Separable Common Spatio-Spectral Patterns for Motor Imagery BCI Systems

Amirhossein S. Aghaei*, *Member, IEEE*, Mohammad Shahin Mahanta, *Student Member, IEEE*, and Konstantinos N. Plataniotis, *Fellow, IEEE*

***Abstract*—*Objective*: Feature extraction is one of the most important steps in any brain–computer interface (BCI) system. In particular, spatio-spectral feature extraction for motor-imagery BCIs (MI-BCI) has been the focus of several works in the past decade. This paper proposes a novel method, called separable common spatio-spectral patterns (SCSSP), for extraction of discriminant spatio-spectral EEG features in MI-BCIs. *Methods:* Assuming a binary classification problem, SCSSP uses a heteroscedastic matrix-variate Gaussian model for the multiband EEG rhythms, and seeks the spatio-spectral features whose variance is maximized for one brain task and minimized for the other task. Therefore, SCSSP can be considered as a spatio-spectral generalization of the conventional common spatial patterns (CSP) algorithm. *Results:* The experimental results on two-class and multiclass motor-imagery data from publicly available BCI Competition datasets demonstrate that the proposed computationally efficient method competes closely with filter-bank CSP (FBCSP), and can even outperform the FBCSP if enough training data are available. Furthermore, SCSSP provides us with a simple measure for ranking the discriminant power of extracted spatio-spectral features, which is not possible in FBCSP. *Conclusion:* The matrix-variate Gaussian assumption allows the SCSSP method to jointly process the EEG data in both spatial and spectral domains. As a result, compared to the similar solutions in the literature such as FBCSP, the proposed SCSSP method requires significantly lower computations. *Significance:* The proposed computationally efficient spatio-spectral feature extractor is particularly suitable for applications in which the computational power is limited, such as emerging wearable mobile BCI systems.**

***Index Terms*—Brain–computer interface (BCI), common spatial patterns, matrix-variate Gaussian, spatio-spectral features, separability.**

I. Introduction

BRAIN–COMPUTER interface (BCI) systems aim to provide a nonmuscular channel for the brain to control external devices using electrical activities of the brain. These BCIs can be used in various applications, such as controlling a wheelchair or neuroprosthesis for disabled individuals, navigation in virtual environment, and assisting healthy individuals in performing highly demanding tasks or controlling devices such as quadcopters in 2-D/3-D space [2]–[10]. A comprehensive review of such applications is provided by Yuan and He [11]. Motor-imagery BCI systems in particular are based on decoding imagination of motor tasks, e.g., to control the movement of a wheelchair, or a mouse curser on computer screen, and move it to the right/left directions by imagining right/left hand movement. During the past decade, there has been a growing interest in utilization of electroencephalogram (EEG) signals for noninvasive motor-imagery BCI (MI-BCI) systems, due to their low cost, ease of use, and widespread availability.

The EEG signal recorded during motor imagination exhibits task-specific features in both spatial domain and spectral (or frequency) domains [12]–[15]. Several spatial and spectral processing methods have been used in the literature to extract the most discriminant features from these EEG signals. The common spatial patterns (CSP) algorithm is one of the most successful solutions which has been widely used in MI-BCIs [13], [16], [17]. CSP was first used in BCIs with two-class problem, such as left hand versus right hand movement. Given a set of training data, this algorithm tries to find spatial filters that maximize the variance for one class, while minimizing the variance of the other class. In the case of left/right hand movement, this criterion very well matches the characteristics of EEG signals, since during the hand movement imagination, the power of ipsilateral channels is maximized (event-related synchronization), while the power of contralateral channels is minimized (even-related desynchronization) [12], [18], [19].

Despite its power in extracting spatial features, CSP completely ignores the spectral characteristics of the EEG signal. Several variants of CSP have been proposed in the literature to resolve this problem [20]–[29]. The work in [20], called common spatio-spectral patterns (CSSP) method, applies CSP to the first-order time delayed version of the EEG data to incorporate the information in the frequency domain. The work in [21], extends the CSSP solution by utilizing higher order finite impulse response (FIR) filters that provide more degrees of freedom for extraction of spectral features. To jointly optimize both the FIR spectral filter and the CSP module, various iterative procedures have been designed in the literature that alternate between optimization of the spatial and spectral filters. These methods include spectrally weighted CSP [26], which utilizes a criterion similar to CSP to optimize the FIR filter coefficients, and the more recent methods of iterative spatio-spectral patterns learning (ISSPL) [22] and discriminative filter-bank CSP (DFBCSP) [27]. Similarly, spatio-spectral iterative procedures have been utilized to maximize the mutual information between the features and the classes, and hence, minimization of the Bayes classification error [23], [28]. However, these iterative

Manuscript received August 22, 2015; revised September 14, 2015; accepted September 28, 2015. Date of publication September 28, 2015; date of current version December 17, 2015. *Asterisk indicates corresponding author*

*A. S. Aghaei is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto M5S 3G4, Canada (e-mail: aghaei@ece.utoronto.ca).

M. S. Mahanta and K. N. Plataniotis are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto.

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.2015.2487738

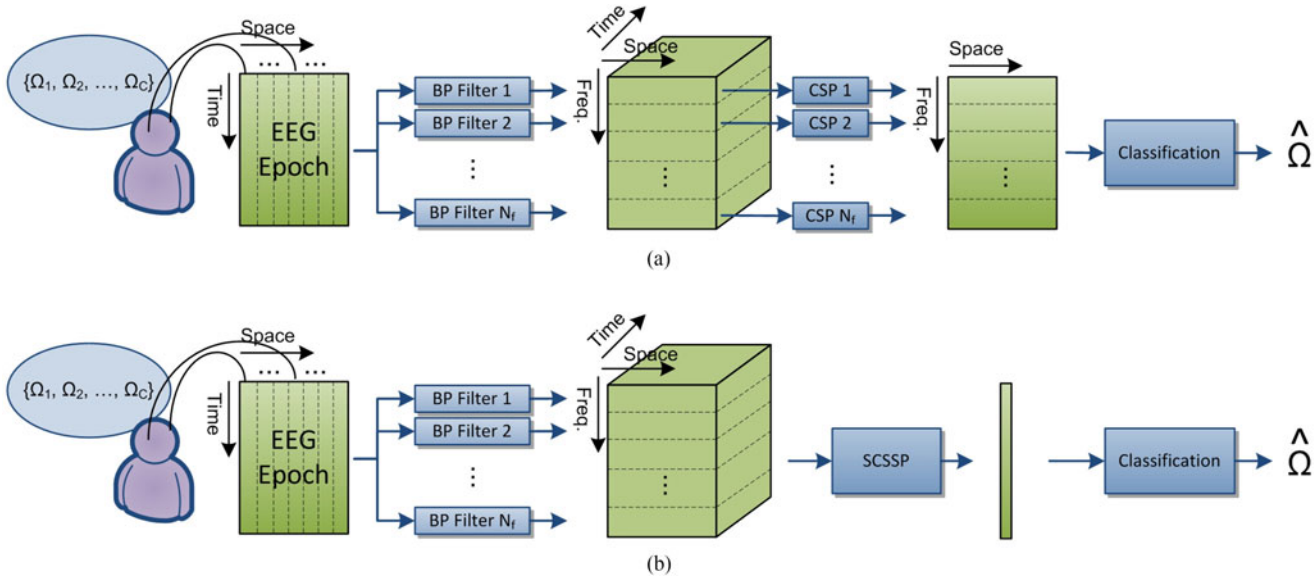


Fig. 1. System model for spatio-spectral feature extraction schemes in (a) Filter-bank common spatial pattern (FBCSP), and (b) Separable common spatio-spectral pattern (SCSSP) methods. Note that FBCSP algorithm utilized in this paper does not deploy the mutual information based feature selector of [25].

methods impose a significantly high computational complexity on the training stage.

The work in [24] provides a probabilistic Bayesian framework to find the optimized frequency bands and the corresponding spatial filters. Also, a recent work [29] has utilized a relatively high-dimensional (and hence, computationally expensive) eigen-decomposition on regularized covariances of concatenated time-delayed data to ensure a joint spatio-spectral optimization of the features. Yet, the most computationally efficient algorithm is the filter-bank CSP (FBCSP) [25], which uses a bank of N_f bandpass filters to obtain the EEG rhythms in different frequency bands, and then, deploys a separate CSP module for each frequency band to extract spatial features from each EEG rhythm [see Fig. 1(a)].

The FBCSP method has been highly successful for feature extraction from motor imagery data in BCI Competition IV, where it has achieved the highest performance for Dataset 2a and Dataset 2b [25]. Despite its high performance, the FBCSP method suffers from a number of shortcomings as follows. 1) FBCSP has a high computational cost at the training phase since it requires a separate feature extractor for each spectral band, each of which requires calculation of generalized eigenvectors for spatial covariance matrices. This issue can be restrictive in applications that require frequent retraining of the BCI system (e.g., adaptive scenarios) or the ones that require the training to be performed on a low power mobile/wearable device. 2) Since each spectral band is treated independently, possible correlations between different EEG rhythms are ignored by the FBCSP method, which in turn causes redundancy in the extracted feature set. 3) FBCSP does not provide any measure for comparing discriminant power of the features obtained from different spectral bands. Although the CSP features within each band are sorted based on their discriminant power, it is not possible to sort the features across different bands.[1]

We propose a novel algorithm that simultaneously processes the EEG rhythmic activities in both spatial and spectral domains, and extracts the most discriminant spatio-spectral features across all the frequency bands. Using a matrix-variate Gaussian model for spatio-spectral EEG patterns, we develop a bilinear feature extractor, called *separable common spatio-spectral patterns (SCSSP)*, which has the following advantages compared to the FBCSP method: First, the SCSSP has significantly less computational cost for training since it requires training of only two CSP-type modules instead of N_f modules in FBCSP. Second, the features are extracted based on joint analysis of both spatial and spectral characteristics of the signal. Therefore, correlations between different spectral bands can be exploited for feature extraction. Third, a measure is provided to rank the discriminatory power of extracted spatio-spectral features, which enables us to directly perform dimensionality reduction without any need to deploy a separate subsequent feature extraction/selection module, such as the one suggested in [25].

The rest of this paper is organized as follows: Section II provides the system model and introduces the matrix-variate Gaussian model used throughout this paper. The proposed SCSSP method is discussed in Section III-A. Section III-B briefly compares the SCSSP with the state of the art FBCSP method to provide the reader with a better understanding of the similarities and the differences between these two solutions. Multiclass extension of the SCSSP method is discussed in Section III-C. Section IV studies the performance of the SCSSP method in different experimental setups. Finally, the summary and concluding remarks are presented in Section V.

[1]It is noteworthy that the argument here is regarding the FBCSP feature extractor, which is composed of bandpass filtering followed by CSP feature extraction, as shown in Fig. 1(a). The work in [25] suggests the use of a separate feature selector to estimate the discriminance power of the extracted features, using the mutual information criteria; however, such feature selection is not an integral part of the FBCSP algorithm and can be substituted with any other feature selector/extractor that suits this purpose. Therefore, this feature selection stage that sorts the features from different frequency bands based on the mutual information criterion is not employed in this paper.

II. System Model

The processing pipeline for the proposed SCSSP method is presented in Fig. 1(b). Consider an EEG epoch with N_t time samples from N_{ch} channels.[2] The EEG signal is first passed through a set of N_f bandpass filters. This process extracts N_f rhythmic activities from each EEG channel, where each rhythm has a length of N_t samples. These extracted rhythms form a three dimensional tensor, as shown in Fig. 1(b), which consists of N_t matrices of size $N_f \times N_{\text{ch}}$. Each of these matrices represents a spatio-spectral EEG pattern at a certain time instant. The proposed SCSSP method operates on these matrix-variate patterns, denoted by $\mathbf{X} \in \mathbb{R}^{N_f \times N_{ch}}$, in order to extract the most discriminant features for classification.

