience learnt from [12], the following six values for β and five values for γ are empirically selected in an approximately even logarithmic scale to cover a wide range, ensure diversity, and also limit the number of values for computational concerns

$$\beta \in \{ \; , \; . \; , \; . \; , \; . \; , \; . \; , \; . \; \}, \quad \gamma \in \{ \; , \; . \; , \; . \; , \; . \; , \; . \; \}$$

where one more value is selected for β than for γ because the effective β values have a wider range, as seen in [21]. The aforementioned selection gives A $\times$ differently regularized CSP feature extractors, indexed by a $, \ldots, A$. This setting for R-CSP-A is used in all the experiments in the following.

Three sets of experiments are carried out as detailed in the following.

I) To study EEG signal classification in SSS, the following 15 values of M (the number of training samples per class) are tested for each subject:

$$M \in \{ \; , \; , \; , \; , \; , \; , \; , \; , \; , \; , \; , \; , \; , \; , \; \}.$$

To ensure the significance of the studies, for each class per subject, the M trials are randomly selected from the 140 trials and the rest $- M$ trials are used for testing. The reported results are the average of 20 such repeated experiments. Thus, there are $\times$ $\times$ experiments in total. For this study, the 7–30 Hz frequency band is used. Besides space limitation, as the commonly used ten-fold cross validation cannot be performed for very small values of M, only the results for conventional CSP and R-CSP-A are reported for this set of experiments.

II) To evaluate the proposed R-CSP-A algorithm against competing solutions, this set of experiments is carried out using subject-specific frequency bands that are used by the winning entry of data set IVa in BCI Competition III [38]. R-CSP-A is compared against the following five competing algorithms:

a) CSP: the conventional CSP [5];
b) LW-CSP: CSP with regularized covariance matrices determined by Ledoit and Wolf's method [39], [40];
c) LRDS: logistic regression with dual spectral regularization, where the regularization parameter is determined by ten-fold cross validation (20 parameters are tested as suggested by the authors) [18];
d) SRCSP: spatially regularized CSP with ten-fold cross validation to determine the regularization parameters (80 parameter combinations are tested as suggested by the authors) [19];
e) R-CSP-CV: the proposed R-CSP with ten-fold cross validation to determine the regularization parameters (30 parameter combinations used by R-CSP-A are tested).

To be more realistic, the first L trials are used for training and the rest are used for testing. The following ten values of L are tested for each subject:

$$\{ \; , \; , \; , \; , \; , \; , \; , \; , \; , \; \}.$$

TABLE I
EEG CLASSIFICATION PERFORMANCE FOR EXPERIMENT I, WITH CCRS IN PERCENTAGE REPORTED (MEAN ± STD)

