ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-1080

Higher-Order Correlation Coefficient Analysis for EEG-Based Brain-Computer Interface

Ye Liu [1] and **Qibin Zhao** [2] and **Min Chen** [3] and **Liqing Zhang**[1]

Abstract. Electroencephalogram (EEG) based brain-computer interface (BCI) has been proved to be an effective communication way between human brain and external devices. In order to effectively recover the cortical dynamics from the EEG signals and improve the classification performance, plenty of studies focused on constructing subject-specific spatial and spectral filters, achieving considerable improvement in classification accuracy. However, almost all the approaches aimed to find one common subspace for projection of all the samples in different classes. Studies have shown that active channels and frequency information were not only subject-dependent but also class-dependent. Thus the variety of class-dependent spatial and spectral characteristics can provide further discriminative information for classification. In this paper, we proposed a tensor-based method which attempted to seek individual spatial and spectral subspaces for each class by which each class was projected into its own subspace separately such that they were easily to be classified. Finally, we added a regularization term in this model to avoid overfitting. We evaluated the effectiveness and robustness of the proposed method on two different datasets including one widely-used benchmark EEG dataset collected from healthy subjects and one self-collected EEG dataset collected from stroke patients. The results demonstrated its superior performance.

1 Introduction

Brain-computer interface (BCI) provides a communication system between human brain and external devices, and hence provides a communication channel for people with severe motor disabilities to reestablish environmental control abilities [35]. Among assorts of brain diffused signals, electroencephalogram (EEG) is the most exploited sensory signal to be studied in BCI researches, due to its low cost and high time resolution compared to other modalities. A number of EEG-based BCI systems have been developed recently [35, 6, 25] in which patterns of EEG in different mental states can be discriminated for information transmission by feature extraction and classification algorithms. Research [9, 22] has shown that their effectiveness and efficiency depend on the quality of EEG feature representation and the accuracy of pattern classification of the recorded single trial EEG.

One of the most successful algorithms for EEG classification, evidenced by the BCI Competition [8], is termed the Common Spatial Pattern (CSP) [28]. CSP is used for discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class. Although CSP proves to be highly successful in BCI, the performance of the spatial filters constructed by CSP algorithm heavily depends on their operational frequency bands and channels configuration [3, 31, 21]. Setting a broad frequency range and channels or manually selecting a subject-specific frequency range and channels are commonly used with the CSP algorithm when applied on real applications.

To address the problem of manually selecting the operational subject-specific frequency band and channels group, several approaches have been proposed. Arvaneh *et al.* proposed a sparse CSP-based channel selection method by sparsifying the CSP projection matrix [4]. Lemm *et al.* proposed a Common Spatio-Spectral Pattern (CSSP) [18] method to optimize a simple filter that employed a one time delayed sample with the CSP algorithm, while Dornhege introduced a Common Sparse Spectral Spatial Pattern (CSSSP) [14] method to improve the CSSP algorithm by performing simultaneous optimization of an arbitrary (Finite Impulse Response) FIR filter within the CSP algorithm. Differently, Wu *et al.* proposed an iterative learning method, called Iterative Spatio-Spectral Patterns Learning (ISSPL) [36], in which an FIR filter and a classifier were simultaneously parameterized and optimized in the spectral domain, alternately with optimization of spatial filters using CSP in the spectral domain. More recently, Filter Bank Common Spatial Pattern (FBCSP) [2], whose efficacy was demonstrated in the latest BCI Competition IV [26], combined a filter bank framework with CSP and selected the most discriminative features using a mutual information based criterion [2].

However, learning optimum spatial-spectral filters is still a challenging and open issue in BCI, especially in some paradigms where motor imagery characteristics are not available (e.g., active channels and frequency information of stroke patients are not well known). Moreover, the previous methods aimed to find *common* projection matrices for all the classes so that all the samples were projected into a lower dimensional space. Despite various studies and recent advances, none of them attempted to learn the optimal individual spatial-spectral filters for each class. e.g., CSP-based methods only learned the individual spatial filters for each class, ignoring finding class-dependent spectral filters. Studies have shown that both active channels and frequency information are not only subject-dependent but also class-dependent [38, 13, 39]. Hence, the variety of class-dependent spatial and spectral characteristics can provide further discriminative information. Theoretically speaking, the classification performance will be improved if we try to seek a set of individual spatial and spectral subspaces for each class and project the samples in each class into its own subspace rather than a common subspace.