Using a *heteroscedastic matrix-variate Gaussian model* [30] for $\mathbf{X}$, we provide a mathematical framework that allows us to take into account the joint characteristics of the spatial and spectral features. Let $f(\mathbf{X}|\Omega_i)$ denote the conditional probability of $\mathbf{X}$ under class Ω_i. The matrix-variate Gaussian model is defined as

$$\mathbf{X}|\Omega_i \sim \mathcal{N}(\mathbf{M}_i, \mathbf{\Phi}_i, \mathbf{\Psi}_i), \qquad 1 \leq i \leq C \tag{1}$$

where C is the total number of classes, $\mathbf{M}_i$ denotes the class mean, $\mathbf{\Phi}_i$ is the spectral covariance, also called column-wise or left covariance, and $\mathbf{\Psi}_i$ is the spatial covariance, also called row-wise or right covariance. Since $\mathbf{X}$ is obtained from bandpass filtering of the EEG signal, all classes have zero mean, i.e., $\mathbf{M}_i = \mathbf{0}$ for $1 \leq i \leq C$. Therefore, the discriminant information is contained in the second-order statistics of the data, defined as follows

$$\mathbf{\Phi}_i = \text{tr}^{-1}(\mathbf{\Psi}_i) * \text{E}_{\mathbf{X}|\Omega_i}\left(\mathbf{X}\mathbf{X}^T\right) \tag{2}$$

$$\mathbf{\Psi}_i = \text{tr}^{-1}(\mathbf{\Phi}_i) * \text{E}_{\mathbf{X}|\Omega_i}\left(\mathbf{X}^T\mathbf{X}\right). \tag{3}$$

Let $\mathbf{x} = \text{vec}(\mathbf{X})$ be the vectorized representation of $\mathbf{X}$ obtained from concatenation of its columns, and $\mathbf{\Sigma}_i$ be the covariance matrix of $\mathbf{x}$ under Ω_i. The model in (1) implies that $\mathbf{x}|\Omega_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_i),\ 1 \leq i \leq C$, where $\mathbf{\Sigma}_i = \mathbf{\Psi}_i \otimes \mathbf{\Phi}_i$. Furthermore, any bilinear operation of the form $\mathbf{W}_L\mathbf{X}\mathbf{W}_R$ is equivalent to the linear operation $\mathbf{W}^T\mathbf{x} = (\mathbf{W}_R \otimes \mathbf{W}_L)^T\mathbf{x}$.

III. Separable Common Spatio-Spectral Patterns

A. Definition and Methodology

Consider a binary classification problem, i.e., $\Omega_i \in \{\Omega_1, \Omega_2\}$. Based on the properties of matrix-variate Gaussian model, and following the general goal of the CSP approach, we look for a bilinear operation on $\mathbf{X}$, which simultaneously diagonalizes both $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. In other words, we look for transformation matrices $\mathbf{W}_L$ and $\mathbf{W}_R$, which are the solutions to the following generalized eigenvalue problem:

$$\mathbf{\Sigma}_1\mathbf{W} = (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\,\mathbf{W}\mathbf{\Lambda}, \tag{4}$$

$$\text{where} \quad \mathbf{W} = \mathbf{W}_R \otimes \mathbf{W}_L, \text{ and } \quad \mathbf{\Sigma}_i = \mathbf{\Psi}_i \otimes \mathbf{\Phi}_i\ .$$

[2]In this paper, scalars, vectors, and matrices are, respectively, shown in regular lowercase/uppercase (e.g., a or A), boldface lowercase (e.g., $\mathbf{a}$), and boldface uppercase (e.g., $\mathbf{A}$). Trace of $\mathbf{A}$ is denoted by $\text{tr}(\mathbf{A})$. Also, the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$.

The next theorem provides the solution for (4).

Theorem 1: Let $\mathbf{x} = \text{vec}(\mathbf{X})$, where $\mathbf{X} \in \mathbb{R}^{N_f \times N_{ch}}$ has a matrix-variate Gaussian distribution as given by (1). Then, the solution to (4) is given as follows: $\mathbf{W} = \mathbf{W}_R \otimes \mathbf{W}_L$, and

$$\mathbf{\Lambda} = (\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L)\left(\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L + (\mathbf{I}_{N_{ch}} - \mathbf{\Lambda}_R) \otimes \left(\mathbf{I}_{N_f} - \mathbf{\Lambda}_L\right)\right)^{-1}$$

where $\mathbf{I}_K$ is the identity matrix of size K and the matrices $\mathbf{\Lambda}_R$, $\mathbf{W}_R$, $\mathbf{\Lambda}_L$, and $\mathbf{W}_L$ are, respectively, the solutions to generalized eigenvalue problems for spatial and spectral covariances

$$\mathbf{\Psi}_1\mathbf{W}_R = (\mathbf{\Psi}_1 + \mathbf{\Psi}_2)\,\mathbf{W}_R\mathbf{\Lambda}_R \tag{5}$$

$$\mathbf{\Phi}_1\mathbf{W}_L = (\mathbf{\Phi}_1 + \mathbf{\Phi}_2)\,\mathbf{W}_L\mathbf{\Lambda}_L. \tag{6}$$

Proof. The proof is provided in Appendix A. ■

Using this theorem, we can break the generalized eigenvalue problem of (4) into two lower dimensional problems presented in (5) and (6). Note that $\mathbf{W}_L$ provides the spectral transformation matrix, whereas $\mathbf{W}_R$ provides the spatial transformation matrix. These two transformations will be simultaneously applied to the matrix-variate data $\mathbf{X}$.

To provide a better insight into the result of Theorem 1, let $\lambda_k, 1 \leq k \leq N_f N_{\text{ch}}$, denote the diagonal entries of $\mathbf{\Lambda}$ sorted in descending order. Theorem 1 implies each λ_k corresponds to a pair of eigenvalues from $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$ as follows:

$$\lambda_k = \frac{\lambda_{L,l[k]}\,\lambda_{R,j[k]}}{\lambda_{L,l[k]}\,\lambda_{R,j[k]} + (1-\lambda_{L,l[k]})(1-\lambda_{R,j[k]})}. \tag{7}$$

Here, $\lambda_{L,l[k]}$ and $\lambda_{R,j[k]}$ are the corresponding eigenvalues in $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$, with $1 \leq l[k] \leq N_f$ and $1 \leq j[k] \leq N_{ch}$.[3] Also, the eigenvectors corresponding to λ_k are expressed as $\mathbf{w}_k = \mathbf{w}_{R,j[k]} \otimes \mathbf{w}_{L,l[k]}$, where $\mathbf{w}_{R,j[k]}$ and $\mathbf{w}_{L,l[k]}$ are the eigenvectors in $\mathbf{W}_R$ and $\mathbf{W}_L$ corresponding to $\lambda_{R,j[k]}$ and $\lambda_{L,l[k]}$, respectively. Based on these results, the following algorithm will be used for extracting the "d" most discriminant spatio-spectral features.

1) Assuming that N_i training samples $\mathbf{X}_{i,n}, 1 \leq n \leq N_i$, are available for each class Ω_i, estimate the spectral covariance and spatial covariance of the data, using[4]

$$\hat{\mathbf{\Phi}}_i = \frac{1}{N_{ch}N_i}\sum_{n=1}^{N_i}\mathbf{X}_{i,n}\mathbf{X}_{i,n}^T \tag{8}$$

$$\hat{\mathbf{\Psi}}_i = \frac{1}{N_f N_i}\sum_{n=1}^{N_i}\mathbf{X}_{i,n}^T\mathbf{X}_{i,n}. \tag{9}$$

2) Solve the generalized eigenvalue problems in (5) and (6) for the estimated spatial and spectral covariances.
3) Using (7), calculate the eigenvalues λ_k and sort them in descending order to determine the indices $l[k]$ and $j[k]$.

[3]The indices $l[k]$ and $j[k]$ are determined as follows. Let $\lambda_{L,p}$ and $\lambda_{R,q}$ be the eigenvalues of $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$, respectively, where $1 \leq p \leq N_f$ and $1 \leq q \leq N_{ch}$. First, take all the possible pairs of $\lambda_{L,p}$ and $\lambda_{R,q}$, and form the corresponding λ values using (7). Then, sort these resulting values in descending order and denote them by λ_k, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_f N_{ch}}$. For each λ_k, the corresponding value of index p will be denoted by $l[k]$, and the corresponding value of index q will be denoted by $j[k]$.

[4]Note that in Section II, the temporal length of each EEG epoch was denoted by N_t. Therefore, N_i equals N_t times the total number of training EEG epochs for class Ω_i.

4) Extract the d most discriminant features by calculating

$$y_k = \mathbf{w}_{L,l[k]}^T \mathbf{X} \mathbf{w}_{R,j[k]} \tag{10}$$

$$k \in \{1,\, N_f N_{ch},\, 2, (N_f N_{ch} - 1),\, \ldots,\, \tfrac{d}{2}, (N_f N_{ch} - \tfrac{d}{2} + 1)\}.$$

Note that d is an even number here, similar to the CSP.

5) Calculate the normalized power of features over the length of epoch, in logarithmic scale, as follow:

$$z_k = \log\left(\frac{\mathrm{var}\,(y_k)}{\Sigma_k \mathrm{var}\,(y_k)}\right) \tag{11}$$

where $\mathrm{var}(y_k)$ function calculates the variance or power of y_k over N_t samples, and $\Sigma_k \mathrm{var}(y_k)$ represents the total power of the d features.

6) Construct the feature vector $\mathbf{z} = [z_1, z_{N_f N_{ch}}, \ldots]^T \in \mathbb{R}^{d\times 1}$ as the output of the SCSSP algorithm.

In the aforementioned algorithm, the values of λ_k can be considered as a measure to determine the discriminant power of feature y_k, as follows. Similar to the conventional CSP method, λ_k ranges between zero and one. Values close to zero or one correspond to high discriminant features, whereas values close to $\frac{1}{2}$ correspond to low discriminant features. In other words, the pair of features $[y_1, y_{N_f N_{ch}}]^T$ provide the most discriminant power. Similarly, the features corresponding to $k = 2$ and $k = (N_f N_{ch} - 1)$ are the second most discriminant features, and so on. Thus, the pairs of extracted spatio-spectral features in $\mathbf{z}$ are sorted according to their discriminant power in descending order. These features are then passed to a classifier to determine the $\hat{\Omega}$. In our experimental studies, we consider two simple possible choices for the classifier: (a) naive Bayes Parzen window (NBPW) classifier [25] and (b) linear minimum mean distance classifier [16].

B. Comparative Discussion on the Theoretical Assumptions of FBCSP and SCSSP

Consider the matrix-variate data $\mathbf{X}$ at the output of the band-pass filter bank, and denote the f^{th} row vector of $\mathbf{X}$ by $\mathbf{x}_f$, where $1 \le f \le N_f$. Also, let $\mathbf{x}' = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_f}]$ denote the row-vector that is generated from the row-wise concatenation of the elements in $\mathbf{X}$, i.e., $\mathbf{x}' = (\mathrm{vec}(\mathbf{X}^T))^T$. The class conditional covariance matrix of each row-vector $\mathbf{x}_f$ will be represented by $\mathbf{\Psi}_i^f$, and the class conditional covariance matrix of $\mathbf{x}'$ is represented by $\mathbf{\Sigma}_i{}'$.