M	Subject	aa	al	av	aw	ay	Average
2	CSP	51.3±4.0	70.0±14.8	49.1±2.0	49.7±4.5	57.4±7.7	55.5±11.0
	R-CSP-A	58.8±6.1	76.3±9.2	52.6±3.8	57.1±9.2	73.9±7.8	63.8±12.0
3	CSP	50.8±4.0	63.1±16.4	51.8±4.5	54.0±6.1	58.5±10.1	55.7±10.3
	R-CSP-A	58.6±5.4	79.1±8.3	56.2±4.6	58.9±8.4	71.5±8.4	64.9±11.3
4	CSP	52.6±4.8	77.0±11.7	50.6±4.0	57.8±5.3	56.4±9.3	58.9±12.0
	R-CSP-A	63.6±7.7	83.0±8.0	54.7±4.2	66.6±5.9	75.9±7.7	68.8±11.9
5	CSP	51.7±5.1	81.2±11.3	51.8±5.0	58.9±8.3	59.4±10.3	60.6±13.6
	R-CSP-A	62.9±5.7	85.4±5.8	56.1±5.4	65.6±6.8	78.6±5.3	69.7±12.1
6	CSP	55.0±5.9	80.4±8.2	51.6±4.1	58.3±8.7	65.0±9.9	62.1±12.6
	R-CSP-A	65.1±5.7	85.8±3.6	56.1±5.4	67.8±5.6	80.6±4.9	71.1±11.9
8	CSP	54.6±6.3	86.6±4.5	51.1±3.9	64.3±8.9	68.5±8.9	65.0±14.1
	R-CSP-A	67.7±5.6	86.6±3.8	58.8±4.4	71.1±5.7	80.1±5.5	72.9±10.9
10	CSP	56.7±5.9	85.0±5.9	52.3±4.5	64.9±8.2	76.9±8.1	67.2±13.9
	R-CSP-A	65.4±3.3	86.3±2.3	59.9±5.0	74.0±3.4	84.1±2.8	73.9±10.8
20	CSP	61.9±5.6	87.8±4.5	56.2±6.4	72.4±9.7	81.3±5.4	71.9±13.3
	R-CSP-A	71.3±3.2	89.4±3.0	62.6±6.3	75.8±3.0	86.4±1.9	77.1±10.5
30	CSP	64.4±4.8	90.0±4.6	59.5±5.1	77.4±6.9	82.0±5.1	74.7±12.4
	R-CSP-A	73.7±3.9	88.9±3.1	64.8±4.0	78.0±2.4	86.6±2.9	78.4±9.3
40	CSP	67.7±6.0	***90.4***±4.4	59.3±5.3	77.4±7.8	83.9±4.6	75.7±12.5
	R-CSP-A	74.8±2.7	***89.6***±2.2	65.3±4.3	77.1±3.5	87.6±2.4	78.9±9.4
50	CSP	69.2±5.5	88.4±4.0	59.7±5.7	83.0±6.3	83.9±4.3	76.8±11.9
	R-CSP-A	75.7±3.8	88.1±2.4	66.1±5.0	81.6±3.6	88.6±2.7	80.0±9.1
60	CSP	68.2±5.9	88.8±3.7	62.1±4.3	80.9±7.7	***86.5***±2.9	77.3±11.6
	R-CSP-A	74.6±4.3	88.5±3.3	68.5±3.5	82.1±4.3	***89.8***±1.5	80.7±8.8
80	CSP	68.6±8.0	89.6±5.0	59.7±5.1	***84.5***±9.4	85.3±4.6	77.5±13.2
	R-CSP-A	77.6±4.1	89.0±3.0	68.5±4.6	81.9±5.3	88.3±2.9	81.1±8.6
100	CSP	***71.1***±6.9	88.6±4.8	59.8±6.2	84.4±9.1	85.8±3.4	***77.9***±12.5
	R-CSP-A	***79.0***±5.1	88.7±4.1	69.7±5.8	***83.9***±4.8	89.6±3.5	***82.2***±8.6
120	CSP	69.0±9.5	88.8±6.7	***62.6***±7.7	79.1±10.7	83.9±4.4	76.7±12.4
	R-CSP-A	75.9±7.2	89.0±6.4	***71.9***±6.3	79.8±6.7	88.1±3.5	80.9±9.0
Average	CSP	60.8±9.6	83.7±11.4	55.8±6.7	69.8±14.0	74.3±13.4	68.9±15.0
	R-CSP-A	69.6±8.3	86.2±6.3	62.1±7.7	73.4±9.9	83.3±7.5	74.9±11.9

III) In addition, the EEG classification experiments are carried out in the setting of BCI Competition III for completeness, where L , , , , and for subject "aa," "al," "av," "aw," and "ay," respectively. The results for CSP, LW-CSP, LRDS, SRCSP, R-CSP-CV, and R-CSP-A are reported. Similar to experiment II, subject-specific frequency bands by the winner of data set IVa [38] are used in this set of experiments.