[1] Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China. Corresponding author email: zhang-lq@cs.sjtu.edu.cn
[2] Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, Japan.
[3] Bank of Communications Pacific Credit Card Center, Shanghai, China.

To address the problems mentioned above, in this paper, a tensor-based method, called Higher-order Correlation Coefficient Analysis (HOCCA), is proposed with the aim of simultaneously seeking *individual* spatial and spectral subspaces for each class so that each class is projected into its own subspace separately. Many real-world data is formed in a multidimensional way rather than a vector form. i.e., EEG can be represented by high order tensors with multiple-modality patterns in the spatial-spectral-temporal domain. Tensor provides an effective and faithful representation of the structural properties of data, in particular, when multidimensional data or a data ensemble affected by multiple factors are involved. Tensor representation has attracted growing attention in multidimensional data analysis, such as functional magnetic resonance (fMRI), electrocorticography (ECoG) and EEG [37, 16].

In brief, the main contributions in this paper can be summarized as follows:

- To the best of our knowledge, this is the first study for EEG classification which attempts to seek a set of individual multilinear subspaces for each class rather than a common subspace for all the classes, by which the samples belonging to different classes are projected into their own subspace separately such that they are easily to be classified.
- The proposed method extends the vector-based subspace learning method to a tensor variate input space in order to consider the multiway structure of inputs into the model learning and predictions, which is effective and robust for multidimensional EEG data classification when lacking of the prior knowledge like channels configuration and active frequency bands.
- The discriminative spatial and spectral patterns can be simultaneously identified from limited training dataset, which may provide insights to the underlying cortical activity patterns, e.g. for brain source localization and physiological knowledge exploration.

In the rest part of the paper, Section 2 introduces the notations and basic multilinear algebra operations. Details of HOCCA is given in Section 3, followed by the experimental configuration in Section 4. Section 5 evaluates the effectiveness of HOCCA when applied on one widely-used benchmark EEG dataset and one self-collected EEG dataset. Finally, we give a brief discussion and conclusion about our work in Section 6 and Section 7, respectively.

2 Preliminary

2.1 Notations and tensor algebra

Many real-world data is formed in a multidimensional way rather than a vector form. Tensor representation has attracted growing attention in multidimensional data analysis, such as functional magnetic resonance (fMRI), electrocorticography (ECoG) and EEG [16, 20, 19].

Nth-order tensors (multi-way arrays) are denoted by underlined boldface capital letters, matrices (two-way arrays) by boldface capital letters, and vectors by boldface lower-case letters, e.g., $\underline{\mathbf{X}}$, $\mathbf{P}$ and $\mathbf{t}$ are examples of a tensor, a matrix and a vector, respectively. The nth-mode ***matricization*** of a tensor $\underline{\mathbf{X}}$ is denoted by $\underline{\mathbf{X}}_{(n)}$.

Definition 1: (*Inner product*) The inner product of two same-sized tensors $\underline{\mathbf{X}}$,$\underline{\mathbf{Y}}$ is defined by:

$$\langle\underline{\mathbf{X}},\underline{\mathbf{Y}}\rangle = \sum_{i_1 i_2\cdots i_M} x_{i_1 i_2\cdots i_M} y_{i_1 i_2\cdots i_M} \tag{1}$$

The Frobenius norm by $\|\underline{\mathbf{X}}\|_F = \sqrt{\langle\underline{\mathbf{X}},\underline{\mathbf{X}}\rangle}$.