Recall that in the FBCSP approach, each row-vector $\mathbf{x}_f$ is processed independently from the other rows, using the projection matrix $\mathbf{W}_R^f$ containing the generalized eigenvectors of $\mathbf{\Psi}_1^f$ and $\mathbf{\Psi}_1^f + \mathbf{\Psi}_2^f$. The projected feature pairs are then sorted in descending order of significance. Finally, the log-power of the resulting features are calculated during the epoch length and form the f^{th} row of the output feature matrix. If we compare this approach with SCSSP's approach, the following differences can be pointed out.

The separability assumption in the matrix-variate Gaussian model, which is used by the SCSSP, implies that the covariance matrix of each row-vector $\mathbf{x}_f$ is equal, up to a scale, to the covariance matrix of other row vectors. As a result, the SCSSP only looks for one spatial filtering matrix $\mathbf{W}_R$, which will be commonly applied to all the row vectors in $\mathbf{X}$. In contrast, the FBCSP method assumes that each row-vector $\mathbf{x}_f$ has a unique covariance matrix, and hence, looks for a unique spatial filtering matrix $\mathbf{W}_R^f$ for each row.

The other important difference between FBCSP and SCSSP is in the spectral processing of the data. FBCSP assumes that different EEG rhythms in different frequency bands are independent from each other, and thus, independently processes each rhythm. However, the SCSSP calculates the class conditional spectral covariance matrix $\mathbf{\Phi}_i$ and uses this information together with the information from the spatial covariance matrix $\mathbf{\Psi}_i$ for extraction of the most discriminant spatio-spectral features. Note that owing to the matrix-variate Gaussianity assumption, the SCSSP assumes all EEG channels have the same spectral covariance matrices, up to a scale, and hence, calculates a common spectral covariance matrix for all channels.

In order to further clarify these points, consider the row vector $\mathbf{x}' = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_f}]$, which contains all the elements of $\mathbf{X}$. The FBCSP method assumes a block-diagonal structure for the class conditional covariance of $\mathbf{x}'$ as follows:

$$\mathbf{\Sigma}_i{}' = \begin{bmatrix} \mathbf{\Psi}_i^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_i^2 & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{\Psi}_i^{N_f} \end{bmatrix} \tag{12}$$

whereas the SCSSP assumes the following block-wise structure

$$\mathbf{\Sigma}_i{}' = \begin{bmatrix} \phi_{1,1}\mathbf{\Psi}_i & \phi_{1,2}\mathbf{\Psi}_i & \cdots & \phi_{1,N_f}\mathbf{\Psi}_i \\ \phi_{2,1}\mathbf{\Psi}_i & \phi_{2,2}\mathbf{\Psi}_i & \cdots & \phi_{2,N_f}\mathbf{\Psi}_i \\ \vdots & & \ddots & \\ \phi_{N_f,1}\mathbf{\Psi}_i & \phi_{N_f,2}\mathbf{\Psi}_i & \cdots & \phi_{N_f,N_f}\mathbf{\Psi}_i \end{bmatrix} \tag{13}$$

where $\phi_{m,n}$ represents the $(m,n)^{th}$ element of the spectral covariance matrix $\mathbf{\Phi}$.

It is noteworthy that both assumption in (12) and (13) are restrictive models for the spatio-spectral covariance of the data. The FBCSP completely ignores the off-diagonal blocks of the $\mathbf{\Sigma}_i{}'$, while trying to provide an accurate estimate for the diagonal blocks. In contrast, the SCSSP takes into account the off-diagonal blocks of $\mathbf{\Sigma}_i{}'$ by making the simplifying assumption that all the blocks in $\mathbf{\Sigma}_i{}'$ are up to a scale equal to each other, where the scaling factor is determined by the elements of spectral covariance matrix.

For further clarification, note that FBCSP uses $N_f * N_{ch} * (N_{ch}+1)/2$ parameters to estimate the covariance matrix $\mathbf{\Sigma}_i{}'$, whereas the SCSSP uses $N_f * (N_f+1)/2 + N_{ch} * (N_{ch}+1)/2$ parameters. In our experimental setups $N_{ch} > N_f > 1$; therefore, the FBCSP always uses more parameters compared to SCSSP (ref. Tables VI and VII for the numerical values). Similarly, the number of weighting coefficients used for extraction of EEG features is also larger for FBCSP in comparison with the SCSSP. Indeed, FBCSP uses $N_f * N_{ch}$ coefficients, whereas SCSSP uses $N_f + N_{ch}$ coefficients.[5] Therefore, SCSSP

[5]For multiclass extension of FBCSP/SCSSP, where $C > 2$, the aforementioned values will be multiplied by a factor of C, as explained in Section III-C.

can be viewed as a more restrictive regularizer compared to the FBCSP method.

C. Multiclass Extension of the SCSSP Method

Although SCSSP is derived for a binary classification scenario, there exist several solutions to utilize a binary feature extractor in a multiclass scenario, such as the divide-and-conquer, pairwise, and one-versus-rest (OVR) strategies reviewed in [25]. Since OVR has been shown to provide best performance for FBCSP [25], we utilize OVR strategy for multiclass extension of SCSSP too.[6] Consider the training phase of the SCSSP, and let Ω_i' be the set of all motor-imagery tasks excluding the i^{th} task Ω_i. Starting from $i = 1$, SCSSP finds the bilinear transformation matrices $\mathbf{W}_L^{(i)}$ and $\mathbf{W}_R^{(i)}$ to extract d_0 features that provide high discriminance power for classification of Ω_i versus Ω_i'. This procedure is repeated for $i \in \{1, \ldots, C\}$, which results in a set of C spectral transformation matrices, and C spatial transformation matrices.

Now, consider the testing phase, and let $\mathbf{X} \in \mathbb{R}^{N_f \times N_{ch}}$ represent a test sample. The matrix $\mathbf{X}$ will be passed through C pairs of joint spatio-spectral transformation matrices, i.e., $\mathbf{T}_i = \{\mathbf{W}_L^{(i)}, \mathbf{W}_R^{(i)}\}, i \in \{1, \ldots, C\}$ to generate a set of $d_0 * C$ features. The most discriminant features in this set consists of the first pair of discriminant features obtained from each $\mathbf{T}_i$, which forms a set of $2 * C$ features. Similarly, the second pair of features from each $\mathbf{T}_i$ form the next $2 * C$ discriminant features, and so on. Therefore, in the resulting feature vector, the first $2 * C$ features will correspond to the most discriminant group of features, and similarly the n^{th} group of $2 * C$ features represent the n^{th} most discriminant features.

To classify the extracted features in the multiclass scenario, the following two approaches are used. For the NBPW classifier, the OVR approach of [25] is utilized, where C binary NBPW classifiers are deployed to estimate the posterior probability of each class versus the rest of classes, and the one with the highest probability is selected as the classifier's output. For the linear minimum mean distance classifier, all the C sets of features extracted from different choices of one class versus the rest of classes are concatenated to form a feature vector, which is then passed to the classifier. The classifier's output in this case will be the label of the class whose mean has the minimum Euclidean distance with the classifier's input feature vector.

IV. Experimental Analysis

In this section, we will study the performance of the proposed SCSSP method in both two-class and multiclass motor-imagery scenarios and compare it with the conventional FBCSP method. For these studies, we use Dataset V from BCI competition III [31] and Dataset 2a from BCI competition IV [32].[7] As suggested by the providers of the former dataset, we will also study the effect of surface Laplacian (SL) filtering and channel selection (CS) on the performance of the SCSSP method. The surface Laplacian method can be viewed as a high-pass spatial filter that removes the nonlocalized signal components as well as the interference from neighboring areas caused by volume conduction [33], [34]. The CS is a strategy to reduce the dimensionality of the EEG signal by only selecting the most informative EEG channels. In our experiments, CS is performed by selecting the centroparietal channels located over the motor cortex.

Since the main focus in this paper is on design of the feature extraction step, we will not consider any separate feature selection after the SCSSP or FBCSP, and will directly pass the output of the SCSSP or FBCSP to the classifier. Recall that one of the main motivations behind the design of the SCSSP method is to develop a feature extraction method that can effectively sort the extracted spatio-spectral features based on their discriminance power.

Therefore, the following processing steps will be considered for implementation of the SCSSP and FBCSP methods. First, the multichannel EEG signal will be passed through an optional stage of SL filtering or CS. The resulting signal will then be passed through a bank of bandpass filters, to generate the multiband EEG rhythms. At the next step, either SCSSP method or the FBCSP method will be applied to this multiband EEG data to extract a set of discriminant spatio-spectral features. These extracted features will then be directly passed to a classifier. The classifiers studied in this paper are the NBPW classifier, as suggested by [25], and the simple linear minimum mean distance (Lin) classifier [16]. This procedure results in a total of $16 = 2 \times 2 \times 2 \times 2$ different combinations for feature extraction and classification, namely SL(Yes/No), CS(Yes/No), FBCSP/SCSSP, and NBPW/Lin.

A. Experiment Setup

1) BCI Competition III, Dataset V (Exp. 1): The goal of this experiment is to classify the following mental imagery tasks: left hand movement (Ω_1), right hand movement (Ω_2), and generation of words beginning with a random letter (Ω_3). This dataset contains EEG of three normal subjects recorded in four sessions. Each session consists of sequential 15 second trials of the three tasks. The first three sessions will be used for training/cross-validation purposes, whereas the last session is only used as unseen data for competition, i.e., testing phase. The raw EEG used in our studies is recorded using 32-electrode Biosemi system at 512-Hz sampling rate. The BCI algorithm is required to provide the estimated label $\widehat{\Omega}$ every 0.5 second, using only the last second of EEG recording. The performance measure for this competition is correct classification rate (CCR) of the overall system, defined as the ratio of number of successfully classified samples over the total number of samples. The chance of random classification in this experiment is $P_{\text{rand}} = 1/C = 0.33$. The winning algorithm for this competition in the literature used SL and CS followed by short-time Fourier transformation, and classifies the extracted features using linear discriminant analysis

[6] It should be noted that possible extensions of SCSSP to multiclass scnearios requires a comprehensive future study.

[7] In this paper, we consider the state of the art FBCSP algorithm as the baseline solution for our experimental studies. However, interested readers are referred to [25] for comparative analysis of FBCSP and CSP algorithms.

(LDA) [35]. It achieves an average performance of 62.72% at the classifier output.[8]

2) BCI Competition IV, Dataset 2a (Exp. 2): The goal of this experiment is to classify the following motor-imagery tasks: left hand (Ω_1), right hand (Ω_2), both feet (Ω_3), and tongue (Ω_4) movement. This dataset contains EEG recordings of nine normal subjects recorded in two sessions. The signals are recorded using 22 Ag/AgCl electrodes at 250-Hz sampling rate. Each session consists of six runs, each of which includes 48 trials of length 3 second, yielding a total of 288 trials per session. The first session will be used for training/cross-validation and the second session is only used as unseen data for testing phase. This dataset also contains three electrooclugram (EOG) channel recordings that are provided for subsequent application of artifact processing methods and shall not be used for classification. The competition requires the BCI algorithms to provide a continuous classification output for each sample in the form of the estimated label $\widehat{\Omega}$. The performance measure for this competition is the kappa coefficient (κ) of the overall system [36], which is defined as: $\kappa = (\text{CCR} - P_{\text{rand}})/(1 - P_{\text{rand}})$. Here, P_{rand} is the probability of random classification, i.e., $P_{\text{rand}} = 1/C = 0.25$ for this experiment. Note that the measure κ is normalized such that $\kappa = 0$ for a random classifier. The winning algorithm for this competition in the literature is the FBCSP-NBPW method.