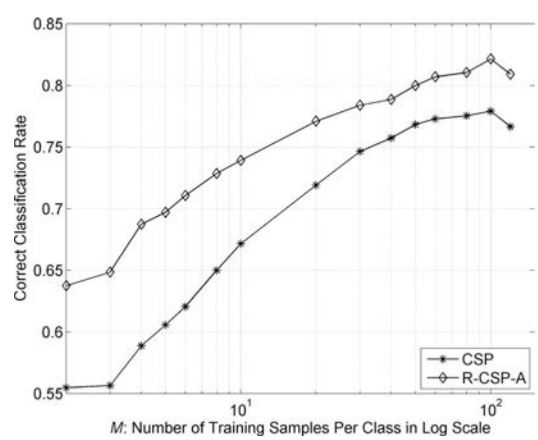
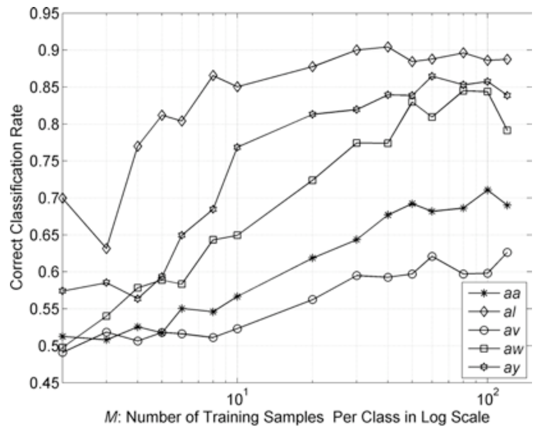


Fig. 5. Illustration of the improvement achieved over CSP by the proposed R-CSP-A algorithm in experiment I.

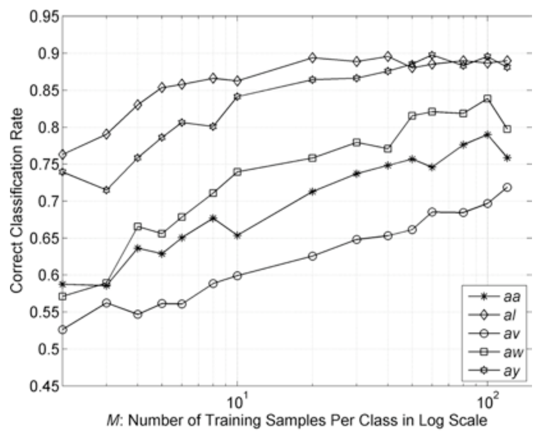
B. Experimental Results

In the following, the experimental results are presented for the experimental settings described previously. For performance evaluation, the correct classification rate (CCR) is used to measure the classification accuracy.

1) Results for Experiment I: The complete experimental results for the first set of experiments are summarized in Table I, where the mean and standard deviation (Std) of the 20 repetitions are reported for the five subjects and for the 15 values of M tested. This table also includes the average over subjects for each value of M and the average over the various Ms for each subject. In all the testing scenarios, the R-CSP-A algorithm outperforms the classical CSP algorithm, with an average improvement of 6% in CCR. This shows that the regularization scheme introduced in this paper is effective in improving the EEG-signal-classification accuracy. Furthermore, on an average, for M ranging from to , R-CSP-A outperforms CSP by 8.6%, while R-CSP-A outperforms CSP by only 3.8% for M ranging from to , indicating that R-CSP-A is particularly powerful when the number of training samples is small. In particular, for subject "ay," the average improvement in CCR is more than for M , , , and . Fig. 5 further illustrates the classification performance difference between R-CSP-A and CSP. Similar to the observations in Table I, the figure shows that the advantage of R-CSP-A over CSP is more significant for small values of M. Due to its ensemble learning nature, R-CSP-A also has lower Std than CSP on an average, as seen from the right bottom of Table I. To study the statistical significance



(a)



(b)

Fig. 6. Demonstration of EEG classification performance difference among subjects in experiment I. (a) CCRs obtained by CSP for each of the five subjects. (b) CCRs obtained by R-CSP-A for each of the five subjects.

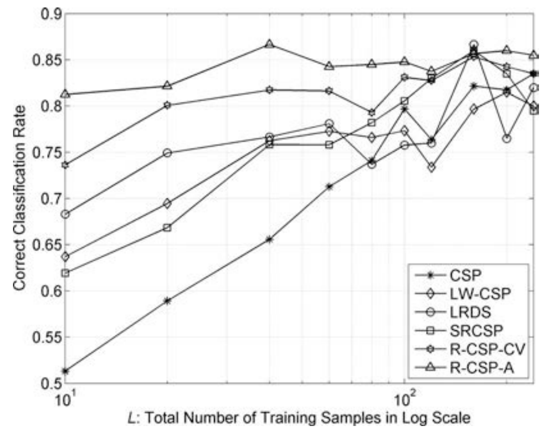


Fig. 7. EEG classification performance comparison for ten values of L, averaged over five subjects in experiment II.

