Feature Weighting and Regularization of Common Spatial Patterns in EEG-Based Motor Imagery BCI

Vasilisa Mishuhina and Xudong Jiang, *Senior Member, IEEE*

***Abstract*—Electroencephalography signals have very low spatial resolution and electrodes capture signals that are overlapping each other. To extract the discriminative features and alleviate overfitting problem for motor imagery brain–computer interface (BCI), spatial filtering is widely applied but often only very few common spatial patterns (CSP) are selected as features while ignoring all others. However, using only few CSP features, though alleviates overfitting problem, loses the discriminating information, which limits the BCI performance. This letter proposes a novel feature weighting and regularization (FWR) method that utilizes all CSP features to avoid information loss. The proposed method can be applied in all CSP-based approaches. Experiments of this letter show the effect of the proposed method applied in the standard CSP and its two extensions, common spatio-spectral patterns and regularized CSP. Results on BCI Competition III Dataset IIIa and IV Dataset IIa demonstrate that the proposed FWR method enhances the classification accuracy comparing to the conventional feature selection approaches.**

***Index Terms*—Brain–computer interface (BCI), common spatial patterns (CSP), electroencephalography (EEG), motor imagery (MI).**

I. Introduction

Brain–computer interface (BCI) [1], [2] is a promising technique that analyzes and translates brain signals to a computer commands. Instead of normal communication through the central nervous system and muscles, people can use BCI to communicate directly from brain to computer. It has been successfully used in various applications, e.g., communication, rehabilitation, and control, such as wheelchair control, prosthesis control, robot control, and computer cursor control. One of the best types of BCI is motor imagery BCI (MI-BCI). Thanks to the high portability and accessible cost, electroencephalography (EEG) is the most popular method for recording the electrical activities of the brain. A critical issue in EEG-based MI-BCI is extracting right features from noisy and nonstationary EEG signal with low spatial resolution.

Common spatial pattern (CSP) [3] is one of the most popular feature extraction algorithms in MI-BCI. It finds spatial filters that maximize the variance of a class and minimize the variance of the other class. However, CSP suffers from problems as overfitting and sensitivity to outliers. Moreover, the method only identifies the spatial information, not the spectral one. To enhance performance of CSP, many extensions were developed. They can be broadly classified in two directions: spatio-spectral analyzing and regularization. The first one tries to find more discriminative information in the time or spectral domain, whereas the other tries to solve the overfitting problem during the optimization of spatial filters.

In the first direction of CSP extension, approaches includes common spatio-spectral patterns (CSSP) [4], common sparse spectral spatial patterns [5], spatial and finite-impulse response filters spectrally weighted CSP [6] and iterative spatio-spectral patterns learning [7]. More complex approaches use multiple spectral filters or filter bank to optimize the spectral filters. They includes the discriminative filter bank CSP [8], subband CSP [9], filter bank CSP [10], optimal spatio-spectral filter network [11], maximizing mutual information of features [12], separable CSSP [13], and probabilistic Bayesian framework [14]. In the second direction of CSP extension, approaches include regularized CSP (RCSP) [15], probabilistic CSP [16], stationary CSP [17], and robust CSP based on divergence [18].

All these approaches select features for classification based on the priorities of features determined by the CSP or its extensions. However, in most cases, the best performances of these approaches are achieved by using only very few features. The loss of discriminating information by discarding most CSP features may adversely affect the classification accuracy. Motivated by this, we propose a novel feature weighting and regularization (FWR) method that utilizes all CSP features to avoid losing information. Another example that uses all data in classification is sparse representation [19], [20]. The proposed method can be applied in all CSP-based approaches. To show this, this letter applies the proposed FWR method to the standard CSP and its two extensions in the two different directions: first, CSSP; and second, RCSP.

II. Feature Extraction From CSP, CSSP, And RCSP

A. Common Spatial Patterns

CSP [3] is one of the most popular algorithms in MI-BCI for spatial filtering to extract features. It finds spatial filters that maximize the variance of one class and minimize the variance of the other class. Let a matrix $\mathbf{X} \in R^{ch \times T}$ capture a trial of bandpass filtered EEG signals that characterizes an imaginary movement, where ch is the number of channels of EEG signals

Manuscript received February 16, 2018; revised March 27, 2018; accepted March 28, 2018. Date of publication April 5, 2018; date of current version April 26, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Waheed U. Bajwa. *(Corresponding author: Xudong Jiang.)*

The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: vasilisa001@e.ntu.edu.sg; exdjiang@ntu.edu.sg).