Table I presents the parameters used to implement the processing steps and extract the spatio-spectral feature matrix $\mathbf{X}_{N_f \times N_{ch}}$. It should be noted that in each experiment, the epoch length and frequency range used by the winning algorithm in the original competition are adopted in this paper to provide a fair comparison between alternative solutions.

3) Comparative Note on the Datasets in Exp. 1 and Exp. 2: Although both databases in Exp. 1 and Exp. 2 contain EEG data from motor-imagery tasks, these two datasets are significantly different in terms of the availability of the training data. In Exp. 1, each trial is of length 15 second, which is significantly longer than the 3 second trial length in Exp. 2. In the context of motor-imagery tasks, the training trial length is of great importance. When the training trials are longer, the subjects will have enough time to concentrate on the desired motor-imagery task and produce stable brain rhythms that can be reliably used for training. Furthermore, the training data in Exp. 1 is collected during three sessions, whereas Exp. 2 only includes one session of EEG recording for training. It is well known in the context of motor-imagery BCIs that the EEG characteristics exhibit intersession variations, which need to be taken into account while training the BCI algorithm [37].

Since motor-imagery BCI systems are mostly designed for long-term utilization by the user, it is usually assumed that the BCI algorithm has access to a training dataset with long enough trials, which are collected over at least two different recording sessions. From this perspective, the training dataset in Exp. 1 can be considered as a typical dataset for motor-imagery applications, whereas the training set in Exp. 2 is an extreme case where only one recording session with very short trials is available for training the algorithms. We have included Exp. 2 in our analysis to study the robustness of different algorithms in the extreme conditions.

Finally, note that SL filter requires the exact locations of EEG sensors. The dataset in Exp. 1 includes the exact coordinates of the sensors, using the standard 10–10 system. In contrast, the dataset in Exp. 2 only contains the approximate relative locations of the sensors. We have mapped these approximate locations to the following closest standard locations: Fz, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2, POz. The effect of this approximate mapping will be discussed later in the experimental results.

B. Epoch Segmentation

Both FBCSP and SCSSP methods operate on the epochs of EEG data [ref. Fig. 1(a)–(b)] and treat the feature set extracted from each epoch as one sample used for training or testing purposes. In testing phase, the EEG epochs are extracted by truncating the data with rectangular windows of length T seconds, where $T = 1$ for Exp. 1 and $T = 2$ for Exp. 2. These epochs are extracted every 0.5 second in Exp. 1 and every 0.1 s in Exp. 2; hence every two consecutive epochs have a 50% (or $\frac{1}{2}T$) overlap in Exp. 1 and a 95% (or $\frac{19}{20}T$) overlap in Exp. 2, respectively.[9]

In training phase, a very similar procedure is used and the only restriction is to make sure that the training epochs in each trial are extracted from the time period that correspond to execution of motor-imagery tasks, explained as follows.

1) In Exp. 1, each trial lasts for approximately 15 second; therefore, approximately a total of 29 training epochs are extracted from each trial. Each of the three training sessions lasts for almost 4 min; therefore, a total of $1392 (= 29 * 16 * 3)$ epochs will be extracted for training purposes.
2) In Exp. 2, each trial lasts for only 3 second, and the FBCSP method only uses the time interval from $t = 0.5$ until $t = 2.5$ second. To have a fair comparison, we have used the same time interval for the SCSSP method as well. The training session has 288 trials; therefore, a total of 288 epochs will be extracted for training purposes.[10]

C. Results for Two-Class Scenario (Left Hand Versus Right Hand)

Since both SCSSP and FBCSP methods are originally designed for two-class scenarios (ref. Section III-A), we first focus on the binary classification problem of classifying left hand movement imagination versus right hand movement imagination. The results for multiclass extensions of the SCSSP and FBCSP methods will be presented in Section IV-D.

1) Cross-Validation Results: The performance of BCI algorithms highly depends on the dimensionality of the feature space

[8]Using a post-processing strategy, called mental tasks transitions detector, the work in [35] improves this performance to 68.65%. However, analysis of the effect of post-processing methods on FBCSP and SCSSP algorithms is outside the scope of this paper. Therefore, to have a fair comparison, the performance of all methods at the classifier's output is studied in this paper.

[9]Note that in Exp. 2, dataset providers have marked some test trials as artifact-contaminated, which are excluded from our analysis.

[10]The aforementioned numbers of training epochs (i.e., 1392 and 288) are rough estimates. For each subject, a few training trials are contaminated by artifacts that are excluded from the training set (these trials are pre-marked in both datasets).

TABLE I
PARAMETERS USED FOR FEATURE EXTRACTION ALGORITHMS IN EXP. 1 AND EXP. 2

		Experiment 1	**Experiment 2**
Raw Data	Sampling Rate	512 Hz	250 Hz
	Number of Channels (N_{ch})	32	22
	Number of Classes (C)	3	4
Surface Laplacian (SL)	Filter Order	2	2
	Regularization Factor (λ)	0.01	0.01
Channel Selection (CS)	Selected Centro-Parietal Channels	C3, Cz, C4, CP1, CP2, P3, Pz, P4	Channels 8-12, 14-21
	Number of Selected Channels (N_{ch})	8	13
Bandpass Filter Bank	Filter Type	Chebyshev Type-II	Chebyshev Type-II
	Filter Order	6	6
	Selected Bands	8-12, 12-16, $\cdots$, 28-32 Hz	4-8, 8-12, $\cdots$, 36-40 Hz
	Number of Bandpass Filters (N_f)	6	9
Epoch Segmentation for CSP/SCSSP	Epoch Length	1 sec	2 sec
	Consecutive Epochs' Overlapping Factor	1/2	19/20
Feature Space's Maximum Dimensionality	FBCSP/SCSSP without CS	576(=6*32*3)	792(=9*22*4)
	FBCSP/SCSSP with CS	144(=6*8*3)	468(=9*13*4)

Note that the parameters used for implementation of SL filter are similar to the ones chosen in [31].

at the classifier's input, denoted by d. To determine the optimal value of d, denoted by d_{opt}, for each feature extraction scheme, we perform cross-validation on the training data. In case of Exp. 1, since we have access to three different training sessions, a threefold cross-validation is performed to make sure that for each validation run the BCI system has access to two distinct sessions for training and one session for analyzing the performance. This strategy is helpful in making sure that intersession variations of the EEG data are taken into account during the validation.

The dataset in Exp. 2, however, only contains one training session, which prevents us from adopting the same cross-validation strategy as Exp. 1. Therefore, we chose to perform a 5×5-fold randomized cross-validation strategy. In this strategy, the training data will be randomized five times. After each randomization, the data will be divided into five folds. In each validation run, four of these folds will be used for training the BCI algorithm and the remaining fold will be used for analyzing the resulting performance. This procedure results in five validation runs for each randomization, which leads to a total of $25 = 5 \times 5$ validation runs.

To determine d_{opt} for each feature extraction method, the overall performance of the BCI is calculated for all possible values of d during each validation run. Thus, for each d, we have a total of three performance results in Exp. 1 and 25 results in Exp. 2. For each d, these performance results are averaged over all validation runs, and their corresponding standard error over these runs are calculated. The value of d that results in the maximum average performance is the d_{opt}.

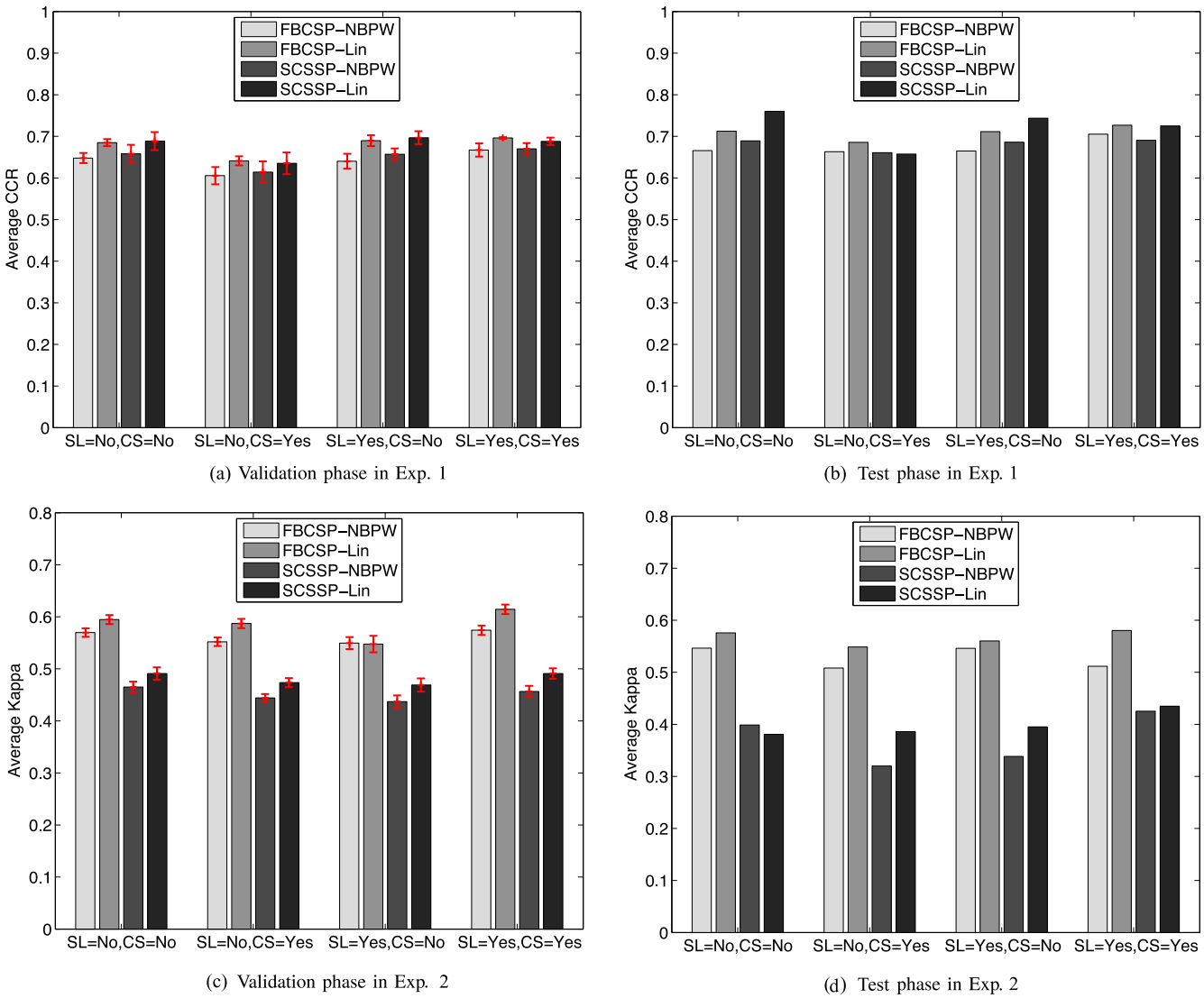
The validation results for Exp. 1 and Exp. 2 are presented in Fig. 2(a) and (c), respectively. These results are presented in groups of size 4, in the following order: FBCSP-NBPW, FBCSP-Lin, SCSSP-NBPW, and SCSSP-Lin. In each figure, the first group of results correspond to the case where no SL or CS is applied. Similarly, the next groups correspond to other possible combinations of SL and CS. For each method, the average validation performance over all the subjects, together with its corresponding standard error, are presented.