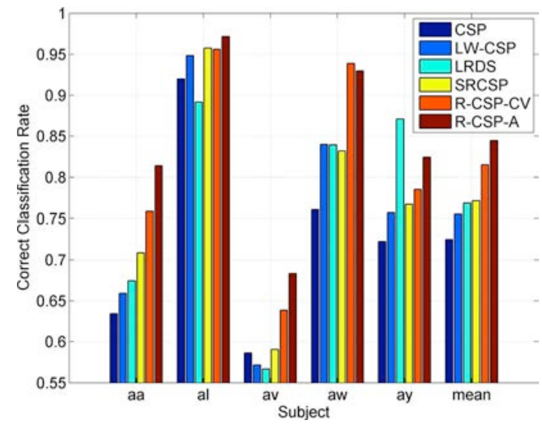


Fig. 8. EEG classification performance comparison in experiment II for the five subjects and their mean, averaged over ten values of L. (Please note that colors are used so this figure is best viewed on screen or in color print.)

of the improvement of R-CSP-A over CSP, paired t-tests were carried out for all the 1500 experiments. The p value obtained is much less than .   , indicating that the performance gain of R-CSP-A over CSP is statistically significant.

Fig. 6 plots the results for the five subjects separately. The figure demonstrates that the classification results are subject dependant. For some subjects such as "al," the classification accuracy is generally higher, while for some other subjects such as "av," the classification accuracy is generally lower. Furthermore, the classification performance does not always increase monotonically with M. In Table I, the best results for each subject and their average over various M are highlighted with italic bold fonts. For CSP, the best results for "aa," "al," "av," "aw," "ay," and their average are obtained with M , , , , , and , respectively. For R-CSP-A, the best results for "aa," "al," "av," "aw," "ay," and their average are obtained with M , , , , , and , respectively. Similar observations can be made from results reported by other researchers, e.g., Figs. 1 and 2 in [3]. This is in contrary to our common belief that better results should be obtained with more training data and the cause needs further investigation. A possible cause could be the increased number of outliers, therefore, effective outlier elimination may mitigate this problem.

2) Results for Experiment II: The results for experiment II are summarized in Figs. 7 and 8. There are a total of 300 experiments (ten experiments on five subjects for six algorithms).

Fig. 7 depicts the EEG classification performance averaged over five subjects for the ten values of L tested. From the figure, it could be seen that R-CSP-A outperforms R-CSP-CV for all averaged cases, illustrating the effectiveness of the proposed aggregation scheme over traditional cross validation. R-CSP-A also outperforms the other four algorithms (CSP, LW-CSP, LRDS, and SRCSP) in most averaged cases except for L , where LRDS obtains better results than R-CSP-A. Furthermore, the figure demonstrates again that R-CSP-A is particularly effective for small values of L. In general, R-CSP-A outperforms the other methods by a greater amount for a smaller value of L. For example, the average CCR over L , , and for CSP, LW-CSP, LRDS, SRCSP, R-CSP-CV, and R-CSP-A are . , . , . , . , . , and . , respectively. R-CSP-A has outperformed all the other methods significantly in this case, with CCR . higher than LRDS and . higher than R-CSP-CV.

