SEPARABLE COMMON SPATIO-SPECTRAL PATTERN ALGORITHM FOR CLASSIFICATION OF EEG SIGNALS

Amirhossein S. Aghaei, Mohammad Shahin Mahanta, and Konstantinos N. Plataniotis

E-mail: {aghaei, mahanta, kostas} @ comm.utoronto.ca
Edward S. Rogers Sr. Department of Electrical and Computer Engineering,
University of Toronto, 10 King's College Road, Toronto, Canada.

ABSTRACT

This paper proposes a novel method for extraction of discriminant spatio-spectral EEG features in motor imagery brain-computer interfaces. Considering a heteroscedastic binary classification setup, this method extracts the spatio-spectral features whose variance is maximized for one brain task and minimized for the other task. Therefore, our method can be considered as a spatio-spectral generalization of the conventional common spatial patterns (CSP) algorithm. In comparison to the similar solutions in the literature, such as filter-bank CSP (FBCSP) method, the proposed method benefits from joint processing of both spatial and spectral features, which improves the overall performance of the BCI while reducing its computational cost. Furthermore, our algorithm provides a simple measure that allows for ranking the discriminant power of extracted spatio-spectral features, which is not possible in FBCSP method. The experimental results demonstrate that the proposed method outperforms FBCSP for both raw EEG and preprocessed EEG data.

***Index Terms*—** brain computer interface, common spatial patterns, spatio-spectral features, matrix-variate Gaussian

1. INTRODUCTION

Electroencephalogram (EEG) signals are widely used in non-invasive brain-computer interfaces (BCIs) to provides a communication channel between the brain and external world. BCIs can be used in various applications such as artificial limbs, speech synthesizers, and navigation in virtual environments. This paper focuses on BCI systems that are based on decoding EEG signals recorded during motor-imagery (MI) tasks. However, our proposed method can be deployed in other BCI systems that utilize multichannel EEG signals.

During motor-imagery tasks, EEG signals exhibit task-specific features in both spatial domain and spectral (or frequency) domain [1–4]. In the literature, various algorithms have been proposed to extract these discriminant features through spatial and spectral processing of the data. One of the most powerful algorithms is the common spatial patterns (CSP) method [2, 5, 6]. CSP is originally designed for binary classification of brain tasks and extracts the spatial features that exhibit maximum variance for one task while having minimum variance for the other task. Then, the variances of these spatial features are used as a new discriminant feature set which can be subsequently passed to a classifier.

A main shortcoming of CSP is that it ignores the spectral characteristics of the EEG signal. To alleviate this problem, several variants of CSP have been proposed in the literature [7–12]. One of the promising solutions is the work in [12] which is illustrated in Fig. 1(a). In this approach, different EEG rhythms are obtained by means of bandpass filtering the EEG signal, and then a bank of CSP modules is deployed to separately extract spatial features from each EEG rhythm; hence the name filter-bank CSP (FBCSP). The resulting features are then used for classification of the EEG data. Despite its high performance, FBCSP suffers from high computational cost since it requires a separate feature extractor for each spectral band. Moreover, since in FBCSP each spectral band is treated independently, possible correlations between different EEG rhythms are completely ignored in the feature extraction stage, which in turn can cause redundancy in the extracted features. Finally, FBCSP does not provide any measure for comparing discriminant power of the features obtained from different spectral bands. As a consequence, [12] suggests to use a feature selection module to reduce the dimensionality of the feature space prior to classification.

This paper proposes a novel algorithm for joint extraction of spatial and spectral features from different EEG rhythms. The proposed method, called *separable common spatio-spectral patterns (SCSSP)*, is based on a matrix-variate Gaussian model for spatio-spectral EEG patterns which allows us to develop a bilinear feature extractor. Compared to the FBCSP method, our algorithm has the following main advantages: First, it involves only two CSP-type modules, regardless of the number of frequency bands (N_f). As a result, the computational cost of training SCSSP algorithm in a practical BCI is significantly less than 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 extrac-