Digital Object Identifier 10.1109/LSP.2018.2823683

and T the signal length of a trial. The normalized covariance matrix of a trial is computed by

$$\boldsymbol{\Sigma} = \frac{\mathbf{X}\mathbf{X}^T}{\mathrm{tr}(\mathbf{X}\mathbf{X}^T)}. \tag{1}$$

If the training set contains N_j trials for class Ω_j, the average covariance matrix of class Ω_j is computed by

$$\mathbf{C}_j = \frac{1}{N_j}\sum_{\Sigma\in\Omega_j}\boldsymbol{\Sigma}. \tag{2}$$

We can find a spatial filter $\mathbf{w} \in R^{ch\times 1}$ that maximizes the variance of Ω_1 and minimizes it of Ω_2 by solving

$$\max J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{C}_1\mathbf{w}}{\mathbf{w}^T(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}}. \tag{3}$$

Solution (3) is obtained by solving the generalized eigenvalue problem as follows:

$$\mathbf{C}_1\mathbf{w} = \lambda(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w} \tag{4}$$

where λ and $\mathbf{w}$ are the generalized eigenvalue and eigenvector, respectively. Equation (4) yields ch eigenvectors $\mathbf{w}_k$ and eigenvalues λ_k, $k = 1,\ldots,ch$.

The output of a spatial filter $\mathbf{w}_k$ to a trial of EEG signals is

$$\mathbf{y}_k = \mathbf{w}_k^T\mathbf{X}. \tag{5}$$

The logarithm of the filtered signal variance serves as a feature f_k for classification

$$f_k = \log(\mathrm{var}(\mathbf{y}_k)) = \log(\mathbf{y}_k\mathbf{y}_k^T) = \log(\mathbf{w}_k^T\mathbf{X}\mathbf{X}^T\mathbf{w}_k) \tag{6}$$

Hence, a feature vector $\mathbf{f} = [f_1,\ldots f_k,\ldots f_{ch}]$ produced by all spatial filters can be computed by

$$\mathbf{f} = \log(\mathrm{diag}(\mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W})) \tag{7}$$

where $\mathbf{f} \in R^{ch\times 1}$ and $\mathbf{W}$ is the eigenvector matrix whose columns are eigenvectors from (4).

All CSP-based approaches in BCI select d features from $\mathbf{f}$, $d < ch$. As $\lambda = J(\mathbf{w})$, spatial filters corresponding to larger eigenvalues yield higher variance ratio of class Ω_1 to class Ω_2 and those of smaller eigenvalues yield higher variance ratio of class Ω_2 to class Ω_1. Therefore, all CSP-based approaches in BCI select d features of the largest eigenvalues and of the smallest eigenvalues.

B. Common Spatio-Spectral Patterns

CSSP [4] is an extension of CSP that uses time delay embedding to optimize both filters: spectral and spatial. It doubles the number of signal channels by concatenating the signals $\mathbf{X}$ and their delayed version $\mathbf{X}_\tau^T$ as

$$\hat{\mathbf{X}} = [\mathbf{X}^T\ \ \mathbf{X}_\tau^T]^T. \tag{8}$$

Then, the above-mentioned CSP is applied to the new augmented data matrix $\hat{\mathbf{X}} \in R^{2ch\times T}$ to produce features for classification. This method can be interpreted as a spatial and spectral (time) filter as the output of the CSSP filter of length $2ch$ is a weighted sum of signals of different channels and different time.

C. Regularized CSP

RCSP [15] reduces the sensitivity of CSP to noise and overfitting by adding prior information and hence regularizing it. One of the best RCSP is Tikhonov regularization CSP. This method adds regularization terms to the objective function. Two objective functions with regularization are formulated as follows:

$$J_{r1}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{C}_1\mathbf{w}}{\mathbf{w}^T(\mathbf{C}_2+\delta\mathbf{K})\mathbf{w}} \tag{9}$$

K=I

$$J_{r2}(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{C}_2\mathbf{w}}{\mathbf{w}^T(\mathbf{C}_1+\delta\mathbf{K})\mathbf{w}} \tag{10}$$

where δ is the regularization parameter and $\mathbf{K}$ is predefined matrix. Here, $\mathbf{K} = \mathbf{I}$ (identity matrix) is used as in [15]. Spatial filters are obtained by solving the generalized eigenvalue problems of J_{r1} and J_{r2} and features are produced by the spatial filters in the same way as CSP. RCSP selects features corresponding to the largest eigenvalues of J_{r1} and J_{r2}.