TABLE II
EEG CLASSIFICATION PERFORMANCE FOR EXPERIMENT III, THE BCI COMPETITION III SETTING, WITH CCRs REPORTED IN PERCENTAGE

Algorithm	aa	al	av	aw	ay	Average
CSP	66.1	98.2	59.2	88.4	61.1	74.6
LW-CSP	69.6	**100.0**	56.6	**93.3**	67.1	77.3
LRDS	**80.4**	94.6	50.0	90.6	**83.3**	79.8
SRCSP	77.7	96.4	59.2	91.1	61.1	77.1
R-CSP-CV	77.7	96.4	58.7	92.0	68.3	78.6
R-CSP-A	76.8	98.2	**74.5**	92.9	77.0	**83.9**
BCI III Winner	95.5	100.0	80.6	100.0	97.6	94.2

Fig. 8 shows the EEG classification performance averaged over the ten values of L tested for five subjects as well as the overall mean. The advantage of R-CSP-A over R-CSP-CV is observed for all subjects except "aw" and R-CSP-A gives higher CCRs than the other four algorithms in most averaged cases except for subject "ay," where LRDS is particularly effective and gives the best results (on the other hand, the performance of LRDS is particularly poor for subject "av"). The overall average CCR for CSP, LW-CSP, LRDS, SRCSP, R-CSP-CV, and R-CSP-A are . , . , . , . , . , . , and . , respectively, as the last bar group in Fig. 8 indicates. R-CSP-A has outperformed all the other methods on average, with CCR . higher than SRCSP and higher than R-CSP-CV. Moreover, even R-CSP-CV produces better results than all the other four algorithms (. higher than SRCSP), demonstrating the effectiveness of the proposed regularization scheme for CSP and also showing that the results from the training set can be well transferred to the test set (though still inferior to the aggregation scheme).

3) Results for Experiment III: Table II reports the results for experiment III, where the classification tasks are carried out in the BCI Competition III setting. The highest CCR among the six algorithms listed in Section IV-A is highlighted in bold font for each subject and their average. On an average, R-CSP-A has outperformed the other five algorithms by at least . In this set of results, its superiority over other methods is mainly on the more difficult subject "av," and its performance on the other subjects has no significant difference over LRDS and R-CSP-CV. In the exceptional case of "aa" (with L), R-CSP-CV gives a better result than R-CSP-A, though this is not the general case, as shown in Section IV-B-2.

The CCRs by the winner for this data set in BCI Competition III are included at the bottom of Table II for easy reference. It could be seen that R-CSP-A is inferior to the winner. However, it should be noted that the winner algorithm involves an ensemble classifier based on three methods: CSP on ERD, autoregressive models on ERD, and LDA on temporal waves of readiness potential. Different methods are used for two groups of subjects with fine-tuned parameters for each subject [38]. CSP is the only method used for all subjects. Furthermore, the winner algorithm uses bootstrap aggregation and extends training samples with former classified test samples for two subjects to achieve the best performance. Thus, R-CSP-A is less complex and not subject customized compared to the winner algorithm, therefore, the performance gap is expected and it should be considered as one significant enhancement of a particular component (CSP) of the winner algorithm.

4) Discussions: One important implication from the experimental results is that the proposed algorithm has positive impact on the data collection effort, processing efficiency and memory/storage requirement in EEG signal-classification applications. This is because for the same level of performance, R-CSP-A needs much fewer training samples than other competing algorithms. For example, from Fig. 7, to achieve an average CCR of at least , R-CSP-A needs only 10 samples in total, while R-CSP-CV needs 20 samples, and the other four algorithms need more than 90 samples.

Finally, since R-CSP-A is an aggregation of A multiple R-CSPs, the computational time of R-CSP-A is about A times of that of CSP. However, since CSP is a very efficient algorithm and only a very small number (six) of features is involved for each R-CSP, the increased computational time will result in little performance degradation in modern computer systems.

V. CONCLUSION

The sample-based covariance-matrix estimation in the conventional CSP algorithm results in limited performance when only a small number of samples is available for training. This paper addresses the problem of discriminative CSP extraction in SSS for EEG signal classification. CSP is regularized using two regularization parameters, with one, to lower the estimation variance, and the other, to reduce the estimation bias. The principle of generic learning is applied in the regularization process. To tackle the problem of regularization parameter determination in SSS, the aggregation method in [12] is adopted. Experiments were performed on data set IVa of BCI Competition III. The experimental results have demonstrated the effectiveness of the proposed R-CSP-A algorithm, especially its superiority over other competing algorithms in SSS. Moreover, R-CSP-A has a positive impact in data collection effort, data storage, and processing efficiency as well.