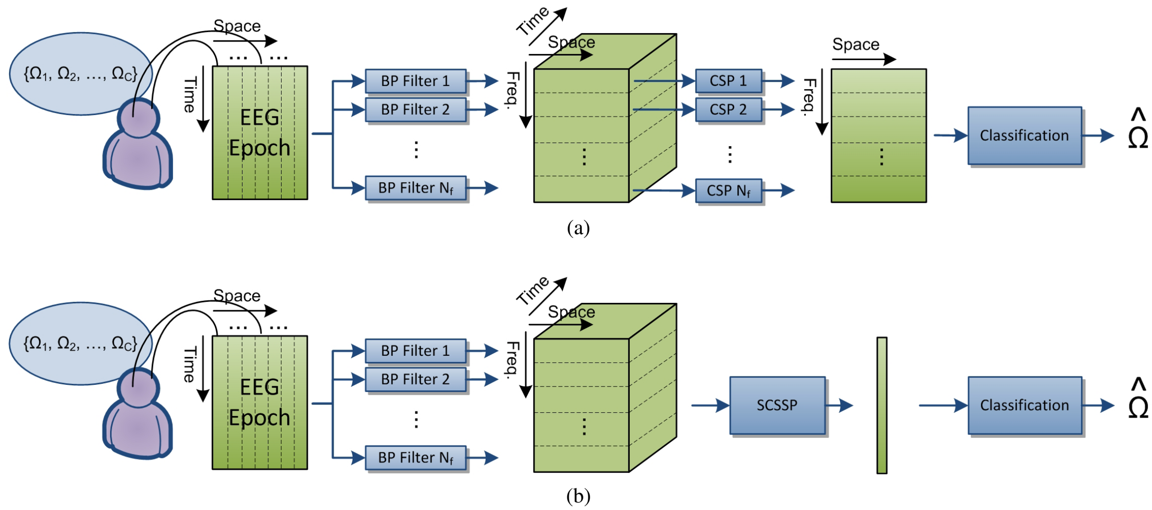

Fig. 1. Spatio-Spectral feature extraction schemes: (a) Filter-bank common spatial pattern (FBCSP), (b) Separable common spatio-spectral pattern (SCSSP)

tion. Third, a measure is provided to rank the discriminatory power of extracted spatio-spectral features, which eliminates the need for a subsequent feature selection stage.

2. SYSTEM MODEL

Fig. 1(b) illustrates the processing pipeline of our proposed algorithm. Consider an EEG epoch with N_t samples from N_{ch} channels (or electrodes)[1]. After passing the EEG epoch through a set of N_f bandpass filters, we get N_t matrices of size $N_f \times N_{ch}$, each of which representing a spatio-spectral EEG pattern. The ultimate goal is to extract the most discriminant features from these matrix-variate patterns.

2.1. Matrix-Variate Gaussian Model for Spatio-Spectral EEG patterns

Let $\mathbf{X} \in \mathbb{R}^{N_f \times N_{ch}}$ denote the matrix-variate EEG pattern. Each motor-imagery task, denoted by class Ω_i, is characterized by the likelihood density $f(\mathbf{X}|\Omega_i)$. We adopt a matrix-variate Gaussian model [13] for these likelihoods:

$$f(\mathbf{X}|\Omega_i) = \mathcal{N}(\mathbf{M}_i, \mathbf{\Phi}_i, \mathbf{\Psi}_i) \quad \text{for} \quad 1 \le i \le C, \tag{1}$$

where $\mathbf{M}_i = \mathrm{E}_{\mathbf{X}|\Omega_i}(\mathbf{X})$, and

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

$$\mathbf{\Psi}_i = \mathrm{tr}^{-1}(\mathbf{\Phi}_i) * \mathrm{E}_{\mathbf{X}|\Omega_i}((\mathbf{X}-\mathbf{M}_i)^T(\mathbf{X}-\mathbf{M}_i)), \tag{3}$$

Here, $\mathbf{M}_i$ 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 $1 \le i \le C$. Therefore, the discriminant information are contained in the second order statistics of the data.