III. Proposed FWR of CSP-Based Features

All CSP-based approaches in BCI select the features produced by the spatial filters. The total number of possible features is the number of EEG channels ch. However, literatures show that the best classification accuracy is achieved at very few number of features, such as 2 to 8 in most cases comparing to, for example, $ch = 60$. The loss of discriminating information by discarding most CSP features may adversely affects the classification accuracy [21]–[24]. To alleviate this problem, we propose a novel FWR method that utilizes all CSP features for classification. As shown in [23] and [24], overfitting problem can be alleviated not only by removing the unreliable features, but also by regularizing the unreliable features.

First, features are sorted in descending order of the eigenvalues of spatial filters obtained from the objective functions (9) and (10). For those obtained from the objective function (3), features are sorted in descending order of $|\lambda - 0.5|$, where λ is the generalized eigenvalues of (4). This is because both the largest and the smallest eigenvalues of (4) correspond to the largest differences of variances of the two classes and $0 \le \lambda \le 1$. All existing CSP-based approaches in BCI select d features and remove the other $ch - d$ features. This is equivalent to weighting the features by a step function that has sharp transition from 1 to 0. We propose to weight the features by a soft function. As the inverse function $1/k$ has long tail and the sigmoid function needs two parameters to control its shift and smoothness, this letter uses a Gaussian function that has short tail to weight the features

$$\gamma_k = e^{-\frac{(k-1)^2}{2\sigma^2}} \tag{11}$$

$$f_k^\gamma = \gamma_k f_k \tag{12}$$

where σ is a parameter, which can be chosen so that the Gaussian function decays the fastest between m and $m+1$. This leads to $\sigma = m - 0.5$, where m is the eigenfeature regularization and extraction (ERE) regularization parameter shown later. The weighted feature vector is $\mathbf{f}_\gamma = [f_1^\gamma,\ldots,f_k^\gamma,\ldots,f_{ch}^\gamma]^T$.

The linear discriminant analysis (LDA) classifier is the most widely used classifier for MI-BCI [3], [4], [15], [16]. It is not difficult to prove that the LDA classifier is identical to the minimum Mahalanobis distance classifier using pooled class-conditional covariance matrix $\mathbf{S}$, also called within-class scatter matrix of the feature vector. The Mahalanobis distance d_j of a feature vector $\mathbf{f}_\gamma$ to the class mean $\boldsymbol{\mu}_j$ of Ω_j is evaluated by

$$d_j = (\mathbf{f}_\gamma - \boldsymbol{\mu}_j)^T \mathbf{S}^{-1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_j). \tag{13}$$

As LDA classifier or the minimum Mahalanobis distance classifier whitens the feature vector by its within-class scatter matrix, feature weighting will not affect the classification results, i.e., feature vectors $\mathbf{f}$ and $\mathbf{f}_\gamma$ will lead to the same classification results. However, feature weighting will greatly affect the classification results if we regularize the classifier, because the regularization techniques modify the feature variance differently based on its value.

As shown in [23] and [25], the Mahalanobis distance can be evaluated in the eigenspace in scalar form as follows:

$$\boxed{d_j = (\mathbf{f}_\gamma - \boldsymbol{\mu}_j)^T \mathbf{S}^{-1} (\mathbf{f}_\gamma - \boldsymbol{\mu}_j) = \sum_{i=1}^{ch} \frac{(z_i - \overline{z}_{ji})^2}{\lambda_i}} \tag{14}$$