ACKNOWLEDGMENT

The authors would like to thank Fraunhofer FIRST, Intelligent Data Analysis Group (K.-R. Müller, Benjamin Blankertz), and Campus Benjamin Franklin of the Charité, University Medicine Berlin, Department of Neurology, Neurophysics Group (G. Curio) for providing the data set used in this paper. The authors also thank the authors of [38] for providing us the subject-specific frequency bands information and the anonymous reviewers for their helpful comments.

REFERENCES

[1] S. Sanei and J. A. Chambers, *EEG Signal Processing*. New York: Wiley, 2007.

[2] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[3] M. Grosse-Wentrup, C. Liefhold, K. Gramann, and M. Buss, "Beamforming in noninvasive brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1209–1219, Apr. 2009.

[4] C. D. Dinnie, A. J. Rowan, and T. Gutter, *A Manual of Electroencephalographic Technology*. Cambridge, U.K.: Cambridge Univ. Press, 1982.

[5] 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, Apr. 2010.
[6] Y. Wang, S. Gao, and X. Gao, "Common spatial pattern method for channel selection in motor imagery based brain-computer interface," in *Proc. Eng. Med. Biol. Soc. (EMBC)*, Sep. 2005, pp. 5392–5395.
[7] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Annu. Int. IEEE Eng. Med. Biol. Soc. (EMBS) Conf. Neural Eng.*, May 2007, pp. 204–207.
[8] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, Aug. 2009.
[9] Data set IVa for the BCI Competition III. (2005). [Online]. Available: http://www.bbci.de/competition/iii/desc_IVa.html
[10] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, Mar. 1989.
[11] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognit. Lett.*, vol. 26, no. 2, pp. 181–191, 2005.
[12] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 103–123, Jan. 2009.
[13] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized discriminant analysis for the small sample size problem in face recognition," *Pattern Recognit. Lett.*, vol. 24, no. 16, pp. 3079–3087, 2003.
[14] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting LDA with regularization on MPCA features for gait recognition," in *Proc. Biometrics Symp. 2007*, pp. 1–6.
[15] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognit.*, vol. 39, no. 9, pp. 1746–1762, 2006.
[16] M. Grosse-Wentrup, K. Gramann, and M. Buss, "Adaptive spatial filters with predefined region of interest for EEG based brain-computer-interfaces," in *Advances in Neural Information Processing Systems*, B. Schölkopf and J. C. Platt, Eds. Cambridge, MA: MIT Press, 2007.
[17] B. Blankertz, M. Kawanabe, R. Tomioka, F. U. Hohlefeld, V. Nikulin, and K. R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *NIPS*: B. Schölkopf and J. C. Platt, Eds., 2007.
[18] R. Tomioka and K. Aihara, "Classifying matrices with a spectral regularization," in *Proc. Int. Conf. Mach. Learning (ICML)*, 2007, pp. 895–902.
[19] F. Lotte and C. Guan, "Spatially regularized common spatial patterns for EEG classification," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, 2010, pp. 3712–3715.
[20] R. Tomioka and K.-R. Muller, "A regularized discriminative framework for EEG based communication," *Neuroimage*, vol. 49, no. 1, pp. 415–432, 2010.
[21] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. Eng. Med. Biol. Soc. (EMBC)*, Sep. 2009, pp. 6599–6602.
[22] X. Yong, R. K. Ward, and G. E. Birch, "Robust common spatial patterns for EEG signal preprocessing," in *Proc. Eng. Med. Biol. Soc. (EMBC)*, Aug. 2008, pp. 2087–2090.
[23] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Uncorrelated multilinear discriminant analysis with regularization for gait recognition," in *Proc. Biometrics Symp. 2007*, pp. 1–6.
[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley Interscience, 2001.
[25] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[26] R. E. Schapire, "The boosting approach to machine learning: An overview," in *Nonlinear Estimation and Classification*, D. D. Denison, M. H. Hansen, C. Holmes, B. Mallick, and B. Yu, Eds. New York: Springer-Verlag, 2003.
[27] M. Skurichina and R. P. W. Duin, "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Anal. Appl.*, vol. 5, no. 2, pp. 121–135, 2002.
[28] L. Breiman, "Arcing classifiers," *Ann. Statist.*, vol. 26, no. 3, pp. 801–849, 1998.
[29] A. Ross and R. Govindarajan, "Feature level fusion of hand and face biometrics," in *Proc. SPIE Conf. Biometric Technol. Human Identification II*, Mar. 2005, pp. 196–204.
[30] A. Ross, A. K. Jain, and J. Z. Qian, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, pp. 2115–2125, 2003.
[31] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.
[32] J. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, and S. Z. Li, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 166–178, Jan. 2006.
[33] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Boosting discriminant learners for gait recognition using MPCA features," *EURASIP J. Image Video Process.*, vol. 2009, pp. 1–11.
[34] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. del R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 153–159, Jun. 2006.
[35] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jun. 2008, pp. 2390–2397.
[36] S. L. Sun, J. H. Xu, L. Y. Yu, Y. G. Chen, and A. L. Fang, "Mixtures of common spatial patterns for feature extraction of EEG signals," in *Proc. Int. Conf. Mach. Learning Cybern.*, Jul. 2008, pp. 2923–2926.
[37] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
[38] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial eeg during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.
[39] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.
[40] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, "Time domain parameters as a feature for eeg-based brain-computer interfaces," *Neural Netw.*, vol. 22, no. 9, pp. 1313–1319, Nov. 2009.