Let $\mathbf{x} = \mathrm{vec}(\mathbf{X})$ be the vectorized representation of $\mathbf{X}$ obtained from concatenation of its columns, and denote the covariance matrix of $\mathbf{x}$ under Ω_i by $\mathbf{\Sigma}_i$. Based on the matrix-variate Gaussian model in (1), $\mathbf{\Sigma}_i$ can be expressed in terms of the spatial and spectral covariances: $\mathbf{\Sigma}_i = \mathbf{\Psi}_i \otimes \mathbf{\Phi}_i$. Moreover, any bilinear transformation of the form $y = \mathbf{a}^T\mathbf{X}\mathbf{b}$ is equivalent to a linear transformation on $\mathbf{x}$ as follows: $y = \mathrm{vec}(\mathbf{b} \otimes \mathbf{a})^T\mathbf{x}$. These two *separability* properties are used in the next section to derive the SCSSP algorithm.

2.2. Separable Common Spatio-Spectral Patterns Method

Consider a binary classification problem (i.e., $C = 2$). Following the general approach of CSP algorithm in [2, 5], we look for linear transformations $\mathbf{w}_k$ that provide uncorrelated spatio-spectral features $y_k = \mathbf{w}_k^T\mathbf{x}$, whose variance is maximum in one class and minimum in the other class. Similar to CSP method, $\mathbf{w}_k$ are the eigenvectors obtained from the following generalized eigenvalue problem:

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

The next theorem provides the solution for (4).

Theorem 1. *Let* $\mathbf{x} = \mathrm{vec}(\mathbf{X})$*, 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:*

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

[1]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 $\mathrm{tr}(\mathbf{A})$. Also, the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$.

where $\mathbf{\Lambda}_R$, $\mathbf{W}_R$, $\mathbf{\Lambda}_L$ and $\mathbf{W}_L$ are 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. Proof by substitution. □

Let λ^k, $1 \leq k \leq N_f N_{ch}$, denote the diagonal entries of $\mathbf{\Lambda}$ sorted in descending order. Theorem 1 implies that

$$\lambda^k = \frac{\lambda_L^{i[k]} \lambda_R^{j[k]}}{\lambda_L^{i[k]} \lambda_R^{j[k]} + (1 - \lambda_L^{i[k]})(1 - \lambda_R^{j[k]})} \tag{7}$$

where $\lambda_L^{i[k]}$ and $\lambda_R^{j[k]}$ are the corresponding eigenvalues in $\mathbf{\Lambda}_L$ and $\mathbf{\Lambda}_R$, with $1 \leq i[k] \leq N_f$ and $1 \leq j[k] \leq N_{ch}$. Also, the eigenvectors corresponding to λ^k are expressed as $\mathbf{w}_k = \mathbf{w}_{R,j[k]} \otimes \mathbf{w}_{L,i[k]}$, where $\mathbf{w}_{R,j[k]}$ and $\mathbf{w}_{L,i[k]}$ are the eigenvectors in $\mathbf{W}_R$ and $\mathbf{W}_L$ corresponding to $\lambda_R^{j[k]}$ and $\lambda_L^{i[k]}$.

Note that for $k = 1$ and $k = N_f N_{ch}$ 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. Using these results, together with the separability property explained in Section 2.1, we propose the following algorithm for extraction of the "d" most discriminant spatio-spectral features:

1. Solve the generalized eigenvalue problems in (5) and (6) for spatial and spectral covariance matrices.

2. Using (7), calculate the eigenvalues λ^k and sort them in descending order to determine the corresponding indices $i[k]$ and $j[k]$.

3. Extract the d most discriminant features by calculating

$$y_k = \mathbf{w}_{L,i[k]}^T \mathbf{X} \mathbf{w}_{R,j[k]} \quad \text{for} \quad k \in \mathfrak{K} \tag{8}$$

$$\mathfrak{K} = \{1, N_f N_{ch}, 2, (N_f N_{ch} - 1), \cdots, \frac{d}{2}, (N_f N_{ch} - \frac{d}{2} + 1)\}.$$

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

4. Calculate the normalized log-power features

$$z_k = \log\left(\frac{\operatorname{var}(y_k)}{\Sigma_{k \in \mathfrak{K}} \operatorname{var}(y_k)}\right) \tag{9}$$

where $\operatorname{var}(y_k)$ function calculates the variance or power of y_k over N_t samples.