where $z_i = \phi_i^T \mathbf{f}_\gamma$ and $\overline{z}_{ji} = \phi_i^T \boldsymbol{\mu}_j$, ϕ_i and λ_i are the ith eigenvector and eigenvalue of $\mathbf{S}$, respectively. If we define a whitened eigenfeature as $z_i/\sqrt{\lambda_i}$, the classifier becomes a simple minimum Euclidian distance classifier. As all CSP features are used in the classification, we need to regularize the classifier, or equivalently, regularize the features to alleviate the problems of instability, overfitting, or poor generalization. There are a few regularization techniques in the literature [23], such as "adding a constant to all eigenvalues" [26], the probabilistic subspace learning and its enhanced version [27], and ERE eigenspectrum model [24]. As ERE has shown good performance in many applications [24], [28]–[30], we combine it with the "adding a constant to all eigenvalues" to strengthen the regularization effect in this letter as follows:

$$\boxed{\tilde{\lambda}_i = \begin{cases} \lambda_i + c, & i < m \\ \frac{\alpha}{i+\beta} + c, & m \le i \le ch \end{cases}} \tag{15}$$

where

$$\alpha = \frac{\lambda_1 \lambda_m (m-1)}{\lambda_1 - \lambda_m} \tag{16}$$

$$\beta = \frac{m\lambda_m - \lambda_1}{\lambda_1 - \lambda_m}. \tag{17}$$

The above-mentioned eigenvalues λ_i of $\mathbf{S}$ are sorted in descending order before the regularization. c and m are two parameters to choose for regularization. This letter chooses $c = \lambda_{ch/2}$ to reduce the free parameter. It is used to ensure a minimum amount of regularization [23], [26] in case ERE fades caused by a large value of m.

The above-mentioned procedure of FWR can be applied in all CSP-based approaches of MI-BCI to replace the conventional feature selection (FS). The proposed FWR approach is summarized in the following.

1) *At the training stage:*
   a) Compute and sort the spatial filters $\mathbf{W}$ by the CSP algorithm or one of its extensions;
   b) Apply all ch spatial filters to EEG signals to extract feature vector $\mathbf{f}$ as (6) or (7);
   c) Compute the weighted feature vector $\mathbf{f}_\gamma$ by the Gaussian function as (11) and (12);
   d) Compute the within-class scatter matrix $\mathbf{S}$ of feature vector $\mathbf{f}_\gamma$ and regularize the descending order sorted eigenvalues λ_i by (15)–(17) to get the regularized eigenvalues $\tilde{\lambda}_i$.
2) *At the recognition stage:*
   a) Apply the steps b) and c) of the training stage to compute the weighted feature vector $\mathbf{f}_\gamma$ for the query trial of EEG signals;
   b) Classify the weighted feature vector $\mathbf{f}_\gamma$ by minimum Mahalanobis distance classifier (14) but using the regularized eigenvalues $\tilde{\lambda}_i$ instead of λ_i.

IV. Experiments

A. EEG Datasets

Two publicly most widely used BCI datasets are used, which are shown as follows.

1) BCI competition IV, Dataset IIa [31] contains EEG data of 22 channels recorded from 9 healthy subjects ($A01$–$A09$). Data contain time point markers of 72 trials per class for each subject for training and testing. The EEG signal was sampled with 250 Hz and later bandpass filtered between 0.5 Hz and 100 Hz (with 50 Hz notch filter enabled).

2) BCI competition III, Dataset IIIa [32] contains EEG data of 60 channels recorded from 3 subjects. Data contain time markers of 90, 60, and 60 trials per class for subjects k3b, k6b, and l1b, respectively. For each subject, a training set and a test set are available. The EEG signal was sampled with 250 Hz and filtered between 1 Hz and 50 Hz with notch filter.

Both datasets contain four motor imagery (MI) actions: left hand, right hand, foot, and tongue. Since CSP-based approaches work on discriminating two classes, we pair the four MIs to construct six binary classification tasks: task 1—left hand versus right hand, task 2—left hand versus foot, task 3—left hand versus tongue, task 4—right hand versus foot, task 5—right hand versus tongue, and task 6—foot versus tongue.

B. Preprocessing and Experiment Settings

At first the raw EEG signals are filtered between 8 and 30 Hz as in [15] with the fifth-order Butterworth filter as preprocessing. Data of each trial are captured from the time segment starting from 0.5 to 2.5 s after the beginning of each trial (as done by the winner of the BCI Competition IV, Dataset IIa). The first 0.5 s is excluded due to reaction time on visual clue. Feature vector extracted from each trial is used as one sample in training or testing. For fair comparison, homogenous setting of filtering and time window is adopted for all methods.