Haiping Lu (S'02–M'09) received the B.Eng. and M.Eng. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2001 and 2004, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Canada, in 2008.

He is currently a Research Fellow at the Institute for Infocomm Research, Singapore. His current research interests include pattern recognition, machine learning, biometrics, and biomedical engineering.

How-Lung Eng (M'03) received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from Nanyang Technological University, Singapore, in 1998 and 2002, respectively. He is currently a Research Scientist with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. His research interest includes real-time vision, pattern classification and machine learning for human behavior analysis, and abnormal event detection. He has made several Patent Cooperation Treaty fillings related to video surveillance applications.

Cuntai Guan (S'92–M'97–SM'03) received the Ph.D. degree in electrical and electronic engineering from Southeast University, Nanjing, China, in 1993.

He currently is a Principal Scientist and Program Manager at the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore. His current research interests include brain-computer interface, neural signal processing, machine learning, pattern classification, and statistical signal processing, with applications to neurorehabilitation, health monitoring, and cognitive training. He is an Associate Editor of Frontiers in Neuroprosthetics.

Konstantinos N. Plataniotis (S'93–M'95–SM'03) received the B.Eng. degree in computer engineering from the University of Patras, Patras, Greece, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the Florida Institute of Technology, Melbourne, in 1992 and 1994, respectively.

He is currently a Professor in the Department of Electrical and Computer Engineering and the Director of the Knowledge Media Design Institute at the University of Toronto. His research interests include multimedia systems, biometrics, image and signal processing, communications systems, and pattern recognition.

Dr. Plataniotis is a Registered Professional Engineer in Ontario, and the Editor-in-Chief (2009-2011) for the IEEE SIGNAL PROCESSING LETTERS.

Anastasios N. Venetsanopoulos (S'66–M'69–SM'79–F'88) received the B.Eng. degree in electrical and mechanical engineering from the National Technical University of Athens, Athens, Greece, in 1965, and the M.S., M.Phil., and Ph.D. degrees in electrical engineering from Yale University, New Haven, CT, in 1966, 1968, and 1969, respectively.

He is currently a Professor of Electrical and Computer Engineering at Ryerson University, Toronto, ON, Canada, and a Professor Emeritus at the Department of Electrical and Computer Engineering, University of Toronto, ON. His research interests include multimedia, digital signal/image processing, telecommunications, and biometrics.

Dr. Venetsanopoulos is a Fellow of the Engineering Institute of Canada, the Canadian Academy of Engineering, and the Royal Society of Canada.