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

It is worth mentioning that λ^k ranges between zero and one, and its value provides a measure for discriminant power of feature y_k. Similar to the conventional CSP method, values close to zero or one correspond to high discriminant features, whereas values close to $\frac{1}{2}$ correspond to low discriminant features. 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 this paper, we consider two possible choices for classifier: (a) Naive Bayes classifier, (b) linear classifier.

Multiclass Extension: The algorithm proposed in this section is derived for a binary classification problem. In case that $C > 2$, various different binary-to-multiclass extension approaches can be used, such as one-versus-rest (OVR), pairwise, and divide-and-conqure methods [14]. Comparison of these multiclass extension techniques is outside the scope of this paper and we only consider the OVR approach. In the OVR method, the SCSSP module extracts d spatio-spectral features for discrimination of each class from the rest of classes. Therefore, a total number of $d * C$ features will be passed to the classifier.

Parameter Estimation: In (5) and (6), we need the spatial and spectral covariances, which can be obtained from the following moment estimators [15]:

$$\widehat{\mathbf{\Phi}}_i = \frac{1}{N_{ch} N_i} \sum_{n=1}^{N_i} \mathbf{X}_n \mathbf{X}_n^T, \quad \widehat{\mathbf{\Psi}}_i = \frac{1}{N_f N_i} \sum_{n=1}^{N_i} \mathbf{X}_n^T \mathbf{X}_n$$

where N_i is the number of training samples $\mathbf{X}_n, 1 \leq n \leq N_i$, available for each class Ω_i.

3. EXPERIMENTAL RESULTS

In this section the performance of proposed SCSSP method will be compared against the FBCSP method, using data set V from BCI competition III [16]. To provide a fair comparison, we use both Naive Bayes (NB) classifier and linear classifier for each method. Here, the performance of the overall BCI system is measured by correct classification rate (CCR). This data set contains EEG signals recorded from three subjects in four sessions. The first three sessions are used for training purposes, and the fourth one is used for competition, i.e. testing phase. The goal is to classify the imagined tasks every 0.5 second, using the last second of data. The signals are collected using 32-electrode Biosemi system at 512Hz sampling rate. Each session consists of sequential 15-second trials of three tasks: left-hand movement, right-hand movement, and generation of words beginning with a random letter.

The database providers have suggested to perform surface Laplacian (SL) spatial filtering on EEG and then select the 8-centro-parietal channels to reduce the dimensionality of the data. As a result, four versions of preprocessed data will studied in this section for comparative purposes: (a) 32-channel raw EEG, (b) 8-centro-parietal channels from raw EEG, (c) 32-channel SL-filtered EEG, (d) 8-centro-parietal channels from SL-filtered EEG. In order to perform bandpass filtering of the signals, we use Chebyshev Type II filters of order 6 and bandwidth of 4 Hz. A total of $N_f = 6$ filters are used to cover the α and β rhythms ($8 - 32$ Hz).

Table 1. Correct classification rate (CCR) results in the cross-validation phase.