Despite the fact that SCSSP has less computational cost compared to FBCSP, Fig. 2(a) shows that for all combinations of SL and CS in Exp. 1, SCSSP-NBPW outperforms FBCSP-NBPW and SCSSP-Lin either outperforms or closely competes with FBCSP-Lin. In Exp. 2, however, Fig. 2(c) illustrates that SCSSP cannot compete with FBCSP-based methods. In order to describe this difference between Exp. 1 and Exp. 2, recall that the most important difference between these two experiments is the availability of training data. Considering these differences discussed in Section IV-A3, the cross-validation results reveal that the performance competency and computational cost efficiency of the SCSSP are achieved at the cost of requiring more training samples, compared to the FBCSP.

Based on our discussions in Section III-B, the higher sensitivity of the SCSSP to the number of training samples can be explained as follows. The FBCSP only focuses on the diagonal block matrices of $\mathbf{\Sigma}_i{}'$ matrices, as defined in (12), whereas SCSSP aims to provide an estimate of both diagonal and off-diagonal block matrices of $\mathbf{\Sigma}_i{}'$ matrices, as defined in (13). Therefore, when the number of training samples is extremely small, SCSSP cannot reliably estimate $\mathbf{\Sigma}_i{}'$, and consequently, does not succeed in extracting discriminant features from the EEG data. The high performance of the SCSSP method in Exp. 1 shows that the matrix-variate Gaussian model deployed by SCSSP algorithm can very well describe the statistical characteristics of the EEG signals; however, reliable estimation of its parameters requires access to a large training set.

These results suggest that in order to benefit from the low computational cost and high performance of the SCSSP, we need to provide this algorithm with enough training samples. This condition is not restrictive in most motor-imagery BCI applications, since these BCIs are generally designed for long-term utilization, which guarantees access to large enough training sets. In such cases, SCSSP can reliably estimate signal parameters, which allows for reducing the computational cost while improving the performance of the BCI system.

2) Test (Competition) Results: The average test performance over all the subjects for unseen competition data in Exp. 1 and



(a) Validation phase in Exp. 1

(b) Test phase in Exp. 1

(c) Validation phase in Exp. 2

(d) Test phase in Exp. 2

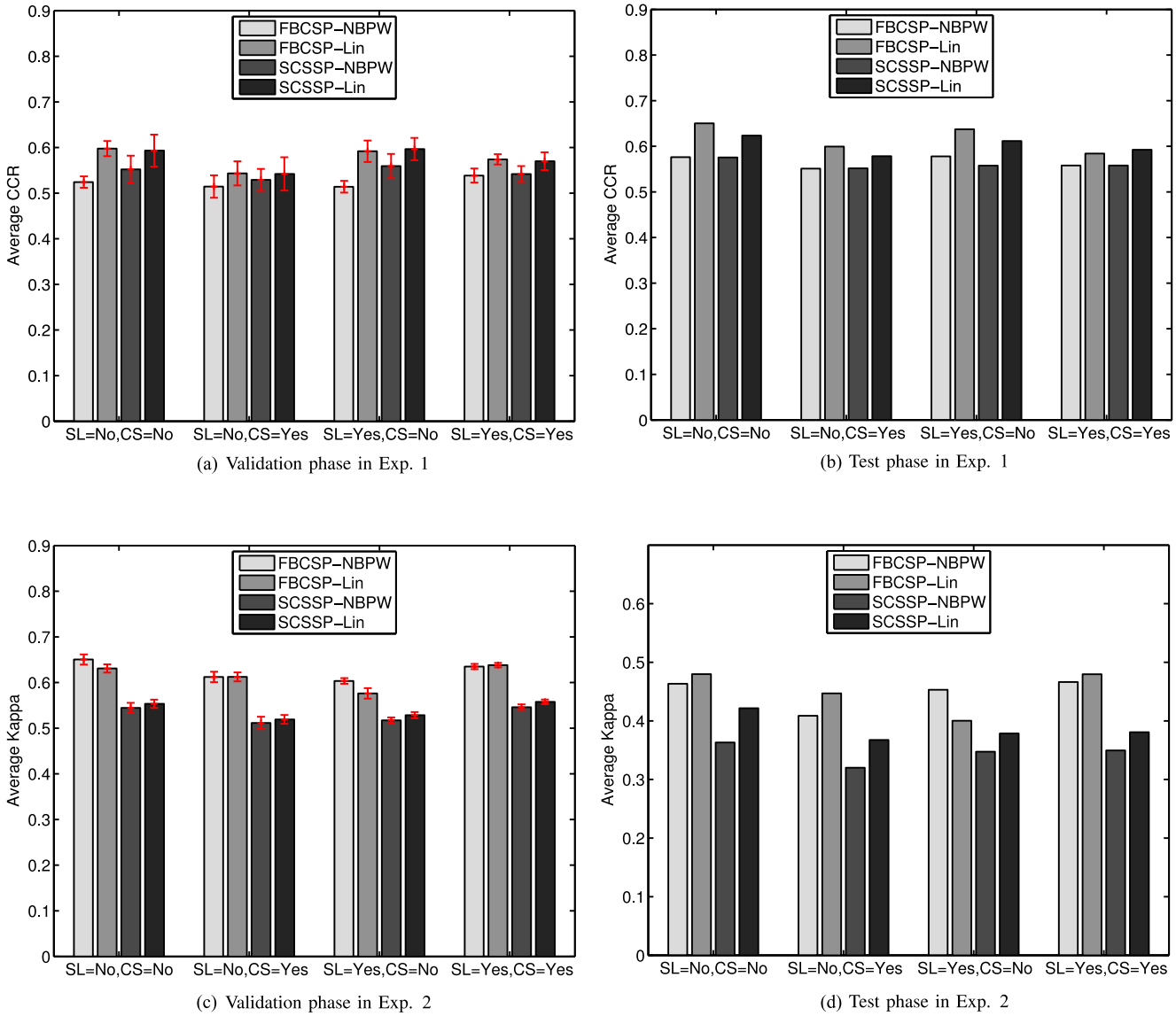
Fig. 2. Comparison of the average performance in two-class motor-imagery classification scenarios (Left hand versus Right hand movement) in Exp. 1 and Exp. 2. The performance results for SCSSP-based and FBCSP-based solutions in (a) validation phase of Exp. 1, (b) testing phase of Exp. 1, (c) validation phase of Exp. 2, and (d) testing phase of Exp. 2 are presented, respectively. For validation results, the average performance of each method over all the subjects and all validation runs, together with its corresponding standard error, is plotted. For more clarity, the results are illustrated in four groups, depending on whether or not the SL and CS are applied to the EEG data. Note that the performance measure is the CCR in Experiment-1, and the Kappa coefficient (κ) in Experiment-2.

Exp. 2 is presented in Fig. 2(b) and (d). The different methods are categorized in groups of size four, depending on whether or not the SL and CS are applied to the data, similar to Fig. 2(a) and (c). For each method and each subject, the value of d_{opt} obtained from the validation phase is used to set the feature space dimensionality in the test phase, and then, the feature extractor and classifier are retrained by the whole training dataset. Since the training/testing procedure has only been performed once in this phase, the results do not include any standard error analysis.

These results show a trend similar to the cross-validation results. In Exp. 1, SCSSP-based methods outperform FBCSP-based methods when no CS is performed and closely compete with FBCSP-based methods in the presence of CS. In Exp. 2, the SCSSP cannot compete with other methods due to the lack of access to sufficient training information for reliable estimation of model parameters.

3) Effect of SL and CS: Comparison of the average performances in Fig. 2 shows that combination of SL with SCSSP has different effects on the classification performance depending on whether or not CS has been applied, as follows. When CS is applied, the SL slightly improves the performance of SCSSP-based methods in both Exp. 1 and Exp. 2. However, when there is no CS, the SL can degrade the performance of SCSSP-based methods.

It can be seen that the FBCSP-Lin method achieves its highest performance when both SL and CS are deployed, whereas SCSSP-Lin achieves its highest performance when it is applied to the raw data, with only one exception which is the test phase of Exp. 2. In case of the FBCSP-Lin method, the combination of SL and CS helps to manually reduce the dimensionality of the space in which the *spatial covariances* $\mathbf{\Psi}_i^f$ are calculated, without losing the local information relevant to the motor cortex



(a) Validation phase in Exp. 1

(b) Test phase in Exp. 1

(c) Validation phase in Exp. 2

(d) Test phase in Exp. 2

Fig. 3. Comparison of average performance in multiclass motor-imagery classification scenarios in Exp. 1 and Exp. 2. The performance results for SCSSP-based and FBCSP-based solutions in (a) validation phase of Exp. 1, (b) testing phase of Exp. 1, (c) validation phase of Exp. 2, and (d) testing phase of Exp. 2 are presented, respectively. For validation results, the average performance of each method over all the subjects and all validation runs, together with its corresponding standard error, is plotted.

area. This dimensionality reduction improves the accuracy of the spatial covariance estimation for each band, which in turn improves the performance of the system.

In the case of SCSSP, however, it is not necessarily desired to reduce the dimensionality of the data in the *spatial domain* while having the same dimensionality in the *spectral domain*. The main reason for this effect is as follows. The SCSSP only calculates one common spatial covariance matrix for all the bands. As a result, SCSSP treats different rows of the matrix $\mathbf{X} \in \mathbb{R}^{N_f \times N_{ch}}$ as extra training samples for calculation of the covariance matrix. In other words, SCSSP method has access to $N_f * N_i$ training samples for estimation of $\mathbf{\Psi}_i$, whereas FBCSP has only access to N_i samples for estimation of each $\mathbf{\Psi}_i^f$. On the other hand, SCSSP requires to estimate the common *spectral covariance* matrix $\mathbf{\Phi}_i$ by treating different columns of $\mathbf{X}$ as extra training samples, which leads to a total of $N_{ch} * N_i$ samples. As a consequence, any reduction in the number of EEG channels results in a significant reduction in the number of training samples for $\mathbf{\Phi}_i$.

In other words, in SCSSP, the CS results in higher accuracy for spatial covariance estimation at the cost of reducing the accuracy for spectral covariance estimation. The experimental results suggest that these two opposite effects almost cancel out each other and there is marginal change in the performance of SCSSP-based methods when CS and SL are utilized together with the SCSSP, as opposed to when SCSSP is directly applied to the raw data. Note that as long as CS does not deteriorate the overall performance, it might still be beneficial since it reduces the computational cost of the feature extraction.