For the FS method, the first d features are selected while the proposed FWR regularizes features controlled by m. Experiments are conducted for values of d and m from 2 to 16. The parameters δ for RCSP, τ for CSSP, and σ for FWR are all selected by five-fold cross validation on training data. As mentioned earlier, σ should be around $m - 0.5$ so its value is

TABLE I
PERFORMANCE COMPARISON OF VARIOUS SUBJECTS ON BCI COMPETITION IV, DATASET IIA AND BCI COMPETITION III, DATASET IIIA

	A01	A02	A03	A04	A05	A06	A07	A08	A09	k3b	k6b	l1b
CSP-FS	92.94	68.06	90.63	81.25	68.87	67.82	**92.36**	88.43	90.86	96.85	76.11	91.39
(d, AUC)	(10, 3.5)	(10, 21.3)	(4, 3.2)	(2, 10.3)	(4, 24.2)	(2, 26.3)	(8, 3.5)	(4, 2.5)	(4, 2.0)	(6, 0.5)	(10, 16.1)	(4, 2.9)
CSP-FWR	**93.75**	**70.6**	**90.97**	**82.18**	**69.1**	**68.29**	91.78	**90.28**	**92.59**	96.85	**76.94**	**91.94**
(m, AUC)	(13, 3.1)	(7, 19.8)	(4, 3.1)	(2, 10.0)	(3, 23.7)	(2, 24.6)	(7, 3.7)	(3, 2.0)	(15, 1.4)	(6, 0.3)	(6, 14.7)	(2, 2.1)
CSSP-FS	95.25	69.44	93.98	84.26	72.34	70.95	92.82	89	90.63	97.41	81.94	92.78
(d, AUC)	(10, 2.6)	(14, 19.2)	(6, 1.3)	(4, 7.7)	(6, 20.1)	(6, 21.8)	(16, 3.0)	(6, 2.1)	(16, 2.3)	(4, 0.3)	(6, 7.9)	(4, 3.2)
CSSP-FWR	**95.49**	**72.11**	93.98	**84.84**	**72.69**	**71.18**	92.82	**91.2**	**91.2**	97.41	**83.89**	**94.44**
(m, AUC)	(7, 2.4)	(2, 19.5)	(3, 1.3)	(2, 7.9)	(3, 20.4)	(5, 20.7)	(16, 2.6)	(3, 2.1)	(2, 2.2)	(5, 0.5)	(4, 8.3)	(3, 2.0)
RCSP-FS	93.52	68.63	**91.67**	80.56	66.09	67.71	92.36	89.7	92.13	95.93	79.72	91.39
(d, AUC)	(10, 3.3)	(10, 23.3)	(6, 2.7)	(2, 10.6)	(4, 25.4)	(8, 25.6)	(10, 3.5)	(4, 2.7)	(2, 2.3)	(6, 0.6)	(12, 10.9)	(4, 3.3)
RCSP-FWR	**93.63**	**69.91**	91.2	**81.37**	**69.21**	**68.52**	92.36	**90.28**	**92.71**	**96.85**	**81.39**	**91.94**
(m, AUC)	(9, 3.0)	(4, 19.0)	(6, 2.8)	(2, 10.4)	(4, 24.0)	(2, 23.9)	(6, 3.8)	(5, 2.3)	(14, 1.6)	(6, 0.3)	(2, 10.6)	(3, 3.0)

TABLE II
PERFORMANCE COMPARISON OF VARIOUS TASKS ON BCI COMPETITION IV, DATASET IIA AND BCI COMPETITION III, DATASET IIIA