EEG Data	FE Method	Classifier	Subject a		Subject b		Subject c		Average
			%CCR	d	%CCR	d	%CCR	d	%CCR
Raw EEG (32 Channel)	FBCSP	NB	58.22 ± 7.26	108 (=6*18)	48.10 ± 5.86	48 (=6*8)	45.62 ± 0.80	60 (=6*10)	50.65 ± 6.67
		Linear	63.25 ± 3.46	72 (=6*12)	50.42 ± 5.85	72 (=6*12)	48.73 ± 1.73	60 (=6*10)	54.13 ± 7.94
	SCSSP	NB	56.88 ± 13.06	100	48.66 ± 3.23	168	47.60 ± 2.63	48	51.05 ± 5.08
		Linear	**72.17 ± 5.97**	**34**	**59.85 ± 5.65**	**46**	**49.72 ± 2.34**	**40**	**60.58 ± 11.24**
Raw EEG (8 Channel)	FBCSP	NB	60.64 ± 4.45	36 (=6*6)	46.62 ± 5.66	48 (=6*8)	45.83 ± 4.01	36 (=6*6)	50.97 ± 8.22
		Linear	65.99 ± 3.96	36 (=6*6)	49.79 ± 1.84	48 (=6*8)	43.50 ± 3.74	48 (=6*8)	53.10 ± 11.60
	SCSSP	NB	61.21 ± 9.44	48	48.59 ± 11.53	48	45.34 ± 2.02	46	51.71 ± 8.38
		Linear	**70.55 ± 1.11**	**38**	**50.56 ± 2.96**	**30**	**46.89 ± 5.93**	**14**	**56.00 ± 12.73**
SL-filtered EEG (32 Channel)	FBCSP	NB	58.71 ± 6.96	108 (=6*18)	48.10 ± 4.96	84 (=6*14)	46.40 ± 2.97	96 (=6*16)	51.07 ± 6.67
		Linear	59.26 ± 5.25	96 (=6*16)	52.60 ± 8.61	36 (=6*6)	48.45 ± 1.92	72 (=6*12)	53.43 ± 5.45
	SCSSP	NB	59.06 ± 12.90	174	52.25 ± 3.72	44	49.65 ± 4.57	106	53.65 ± 4.86
		Linear	**70.41 ± 5.86**	**38**	**61.18 ± 8.50**	**40**	**48.73 ± 0.76**	**52**	**60.11 ± 10.88**
SL-filtered EEG (8 Channel)	FBCSP	NB	61.99 ± 3.30	48 (=6*8)	44.66 ± 8.90	48 (=6*8)	47.53 ± 3.82	48 (=6*8)	51.39 ± 9.29
		Linear	65.14 ± 2.65	48 (=6*8)	55.20 ± 4.11	48 (=6*8)	48.52 ± 2.02	36 (=6*6)	56.29 ± 8.37
	SCSSP	NB	59.32 ± 5.76	44	45.64 ± 8.55	48	49.93 ± 3.60	38	51.63 ± 7.00
		Linear	**67.39 ± 1.95**	**46**	**58.23 ± 6.55**	**26**	**52.82 ± 2.32**	**46**	**59.48 ± 7.37**

Table 2. Correct classification rates (CCR) in testing phase.

EEG Data	FE Method	Classifier	%CCR			
			Subj. a	Subj. b	Subj. c	Avg.
Raw EEG (32 Channel)	FBCSP	NB	68.09	54.85	39.29	54.07
		Linear	71.06	62.66	48.32	60.68
	SCSSP	NB	68.51	54.64	32.98	52.05
		Linear	**74.89**	**71.10**	**42.02**	**62.67**
Raw EEG (8 Channel)	FBCSP	NB	64.04	55.91	49.37	56.44
		Linear	74.04	53.38	48.74	58.72
	SCSSP	NB	65.32	53.80	47.48	55.53
		Linear	**73.40**	**52.95**	**47.06**	**57.81**
SL-filtered EEG (32 Channel)	FBCSP	NB	66.17	60.55	40.34	55.68
		Linear	66.60	50.00	48.95	55.18
	SCSSP	NB	69.36	54.64	39.08	54.36
		Linear	**75.32**	**73.42**	**45.59**	**64.78**
SL-filtered EEG (8 Channel)	FBCSP	NB	69.79	50.42	45.59	55.27
		Linear	72.77	60.97	48.53	60.76
	SCSSP	NB	65.53	52.53	46.01	54.69
		Linear	**71.91**	**62.03**	**49.37**	**61.10**

A three-fold cross-validation on the training data (i.e., the first three session) is used to determine the optimal value of d for each method. Tab. 1 provides the optimal dimension for different subjects in each method and their corresponding CCR ± its standard deviation. These results reveal that in general, the SCSSP method with linear classifier requires minimum number of features. This can be attributed to the fact that SCSSP jointly processes the spatial and spectral features of the data and takes into account possible correlations between different frequency bands. Moreover, note that the number of features in FBCSP method is constrained to $N_f * d_{csp}$ due to its inability to compare discriminate power of features from different frequency bands. Finally, it can be seen that during the validation phase, SCSSP outperforms the FBCSP method regardless of the classifier type.