D. Results for Multiclass Scenario

Using the OVR strategy described in Section III-C, this section examines the performance of both SCSSP and FBCSP

TABLE II
CROSS-VALIDATION PERFORMANCE RESULTS FOR DIFFERENT ALGORITHMS IN EXPERIMENT-1

Feature Extraction			Classification	Performance in the Cross-validation Stage						
Spatial		Spatio-		Subject 1		Subject 2		Subject 3		Average
SL	CS	Spectral		%CCR	d_{opt}	%CCR	d_{opt}	%CCR	d_{opt}	%CCR
No	No	FBCSP	NBPW	60.17±3.30	288	50.36±2.08	252	46.75±1.29	216	52.42±6.94
			Lin	69.86±5.14	360	59.54±1.97	252	49.93±2.75	252	59.77±9.97
		SCSSP	NBPW	63.09±6.89	192	52.56±2.38	144	50.00±1.33	126	55.21±6.94
			Lin	67.99±6.31	192	59.43±4.53	144	50.58±2.25	144	59.33±8.71
	Yes	FBCSP	NBPW	64.41±4.38	144	47.34±1.06	144	42.59±2.72	144	51.45±11.48
			Lin	66.45±4.78	144	49.75±2.93	144	46.82±2.46	108	54.34±10.59
		SCSSP	NBPW	64.09±2.23	144	48.38±4.90	144	46.24±1.29	144	52.90±9.75
			Lin	66.06±4.08	120	47.12±5.21	144	49.46±1.95	138	54.21±10.33
Yes	No	FBCSP	NBPW	58.90±3.86	576	49.10±1.36	324	46.20±1.15	360	51.40±6.66
			Lin	69.61±4.59	504	57.70±3.46	324	50.25±1.54	324	59.19±9.76
		SCSSP	NBPW	64.21±4.71	192	54.46±2.47	120	49.13±2.32	144	55.94±7.65
			Lin	66.96±5.42	168	60.51±3.82	120	51.52±1.50	192	59.66±7.75
	Yes	FBCSP	NBPW	64.56±3.57	144	48.20±2.14	144	48.81±1.60	144	53.86±9.28
			Lin	68.28±3.37	144	53.13±1.82	144	50.80±2.86	144	57.40±9.49
		SCSSP	NBPW	63.92±2.26	132	49.53±2.22	144	49.10±2.73	102	54.18±8.43
			Lin	68.07±2.55	132	51.37±3.16	144	51.59±2.58	108	57.01±9.58

For each method and each subject, the optimal dimension (d_{opt}) is presented together with the corresponding average correct classification rate ± standard error over all the validation runs. Note that $d_{\text{opt}} = N_f * d_{\text{csp}} * C$ for FBCSP-based methods and $d_{\text{opt}} = d_{\text{scssp}} * C$ for SCSSP-based methods, where $2 \leq d_{\text{csp}} \leq N_{ch}$ and $2 \leq d_{\text{scssp}} \leq N_f * N_{ch}$.

TABLE III
CROSS-VALIDATION PERFORMANCE RESULTS FOR DIFFERENT ALGORITHMS IN EXPERIMENT-2

Feature Extraction			Classification	Performance in the Cross-Validation Stage (κ and d_{opt})									
Spatial		Spatio-											
SL	CS	Spectral		Subj. 1	Subj. 2	Subj. 3	Subj. 4	Subj. 5	Subj. 6	Subj. 7	Subj. 8	Subj. 9	Average
No	No	FBCSP	NBPW	67.88±2.87 360	42.18±3.35 72	77.87±2.78 144	51.77±3.73 576	50.17±4.20 72	45.97±3.13 648	87.50±2.30 72	85.79±2.03 432	76.31±2.86 144	65.05±1.12
			Lin	68.76±2.56 756	29.21±3.02 432	72.07±2.67 72	49.24±2.45 720	52.54±3.82 72	48.91±3.26 72	84.63±2.40 576	85.96±2.41 756	76.46±2.89 72	63.09±0.90
		SCSSP	NBPW	59.89±3.18 216	28.25±3.05 616	73.02±3.37 88	32.37±2.74 400	35.27±3.85 656	38.11±4.63 672	74.43±2.84 40	80.20±2.53 216	68.54±2.60 8	54.45±1.10
			Lin	64.98±3.40 736	25.65±3.97 688	72.44±2.89 40	32.16±2.90 656	34.43±3.66 192	36.82±3.06 736	73.68±3.51 64	85.60±2.62 32	72.28±2.63 8	55.34±0.92
	Yes	FBCSP	NBPW	65.69±3.06 144	42.32±2.36 72	80.37±2.41 72	40.12±3.71 72	39.04±3.59 576	50.58±3.57 360	78.57±2.84 756	77.67±3.16 144	76.60±3.12 144	61.22±1.15
			Lin	66.96±2.99 144	33.32±3.14 576	80.89±2.69 72	49.14±2.47 576	42.29±3.30 144	46.30±3.79 576	73.01±2.90 756	83.04±2.48 432	76.17±3.09 72	61.24±0.98
		SCSSP	NBPW	53.40±2.54 168	28.74±4.46 464	80.32±2.40 48	36.77±3.00 72	19.59±4.04 424	35.03±4.33 248	60.95±3.33 184	75.81±2.75 328	69.79±3.25 160	51.15±1.36
			Lin	48.16±3.97 464	25.57±3.79 376	83.98±2.30 56	30.36±2.74 32	24.22±3.72 352	37.47±3.24 120	63.12±2.99 264	83.98±2.86 40	70.45±2.61 16	51.92±0.97
Yes	No	FBCSP	NBPW	71.45±1.65 144	40.10±1.78 144	71.00±1.72 144	39.30±2.05 144	57.56±1.86 72	30.37±2.26 216	81.43±1.23 72	77.98±0.94 144	73.93±1.71 216	60.35±0.62
			Lin	70.54±4.41 72	35.11±1.71 144	69.43±1.57 216	41.93±2.41 144	40.79±7.40 72	31.71±2.82 360	79.30±1.35 288	79.93±1.00 360	69.79±1.94 216	57.61±1.15
		SCSSP	NBPW	63.48±1.50 192	35.27±1.59 488	65.12±1.68 104	35.21±1.89 752	31.78±2.21 128	25.84±1.94 560	68.78±1.59 128	72.28±1.13 160	67.68±1.66 232	51.72±0.65
			Lin	65.50±1.59 768	31.53±1.63 792	65.45±1.23 88	35.50±2.11 776	38.70±2.14 144	27.96±1.76 784	69.86±1.50 184	75.87±1.33 416	65.19±1.80 80	52.84±0.68
	Yes	FBCSP	NBPW	74.26±1.56 216	40.09±1.30 72	74.26±1.38 72	43.01±2.22 216	61.68±1.64 72	35.86±1.94 216	85.74±0.99 72	81.30±1.14 144	75.35±1.52 72	63.51±0.57
			Lin	76.37±1.52 144	39.70±1.63 216	78.13±1.26 144	46.67±2.22 216	59.49±1.72 144	35.39±2.29 216	84.74±1.50 216	82.38±1.14 144	71.56±1.77 72	63.82±0.49
		SCSSP	NBPW	66.11±1.68 120	36.94±1.68 464	69.22±1.44 8	35.98±1.83 368	33.49±1.75 48	30.77±2.33 120	74.45±1.77 96	74.20±1.08 88	70.29±2.04 224	54.60±0.60
			Lin	67.27±1.44 456	35.93±1.67 448	68.28±1.40 88	39.13±2.05 424	41.76±1.73 448	30.81±2.04 384	74.39±1.38 464	77.16±1.04 40	67.21±1.57 80	55.77±0.51

For each method and each subject, the optimal dimension (d_{opt}) is presented at the bottom row, whereas the top row presents the corresponding average correct classification rate ± standard error over all the validation runs. Note that $d_{\text{opt}} = N_f * d_{\text{csp}} * C$ for FBCSP-based methods and $d_{\text{opt}} = d_{\text{scssp}} * C$ for SCSSP-based methods, where $2 \leq d_{\text{csp}} \leq N_{ch}$ and $2 \leq d_{\text{scssp}} \leq N_f * N_{ch}$.

TABLE IV
TEST PERFORMANCE RESULTS FOR DIFFERENT ALGORITHMS IN EXP. 1

Feature Extraction			Classifier	Performance in Test Stage (%CCR)			
Spatial		Spatio-					
SL	CS	Spectral		Subj. 1	Subj. 2	Subj. 3	Average
No	No	FBCSP	NBPW	72.51	57.99	42.37	57.62
			Lin	78.25	68.90	47.96	65.03
		SCSSP	NBPW	73.92	60.26	38.49	57.56
			Lin	77.81	65.33	43.87	62.34
	Yes	FBCSP	NBPW	69.91	50.43	44.95	55.10
			Lin	73.05	55.08	51.72	59.95
		SCSSP	NBPW	70.24	47.95	47.31	55.17
			Lin	74.13	49.35	50.00	57.83
Yes	No	FBCSP	NBPW	69.70	59.72	43.98	57.80
			Lin	76.41	67.71	47.10	63.74
		SCSSP	NBPW	73.16	55.83	38.28	55.76
			Lin	76.84	60.15	46.45	61.15
	Yes	FBCSP	NBPW	72.08	50.76	44.52	55.78
			Lin	73.48	54.21	47.53	58.41
		SCSSP	NBPW	70.13	52.38	44.84	55.78
			Lin	73.59	55.72	48.39	59.23

methods in multiclass motor-imagery scenarios ($C = 3$ in Exp. 1 and $C = 4$ in Exp. 2).

1) Cross-Validation Results: Fig. 3(a) and (c) present the average validation performance over all the subjects, together with its corresponding standard error, in Exp. 1 and Exp. 2, respectively. Similar to the results presented in Section IV-C1, in the multiclass scenario the SCSSP-based methods outperform FBCSP-based algorithms in Exp. 1 while providing a better computational efficiency. Also, similar to the two-class scenario, in the multiclass case SCSSP cannot compete with FBCSP in Exp. 2 due to small size of training set (ref. Section IV-C1).

For completeness of the reported results, we have provided detailed results of validation phase for each participant in Tables II and III. These results are presented in groups of size 4, in the following order: FBCSP-NBPW, FBCSP-Lin, SCSSP-NBPW, and SCSSP-Lin. In each table, the first group of results correspond to the case where no SL or CS is applied. Similarly, the next groups correspond to other possible combinations of SL and CS. These results reveal that despite the high variations of the performances between different subjects, the relative performance of different methods is consistent among all the subjects.

2) Test (Competition) Results: Fig. 3(b) and (d) present the average test performance over all the subjects for unseen competition data in Exp. 1 and Exp. 2, respectively. These results have a trend similar to the two-class results in Section IV-C2, except for the fact that in Exp. 1 SCSSP-based methods can no longer outperform FBCSP-based ones when no CS is applied. However, it should be noted that even in these cases SCSSP is still closely following FBCSP's performance while providing a significant reduction in the computational cost.

For the reader's reference, the corresponding detailed results are provided in Tables IV and V. Similar to the validation results in Section IV-D1, these test results show a consistent relative performance among all the subjects.

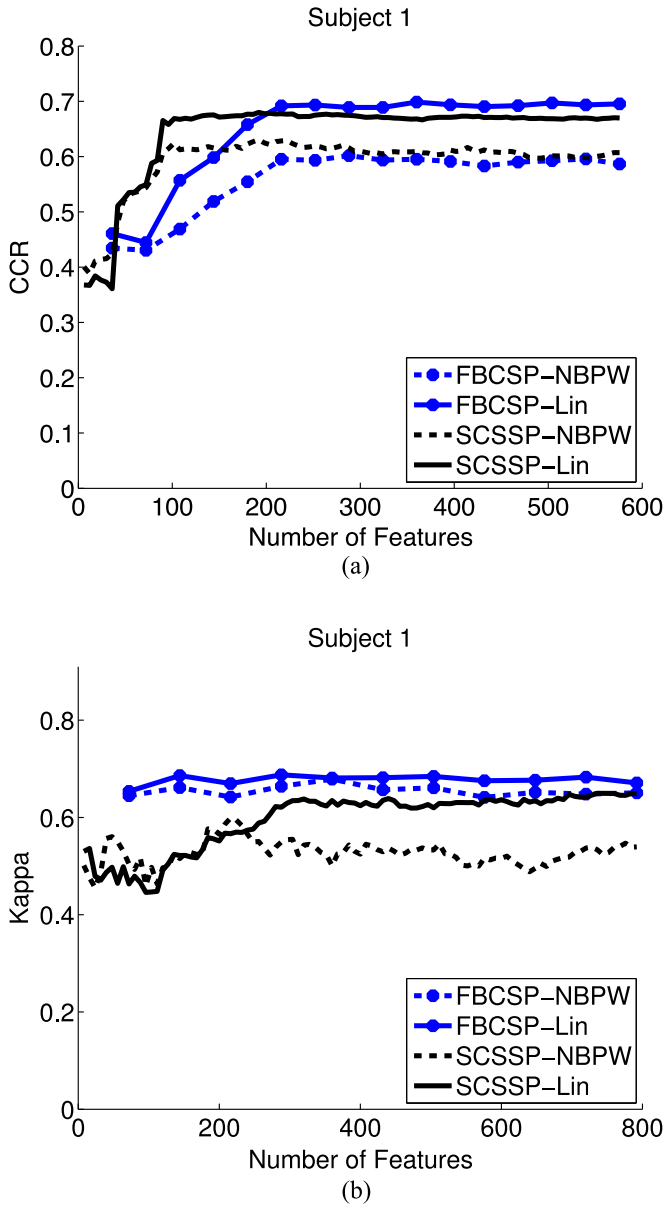


Fig. 4. Validation performance for the SCSSP-based and FBCSP-based methods versus the number of features extracted by the feature extraction method in the validation phase for the first subject in (a) Experiment-1 and (b) Experiment-2.