	Data set IIa						Data set IIIa					
	task 1	task 2	task 3	task 4	task 5	task 6	task 1	task 2	task 3	task 4	task 5	task 6
CSP-FS	79.71	84.03	85.96	83.64	84.18	74.77	**87.96**	**91.85**	**94.44**	83.89	91.85	86.85
(d, AUC)	(4, 13.8)	(4, 7.4)	(8, 6.8)	(2, 9.1)	(6, 9.5)	(4, 15.8)	(10, 8.4)	(6, 4.3)	(2, 1.2)	(4, 10.1)	(4, 4.2)	(6, 6.7)
CSP-FWR	**80.71**	**85.26**	**87.81**	**84.65**	**85.26**	**75.31**	86.67	90.56	93.89	**84.63**	**92.04**	86.85
(m, AUC)	(3, 12.5)	(2, 6.7)	(13, 6.8)	(3, 8.4)	(11, 8.8)	(3, 15.9)	(6, 8.1)	(5, 4.6)	(2, 1.6)	(2, 9.7)	(7, 4.7)	(2, 4.5)
CSSP-FS	80.48	85.26	87.27	85.8	86.03	77.7	90.93	88.7	**95.56**	88.89	94.07	89.44
(d, AUC)	(8, 12.9)	(10, 5.9)	(6, 6.6)	(6, 7.9)	(4, 7.4)	(6, 12.6)	(12, 6.5)	(6, 4.6)	(4, 1.4)	(6, 2.7)	(10, 2.7)	(2, 4.0)
CSSP-FWR	**81.94**	**87.19**	**87.35**	**86.34**	**86.34**	**79.32**	**91.67**	**90.19**	95	**91.85**	**95**	**90.19**
(m, AUC)	(4, 12.3)	(2, 6.3)	(13, 6.3)	(3, 7.4)	(3, 7.6)	(3, 12.4)	(5, 6.1)	(4, 4.4)	(2, 1.2)	(4, 3.2)	(3, 2.0)	(3, 4.0)
RCSP-FS	78.86	84.1	86.19	**85.19**	**85.03**	74.38	89.07	88.52	95.37	88.89	92.41	87.96
(d, AUC)	(6, 14.9)	(4, 9.4)	(10, 8.3)	(6, 9.2)	(4, 9.1)	(4, 16.9)	(12, 4.7)	(14, 6.4)	(4, 0.7)	(10, 3.4)	(2, 4.5)	(8, 6.6)
RCSP-FWR	**79.01**	**85.03**	**87.11**	84.26	84.72	**76.16**	**89.26**	**88.7**	**95.56**	**89.63**	**93.52**	**89.63**
(m, AUC)	(4, 13.5)	(2, 6.8)	(13, 7.9)	(5, 9.0)	(4, 10.0)	(3, 15.4)	(8, 5.7)	(2, 4.9)	(2, 0.7)	(10, 4.0)	(12, 2.6)	(6, 6.4)

limited between $m - 1.5$ and $m + 0.5$. The minimum Mahalanobis distance classifier is used in all experiments.

Training and testing of 6 methods, CSP, CSSP, and RCSP, using FS and FWR, are conducted for 12 subjects and 6 tasks for each subject. In total, 432 training and testing experiments are performed. In each of the 432 experiments, classification results using parameter values (d and m) are recorded. All these results are summarized into the average accuracy over all tasks of each subject using the same parameter value, and the average accuracy over all subjects for each task using the same parameter value. The best of such accuracies among all tested parameters are recorded in Table I for different subjects and are recorded in Tables II for different tasks.

C. Results and Discussion

For traditional FS methods, Tables I and II show that CSSP outperform CSP in 22 out of 24 cases and RCSP in 15 out of 24 cases, though they are the further developments of CSP. Comparing the two extensions of CSP, CSSP is better than RCSP in 21 out of 24 cases, whereas RCSP is better than CSSP in 2 out of 24 cases. Among the three FS approaches, CSSP performs the best in 20 out of 24 cases, RCSP performs the best in 2 out of 24 cases, and the basic CSP performs the best in 1 out of 24 cases. This shows the difficulty of EEG-based MI-BCI. It is very difficult to develop an approach that can consistently perform better than others in all cases.

To assess the performance of the proposed FWR algorithm, we applied it to the standard CSP and its two typical extensions in two different directions: CSSP utilizing the temporal information and RCSP regularizing the CSP. For the comparison between the conventional FS method and the proposed FWR method, the better one is highlighted in bold in Tables I and II. In total of 72 cases, the proposed FWR method outperforms FS in 58 cases and shows the same performance in 6 cases. It underperforms FS only in 8 out 72 cases. The results on the two datasets clearly suggest that the proposed FWR visibly enhances the performance of CSP-based approaches in MI-BCI. The benefit of applying the proposed FWR method in the CSP-based approaches is evident in Tables I and II. To show the overall classification performance at all different decision thresholds, we also present the average area under the curve (AUC) of the receiver operating characteristics (ROC) beside the d or m value in Tables I and II. Note that AUC given in tables are amplified 100 times to save the space. Tables I and II show that AUCs reflect the relative performances of different methods very similar to those of classification accuracies.