Tab. 2 outlines the CCR results for the testing phase. Here, we have used the optimal values of d from Tab. 1. The winning algorithm in this competition achieves a performance of 62.72% without post processing [17]. In comparison to both FBCSP and the method in [17], the proposed SCSSP algorithm provides competitive performance, specially when all the 32-channels are used and hence dimensionality of the original data is high. These results also cofirm that the matrix-variate Gaussian distribution is a reasonable assumption for spatio-spectral EEG patterns.

4. CONCLUSIONS AND REMARKS

A new approach for spatio-spectral feature extraction from EEG signals was proposed in this paper. In order to derive this feature extractor, we adopt a matrix-variate Gaussian model for the spatio-spectral EEG patterns, which allows for a separable structure for the covariance of the data; hence the name separable common spatio-spectral patterns (SCSSP). Using experimental results on data set V from BCI competition III, it was shown that the proposed SCSSP method outperforms the existing FBCSP method, while requiring a relatively lower number of features. These experimental results indirectly validate the matrix-variate Gaussian model for the spatio-spectral patterns which are obtained from *bandpass filtering* the EEG signal. It is noteworthy that the authors' recent study on linear discriminant analysis in [18] has shown that the matrix-variate Gaussian model can also be used for characterization of spatio-spectral patterns which are obtained from applying *short-time Fourier transformation* to the EEG data. Those results in conjunction with the results of the current paper suggest that the covariance between any two rhythmic activities in two different EEG channels can be decomposed into two multiplicative components: (a) A spectral covariance term that only depends on the frequency of these two rhythms, and (b) A spatial component that only depends on the spatial location of these two EEG channels. In order to investigate the validity of this conjecture, further statistical studies on different EEG databases is required.

5. REFERENCES

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

[2] H. Ramoser, J. Mller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[3] G. Pfurtscheller and C. Neuper, "Motor imagery direct communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[4] K. J. Miller, E. C. Leuthardt, G. Schalk, R. P. N. Rao, N. R. Anderson, D. W. Moran, J. W. Miller, and J. G. Ojemann, "Spectral changes in cortical surface potentials during motor movement," *Journal of Neuroscience*, vol. 27, no. 9, pp. 2424–2432, 2007.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press Professional Inc., San Diego, CA, USA, second edition, 1990.

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

[7] S. Lemm, B. Blankertz, G. Curio, and K. . Mller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.

[8] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. . Mller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[9] W. Wu, X. Gao, B. Hong, and S. Gao, "Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 6, pp. 1733–1743, 2008.

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

[11] H. . Suk and S. . Lee, "A probabilistic approach to spatio-spectral filters optimization in brain-computer interface," in *2011 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2011*, October 2011, pp. 19–24.

[12] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in Neuroscience*, , no. MAR, 2012.

[13] A.K. Gupta and D.K. Nagar, *Matrix variate distributions*, Chapman & Hall/CRC, 1999.

[14] Peter E. Hart Richard O. Duda and David G. Stork, *Pattern Classification*, John Wiley, New York, second edition, 2001.

[15] J. Zhao, P.L.H. Yu, L. Shi, and S. Li, "Separable linear discriminant analysis," *Computational Statistics and Data Analysis*, vol. 56, no. 12, pp. 4290–4300, 2012.

[16] Jdel.R. Millan, "On the need for on-line learning in brain-computer interfaces," in *Proc. Int. Joint Conf. on Neural Networks*, July 2004, vol. 4, pp. 2877–2882 vol.4.

[17] F. Galán, F. Oliva, and J. Guàrdia, "Using mental tasks transitions detection to improve spontaneous mental activity classification," *Med. Bio. Eng. Comput.*, vol. 45, no. 6, pp. 603–609, 2007.

[18] M. S. Mahanta, A. S. Aghaei, and K. N. Plataniotis, "A bayes optimal matrix-variate LDA for extraction of spatio-spectral features from EEG signals," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 3955 –3958.