E. Effect of Feature Space Dimensionality

In this section, we study the effect of number of extracted features on the performance of SCSSP-NBPW and SCSSP-Lin algorithms, and compare them with the FBCSP-NBPW and FBCSP-Lin solutions. The results for the first subjects in Exp. 1 and Exp. 2 are shown in Fig. 4. Similar trends are observed for other subjects as well. These results correspond to the case where no SL or CS is applied to the EEG.

The results of Exp. 1 in Fig. 4(a) show that SCSSP-Lin outperforms FBCSP-Lin when $d < 200$ and closely follows its performance with a marginal gap when $d > 200$. In contrast, SCSSP-NBPW has much lower performance than SCSSP-Lin

TABLE V
TEST PERFORMANCE RESULTS FOR DIFFERENT ALGORITHMS IN EXPERIMENT-2

Feature Extraction				Performance in the Test Stage (κ)									
Spatial		Spatio-Spectral	Classification										
SL	CS			Subj. 1	Subj. 2	Subj. 3	Subj. 4	Subj. 5	Subj. 6	Subj. 7	Subj. 8	Subj. 9	Average
No	No	FBCSP	NBPW	63.67	31.20	60.10	36.73	23.33	19.43	64.20	57.07	61.39	46.35
			Lin	68.07	30.25	70.68	39.34	28.42	25.01	56.89	59.13	54.06	47.98
		SCSSP	NBPW	50.12	21.03	62.16	31.40	0.00	12.13	33.35	53.31	63.30	36.31
			Lin	62.25	26.54	65.76	26.91	6.64	25.56	40.65	59.32	65.86	42.17
	Yes	FBCSP	NBPW	60.72	25.06	56.65	34.97	9.61	18.10	55.56	53.92	53.33	40.88
			Lin	64.46	26.79	60.81	39.07	11.19	24.63	66.53	53.97	54.84	44.70
		SCSSP	NBPW	59.42	17.34	57.17	14.83	-1.95	8.77	28.63	52.13	51.78	32.01
			Lin	59.16	24.64	56.83	15.13	5.30	21.14	34.16	58.50	55.90	36.75
Yes	No	FBCSP	NBPW	64.11	32.17	65.95	43.33	27.37	22.15	64.58	31.93	56.23	45.31
			Lin	69.31	36.63	69.52	48.49	14.20	27.82	61.17	-24.06	57.17	40.03
		SCSSP	NBPW	51.19	19.85	52.04	26.60	14.58	18.89	38.41	42.56	48.62	34.75
			Lin	58.12	27.69	62.55	32.57	4.79	21.91	26.52	54.74	51.87	37.86
	Yes	FBCSP	NBPW	64.87	41.45	60.56	40.46	21.82	19.85	65.74	54.56	50.38	46.63
			Lin	67.55	40.17	66.60	43.54	23.69	25.84	62.14	53.12	49.05	47.97
		SCSSP	NBPW	57.19	25.07	59.32	26.10	8.52	22.53	13.28	50.87	51.81	34.97
			Lin	59.09	32.34	58.72	25.41	5.27	23.65	23.72	59.48	54.99	38.07

The average Kappa coefficient (κ) for each subject and the total average over all the subjects are reported.

TABLE VI
RUN TIME FOR TRAINING FBSCP AND SCSSP ALGORITHMS IN EXP. 1

Feature Extraction		Number of	Run Time (millisecond)			
Spatio-Spectral	CS	Parameters	Subj. 1	Subj. 2	Subj. 3	Average
FBCSP	No	9504	12.75	12.82	12.86	12.81
	Yes	648	3.16	3.11	3.22	3.16
SCSSP	No	1647	2.28	2.27	2.28	2.28
	Yes	171	0.73	0.72	0.72	0.72

but still outperforms FBCSP-NBPW. Note that the performance of SCSSP-Lin method peaks at a relatively low dimension, which shows that SCSSP has been able to capture the discriminant information of the data in a small number of features. In Exp. 2, where the training set is extremely limited, SCSSP-based solutions cannot compete with the FBCSP-based ones. This significant difference between the two experiments is mostly due to the lack of training data in Exp. 2, as discussed in Section IV-C1.

F. Run Time

Tables VI and VII provide the average runtimes for training SCSSP versus FBCSP in Exp. 1 and Exp. 2. The reported times are in milliseconds and are based on executing the corresponding MATLAB codes on a system with eight X5355 processors and 8-GB memory. For each experiment, the amount of time required for training each of the subjects during the validation phase has been calculated, and the resulting execution times are averaged over all validation runs and reported in Table VI. To have a fair comparison, the reported times only correspond to the amount of time required for calculation of transformation matrices in the FBCSP and SCSSP methods through eigen-decomposition of spatial/spectral covariance matrices, and they do not include the time required for executing the SL, Bandpass filtering, covariance estimation, and classification since these steps are common in both FBCSP and SCSSP.

This table illustrates the significant difference between the running time for the FBCSP and SCSSP. Indeed, the proposed SCSSP algorithm is approximately five times faster than FBCSP in Exp. 1, and seven times faster in Exp. 2. Note that the use of CS changes the dimensionality of the input data, and hence, the computational cost of the FBCSP/SCSSP. In fact, SCSSP involves only two generalized eigen decompositions (i.e., one for spectral covariance and one for spatial covariance); whereas FBCSP requires a total of N_f generalized eigen decompositions (i.e., one for each frequency band).

It is noteworthy that the difference between the runtimes of FBCSP and SCSSP is crucially important in various applications, such as the following. (a) *Adaptive Scenarios:* In many applications, the statistical properties of EEG changes over time, e.g., during long data collection sessions or across different sessions (ref. [37]). A widely used solution for such scenarios is to utilize adaptive methods to continuously track these changes and accordingly update the feature extractor. For such solutions, it is critically important to reduce the computational cost of the training phase in each update. Although neither FBCSP nor SCSSP are adaptive solutions, they can be modified to be used as a basis for an adaptive algorithm. In such cases, the computational cost of feature extractor's training dictates the overall computational cost of the system. (b) *Mobile/Wearable Scenarios:* Motor-imagery BCI systems are mostly expected to be used in daily life applications where the processing is done in a portable device. Besides, during the past few years there has been a growing interest in processing of EEG signals in mobile phones or tablets for various applications, such as mediation, learning, vigilance monitoring, etc. In such cases, the processing power is essentially restricted, which necessitates the use of computationally efficient methods. Indeed, in the aforementioned scenarios and many other similar applications, matrix-variate treatment of the EEG data provides a promising framework that has the

TABLE VII
RUN TIME FOR TRAINING FBSCP AND SCSSP ALGORITHMS IN EXP. 2

Feature Extraction		Number of	Run Time (ms)									
Spatio-Spectral	CS	Parameters	Subj. 1	Subj. 2	Subj. 3	Subj. 4	Subj. 5	Subj. 6	Subj. 7	Subj. 8	Subj. 9	Average
FBCSP	No	9108	17.40	16.59	17.46	17.78	17.07	17.82	15.85	16.78	16.53	17.03
	Yes	3276	10.68	11.21	12.03	10.43	11.55	12.87	9.85	10.60	10.79	11.11
SCSSP	No	1192	2.08	2.15	2.73	2.04	2.13	3.06	1.98	2.08	2.10	2.26
	Yes	544	1.35	1.37	2.01	1.35	1.35	2.34	1.30	1.37	1.38	1.53

potential to be utilized in various types of BCI algorithms (e.g., ref. [38] and [39]).

V. CONCLUSION

This paper proposed a new feature extraction method based on a heteroscedastic matrix-variate Gaussian model for the multi-band EEG rhythms. In the proposed approach, EEG is first passed through a bank of bandpass filters to extract different bands of EEG rhythms. The resulting signal is then passed through a joint spatio-spectral feature extractor, called SCSSP, which directly operates on matrix-variate data.

SCSSP's main advantage over FBCSP is that SCSSP jointly processes the data in both spectral and spatial domains, and hence, can sort the extracted features across both domains; whereas FBCSP cannot sort the features that are extracted from different frequency bands. As a result, SCSSP does not require a separate subsequent dimensionality reduction stage, and its output can directly be passed to the classifier. The second advantage of SCSSP is its relatively low computational cost.

The aforementioned advantages come at the cost that SCSSP requires a relatively larger training set, compared to the FBCSP. Unlike Exp. 1, SCSSP cannot compete with the FBCSP in Exp. 2 since the amount of training information is extremely limited due to: 1) short length of each trial (3 versus 15 seconds), 2) single session data collection, and 3) limited number of training samples (288 versus 1392 epochs). One possible solution is to use multisubject regularization.

The appropriate length for training trials depends on several factors, such as 1) subject's ability in controlling motor activities, 2) subject's previous experience with BCIs, and 3) presence of real-time feedback. Furthermore, although in longer training trials, the user has more time to concentrate and produce reliable training samples, if the trial length is unreasonably long, the subject might lose concentration toward the end of trial and generate noisy patterns. Further studies are needed to analyze these effects, which are beyond the scope of this paper.