V. CONCLUSION

This letter addresses problems of extracting discriminant features from the EEG signal for MI-BCI based on a set of training samples. To alleviate the problems of overfitting or poor generalization and, hence, boost the classification accuracy, the CSP-based approaches in MI-BCI select some CSP features and remove the others. However, this will result in the loss of discriminative information.

This letter proposes a novel approach that performs FWR to replace the FS in CSP-based approaches. It weights the features extracted by all spatial filters and, then, regularizes the eigenspectrum of the within-scatter matrix of the weighted feature vector for reliable classification. The proposed method can be applied in all CSP-based approaches in MI-BCI.

This letter evaluates the proposed method applying to CSP and its two extensions in two different directions, CSSP and RCSP on EEG data from 12 subjects with 6 tasks performed by each subject from the BCI competition. Experimental results clearly demonstrate that the proposed FWR method yields superior classification accuracy comparing to the existing FS methods.

REFERENCES

[1] J. J. Vidal, "Toward direct brain-computer communication," *Annu. Rev. Biophys. Bioeng.*, vol. 2, pp. 157–180, Jun. 1973.

[2] J. R. Wolpaw and E. W. Wolpaw, *Brain-Computer Interfaces. Principles and Practice.* New York, NY, USA: Oxford Univ. Press, 2012.

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

[4] S. Lemm, B. Blankertz, G. Curio, and K. R. Muller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[5] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 2274–2281, Nov. 2006.

[6] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K.-R. Muller, "Spectrally weighted common spatial pattern algorithm for single trial EEG classification," Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep. METR-2006-40, 2006.

[7] 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.

[8] 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.

[9] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE EMBS Conf. Neural Eng.*, 2007, pp. 204–207.

[10] K. K. Ang, Z. Y. Chin, H. H. Zhang, and C. T. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2008, pp. 2390–2397.

[11] H. Zhang, Z. Y. Chin, K. K. Ang, C. Guan, and C. Wang, "Optimum spatio-spectral filtering network for brain-computer interface," *IEEE Trans. Neural Netw.*, vol. 22, no. 1, pp. 52–63, Jan. 2011.

[12] J. Meng, L. Yao, X. Sheng, D. Zhang, and X. Zhu, "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.

[13] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, "Separable common spatio-spectral patterns for motor imagery BCI systems," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 15–29, Jan. 2016.

[14] 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.

[15] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[16] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 639–653, Mar. 2015.

[17] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing," *J. Neural Eng.*, vol. 9, no. 2, 2012, Art. no. 026013.

[18] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.

[19] X. Jiang and J. Lai, "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1067–1079, May 2015.

[20] J. Lai and X. Jiang, "Class-wise sparse and collaborative patch representation for face recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3261–3272, Jul. 2016.

[21] G. V. Trunk, "Problem of dimensionality: A simple example," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 3, pp. 306–307, Jul. 1979.

[22] A. Jain, R. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[23] X. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 16–26, Mar. 2011.

[24] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.

[25] X. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

[26] J. Friedman, "Regularized discriminant analysis," *J. Amer. Stat. Assoc.*, vol. 84, pp. 165–175, Mar. 1989.

[27] X. Jiang, B. Mandal, and A. Kot, "Enhanced maximum likelihood face recognition," *Electron. Lett.*, vol. 42, pp. 1089–1090, Sep. 2006.

[28] J. Choi, Y. Ro, and K. Plataniotis, "Color local texture features for color face recognition," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1366–1380, Mar. 2012.

[29] J. Choi, W. Neve, K. Plataniotis, and Y. Ro, "Collaborative face recognition for improved face annotation in personal photo collections shared on online social networks," *IEEE Trans. Multimedia*, vol. 13, no. 1, pp. 14–28, Feb. 2011.

[30] Z. Lu, X. Jiang, and A. Kot, "A color channel fusion approach for face recognition," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1839–1843, Nov. 2015.

[31] "BCI competition 4 data set IIa," 2004. [Online]. Available: http://bbci.de/competition/iv/desc_2a.pdf

[32] "BCI competition 3 data set IIIa," 2008. [Online]. Available: http://www.bbci.de/competition/iii/desc_IIIa.pdf