APPENDIX A
PROOF OF THEOREM 1

Let $\mathbf{\Sigma}_i = \mathbf{\Psi}_i \otimes \mathbf{\Phi}_i$ for $i \in \{1,2\}$ and $\mathbf{W} = \mathbf{W}_R \otimes \mathbf{W}_L$, where $\mathbf{W}_R$ and $\mathbf{W}_L$ satisfy the following equations:

$$\mathbf{\Psi}_1 \mathbf{W}_R = (\mathbf{\Psi}_1 + \mathbf{\Psi}_2)\, \mathbf{W}_R \mathbf{\Lambda}_R \tag{14}$$

$$\mathbf{\Phi}_1 \mathbf{W}_L = (\mathbf{\Phi}_1 + \mathbf{\Phi}_2)\, \mathbf{W}_L \mathbf{\Lambda}_L. \tag{15}$$

In this section, we will prove that

$$\mathbf{\Sigma}_1 \mathbf{W} = (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\, \mathbf{W} \mathbf{\Lambda} \tag{16}$$

where

$$\mathbf{\Lambda} = (\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L)\left(\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L + (\mathbf{I}_{N_{ch}} - \mathbf{\Lambda}_R) \otimes (\mathbf{I}_{N_f} - \mathbf{\Lambda}_L)\right)^{-1}.$$

Based on (14), it can be easily shown that $\mathbf{W}_R$ jointly diagonalizes both $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$. Similarly, $\mathbf{W}_L$ jointly diagonalizes both $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$. Therefore, we have

$$\mathbf{\Psi}_1 \mathbf{W}_R = \mathbf{W}_R \mathbf{\Lambda}_R^{(1)}, \quad \mathbf{\Psi}_2 \mathbf{W}_R = \mathbf{W}_R \mathbf{\Lambda}_R^{(2)} \tag{17}$$

$$\mathbf{\Phi}_1 \mathbf{W}_L = \mathbf{W}_L \mathbf{\Lambda}_L^{(1)}, \quad \mathbf{\Phi}_1 \mathbf{W}_L = \mathbf{W}_L \mathbf{\Lambda}_L^{(2)} \tag{18}$$

where $\mathbf{\Lambda}_R^{(i)}$ and $\mathbf{\Lambda}_L^{(i)}$ are the corresponding diagonal eigenvalue matrices. By substituting (17)–(18) into (14)–(15), we have

$$\mathbf{\Lambda}_R^{(1)} = \left(\mathbf{\Lambda}_R^{(1)} + \mathbf{\Lambda}_R^{(2)}\right)\mathbf{\Lambda}_R,\ \mathbf{\Lambda}_L^{(1)} = \left(\mathbf{\Lambda}_L^{(1)} + \mathbf{\Lambda}_L^{(2)}\right)\mathbf{\Lambda}_L \tag{19}$$

or equivalently

$$\mathbf{\Lambda}_R^{(1)}\left(\mathbf{\Lambda}_R^{(1)} + \mathbf{\Lambda}_R^{(2)}\right)^{-1} = \mathbf{\Lambda}_R \tag{20}$$

$$\mathbf{\Lambda}_R^{(2)}\left(\mathbf{\Lambda}_R^{(1)} + \mathbf{\Lambda}_R^{(2)}\right)^{-1} = \mathbf{I}_{N_{ch}} - \mathbf{\Lambda}_R \tag{21}$$

$$\mathbf{\Lambda}_L^{(1)}\left(\mathbf{\Lambda}_L^{(1)} + \mathbf{\Lambda}_L^{(2)}\right)^{-1} = \mathbf{\Lambda}_L \tag{22}$$

$$\mathbf{\Lambda}_L^{(2)}\left(\mathbf{\Lambda}_L^{(1)} + \mathbf{\Lambda}_L^{(2)}\right)^{-1} = \mathbf{I}_{N_f} - \mathbf{\Lambda}_L. \tag{23}$$

Using these definitions, the left-hand side of (16) can be expanded as follows:

$$\begin{aligned}\mathbf{\Sigma}_1 \mathbf{W} &= (\mathbf{\Psi}_1 \otimes \mathbf{\Phi}_1)(\mathbf{W}_R \otimes \mathbf{W}_L) = (\mathbf{\Psi}_1 \mathbf{W}_R) \otimes (\mathbf{\Phi}_1 \mathbf{W}_L) \\ &= \left(\mathbf{W}_R \mathbf{\Lambda}_R^{(1)}\right) \otimes \left(\mathbf{W}_L \mathbf{\Lambda}_L^{(1)}\right) \\ &= (\mathbf{W}_R \otimes \mathbf{W}_L)\left(\mathbf{\Lambda}_R^{(1)} \otimes \mathbf{\Lambda}_L^{(1)}\right).\end{aligned} \tag{24}$$

Similarly, the right-hand side of (16) can be expanded

$$\begin{aligned}(\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2)\, \mathbf{W} \mathbf{\Lambda} &= (\mathbf{W}_R \otimes \mathbf{W}_L)\left(\mathbf{\Lambda}_R^{(1)} \otimes \mathbf{\Lambda}_L^{(1)}\right)\mathbf{\Lambda} \\ &\quad + (\mathbf{W}_R \otimes \mathbf{W}_L)\left(\mathbf{\Lambda}_R^{(2)} \otimes \mathbf{\Lambda}_L^{(2)}\right)\mathbf{\Lambda} \\ &= (\mathbf{W}_R \otimes \mathbf{W}_L) \\ &\quad \cdot \left(\mathbf{\Lambda}_R^{(1)} \otimes \mathbf{\Lambda}_L^{(1)} + \mathbf{\Lambda}_R^{(2)} \otimes \mathbf{\Lambda}_L^{(2)}\right)\mathbf{\Lambda}.\end{aligned} \tag{25}$$

Therefore, (24) is equal to (25) when

$$\mathbf{\Lambda} = \left(\mathbf{\Lambda}_R^{(1)} \otimes \mathbf{\Lambda}_L^{(1)}\right)\left(\mathbf{\Lambda}_R^{(1)} \otimes \mathbf{\Lambda}_L^{(1)} + \mathbf{\Lambda}_R^{(2)} \otimes \mathbf{\Lambda}_L^{(2)}\right)^{-1} \tag{26}$$

$$= \left(\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L\right)\left(\mathbf{\Lambda}_R \otimes \mathbf{\Lambda}_L + \left(\mathbf{I}_{N_{ch}} - \mathbf{\Lambda}_R\right) \otimes \left(\mathbf{I}_{N_f} - \mathbf{\Lambda}_L\right)\right)^{-1} \tag{27}$$

where (27) is obtained by first substituting $\mathbf{\Lambda}_R^{(i)}$ and $\mathbf{\Lambda}_L^{(i)}$ matrices into (26) with the right-hand sides of the two equations in (19), and then, using the four equalities provided by (20)–(23).

REFERENCES

[1] A. Aghaei *et al.*, "Separable common spatio-spectral pattern algorithm for classification of EEG signals," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 988–992.

[2] D. Huang *et al.*, "Electroencephalography (EEG)-based brain-computer interface (BCI): A 2-D virtual wheelchair control based on event-related desynchronization/ synchronization and state control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 3, pp. 379–388, May 2012.

[3] G. Muller-Putz *et al.*, "EEG-based neuroprosthesis control: A step towards clinical practice," *Neurosci. Lett.*, vol. 382, no. 1-2, pp. 169–174, 2005.

[4] R. Scherer *et al.*, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 979–984, Jun. 2004.

[5] G. Pfurtscheller *et al.*, "Walking from thought," *Brain Res.*, vol. 1071, no. 1, pp. 145–152, 2006.

[6] A. D. Gerson *et al.*, "Cortically coupled computer vision for rapid image search," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 174–179, Jun. 2006.

[7] L. C. Parra *et al.*, "Spatiotemporal linear decoding of brain state," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 107–115, 2008.

[8] A. S. Royer *et al.*, " EEG control of a virtual helicopter in 3-dimensional space using intelligent control strategies," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 6, pp. 581–589, Dec. 2010.

[9] A. J. Doud *et al.*, " Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface," *PloS one*, vol. 6, no. 10, p. e26322, 2011.

[10] K. LaFleur *et al.*, "Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface," *J. Neural Eng.*, vol. 10, no. 4, p. 046003, 2013.

[11] H. Yuan and B. He, " Brain–computer interfaces using sensorimotor rhythms: Current state and future perspectives," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1425–1435, May 2014.

[12] G. Pfurtscheller and F. H. Lopes da Silva, " Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[13] H. Ramoser *et al.*, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.

[14] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[15] K. J. Miller *et al.*, " Spectral changes in cortical surface potentials during motor movement," *J. Neurosci.*, vol. 27, no. 9, pp. 2424–2432, 2007.

[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA, USA: Academic, 1990.

[17] B. Blankertz *et al.*, " Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[18] C. Neuper and G. Pfurtscheller, " Event-related dynamics of cortical rhythms: Frequency-specific features and functional correlates," *Int. J. Psychophysiol.*, vol. 43, no. 1, pp. 41–58, 2001.

[19] G. Pfurtscheller *et al.*, " Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[20] S. Lemm *et al.*, " Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[21] G. Dornhege *et al.*, " Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 11, pp. 2274–2281, Nov. 2006.

[22] W. Wu *et al.*, " 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.

[23] H. Zhang *et al.*, " Optimum spatio-spectral filtering network for brain-computer interface," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 52–63, Jan. 2011.

[24] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in brain-computer interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.

[25] K. K. Ang *et al.*, " Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Front. Neurosci.*, Mar. 2012.

[26] R. Tomioka *et al.*, " Spectrally weighted common spatial pattern algorithm for single trial EEG classification," *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, vol. 40, 2006.

[27] H. Higashi and T. Tanaka, " Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.

[28] J. Meng *et al.*, " Simultaneously optimizing spatial spectral features based on mutual information for EEG Classification," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 227–240, Jan. 2015.

[29] F. Qi, Y. Li, and W. Wu, " RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG," *IEEE Trans. Neural Netw. Learn. Syst.*, 2015, doi:10.1109/TNNLS.2015.2402694.

[30] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London, U.K.: Chapman & Hall, 1999.

[31] J. Millan, "On the need for on-line learning in brain-computer interfaces," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2004, vol. 4, pp. 2877–2882.

[32] M. Tangermann *et al.*, "Review of the BCI competition IV," *Front. Neurosci.*, Jul. 2012.

[33] F. Cincotti *et al.*, "The use of EEG modifications due to motor imagery for brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 131–133, Jun. 2003.

[34] P. L. Nunez and R. Srinivasan, *Electric fields of the brain : the neurophysics of EEG*, 2nd ed. New York, NY, USA: Oxford Univ. Press, 2006.

[35] F. Galán *et al.*, "Using mental tasks transitions detection to improve spontaneous mental activity classification," *Med. Bio. Eng. Comput.*, vol. 45, no. 6, pp. 603–609, 2007.

[36] A. Schlogl *et al.*, *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Presss, 2007.

[37] M. Krauledat, " Analysis of nonstationarities in EEG signals for improving brain-computer interface performance," Ph.D. dissertation, Berlin Institute of Technology, 2008.

[38] M. S. Mahanta *et al.*, "Regularized LDA based on separable scatter matrices for classification of spatio-spectral EEG patterns," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 1237–1241.

[39] R. Tomioka and K.-R. Müller, "A regularized discriminative framework for EEG analysis with application to brain–computer interface," *NeuroImage*, vol. 49, no. 1, pp. 415–432, 2010.

Amirhossein S. Aghaei (S'06–M'14) is currently a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Canada. His research interests are machine learning, pattern recognition, data analytics, statistical signal processing, and biomedical engineering. He received his Ph.D. and M.A.Sc degrees from the Department of Electrical and Computer Engineering at the University of Toronto, in 2008 and 2013 respectively, and received his B.Sc. degree in Electrical Engineering from Iran University of Science and Technology, Tehran, Iran, in 2006.

Mohammad Shahin Mahanta (S'09) is a research engineer with Interaxon, Inc. His interests include statistical signal processing and machine learning, with particular emphasis on analysis of biomedical signals. He received his Ph.D. and M.A.Sc. degrees in electrical and computer engineering from the University of Toronto, Toronto, Canada, in 2015 and 2009 respectively, and had received a B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2007.

Konstantinos N. Plataniotis (S'93–M'95–SM'03–F'12) is currently a Professor and the Bell Canada Chair of Multimedia with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada. His research interests include statistical signal processing, knowledge and digital media design, multimedia systems, biometrics, image and signal processing, biomedical signal processing, and pattern recognition. Among his publications in these fields are the recent books *WLAN Positioning Systems* (2012) and *Multilinear Subspace Learning: Reduction of Multidimensional Data* (2013).

Prof. Plataniotis received the IEEE Canada Engineering Educator Award for contributions to engineering education and inspirational guidance of graduate students. He is a registered Professional Engineer in Ontario and a Fellow of the Engineering Institute of Canada. He has served as the Editor-in-Chief of the *IEEE Signal Processing Letters* and the Technical Cochair of the 2013 IEEE International Conference in Acoustics, Speech, and Signal Processing. He is the IEEE Signal Processing Society Vice President for Membership (2014–2016) and the General Chair of the forthcoming 2018 IEEE International Conference on Image Processing.