Definition 2: (*Outer product*) The outer product of the tensors $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ is given by

$$(\underline{\mathbf{X}}\circ\underline{\mathbf{Y}})_{i_1 i_2\ldots i_M j_1 j_2\ldots j_N} = x_{i_1 i_2\ldots i_M} y_{j_1 j_2\ldots j_N} \tag{2}$$

Definition 3: (*Mode-n product*) The *mode-n product* of a tensor $\underline{\mathbf{X}} \in R^{I_1\times I_2\times\cdots\times I_N}$ and matrix $\mathbf{A}\in R^{J_n\times I_n}$ is denoted by $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\times_n \mathbf{A} \in R^{I_1\times I_2\times\cdots\times I_{n-1}\times J_n\times I_{n+1}\times\cdots\times I_N}$ and is defined as:

$$y_{i_1\cdots i_{n-1} j_n i_{n+1}\cdots i_N} = \sum_{i_n} x_{i_1\ldots i_N}\mathbf{a}_{j_n i_n}$$

The *mode-n product* of a tensor $\underline{\mathbf{X}} \in R^{I_1\times I_2\times\cdots\times I_N}$ and multiple matrices $\{\mathbf{A}^{(n)} \in R^{J_n\times I_n}, n = 1,\ldots,N\}$ is denoted by $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\prod_{n=1}^{N}\times_n\mathbf{A}^{(n)} \in R^{J_1\times J_2\times\cdots\times J_N}$. Especially, the product of $\underline{\mathbf{X}}$ and multiple matrices $\{\mathbf{A}^{(n)}\}_{n=1}^N$ except the k-th one is denoted as

$$\underline{\mathbf{X}}^{(\bar{k})} = \underline{\mathbf{X}}\prod_{n=1,n\neq k}^{N}\times_n\mathbf{A}^{(n)} \in R^{J_1\times\cdots\times J_{k-1}\times I_k\times J_{k+1}\cdots\times J_N}$$

Definition 4: (*Tensor Contraction*) The contraction of a tensor is obtained by equating two indices and summing over all values of the repeated indices. Contraction reduces the tensor order by 2. Given two tensors $\underline{\mathbf{X}} \in R^{I_1\times I_2\times\cdots\times I_M\times J_1\times J_2\times\cdots\times J_N}$ and $\underline{\mathbf{Y}} \in R^{I_1\times I_2\times\cdots\times I_M\times K_1\times K_2\times\cdots\times K_P}$, the contraction on the tensor product $\underline{\mathbf{X}}\otimes\underline{\mathbf{Y}}$ along the first M modes is

$$[\![\underline{\mathbf{X}}\otimes\underline{\mathbf{Y}};(1:M)(1:M)]\!] = \sum_{i_1}^{I_1}\cdots\sum_{i_M}^{I_M} x_{i_1\ldots i_M j_1\ldots j_N} y_{i_1\ldots i_M k_1\ldots k_P}$$

Especially, contracted product of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ on all indices except the k-th index is denoted as $[\![\underline{\mathbf{X}}\otimes\underline{\mathbf{Y}};(\bar{k})(\bar{k})]\!]$.

3 Higher-order Correlation Coefficient Analysis

HOCCA is proposed with the aim of finding multilinear subspaces for each class so that high-dimensional samples belonging to different classes are projected into their individual subspaces where the mutual correlation between them is minimized. We first formulate the HOCCA model and then adopt an alternating solution to solve this problem in iterations.

3.1 Problem Formulation

(1) Vector-based Correlation Coefficient Analysis

For each subject, denote $\mathbf{X}_t \in R^{I_1\times M}$ and $\mathbf{Y}_t \in R^{J_1\times M}$ as EEG training samples of left and right motor imagination at time t, respectively, where M is the number of the training samples and I_1 (J_1) is number of the channels. Motivated by the definition of correlation coefficient [7], we aim to seek two projection vectors, $\mathbf{u}\in R^{I_1}$ and $\mathbf{v}\in R^{J_1}$, such that the following correlation coefficient constructed by $\mathbf{X}_t$ and $\mathbf{Y}_t$:

$$\rho = \frac{\mathbf{u}^T\mathbf{X}_t\mathbf{Y}_t^T\mathbf{v}}{\sqrt{(\mathbf{u}^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{u})(\mathbf{v}^T\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{v})}} \tag{3}$$

is minimized.

Since ρ is invariant to the scaling of $\mathbf{u}$ and $\mathbf{v}$, minimizing ρ can be formulated equivalently as

$$\begin{aligned}&\min_{\mathbf{u},\mathbf{v}} \ \mathbf{u}^T\mathbf{X}_t\mathbf{Y}_t^T\mathbf{v}\\ &s.\ t.\ \ \mathbf{u}^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{u}=1,\ \mathbf{v}^T\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{v}=1\end{aligned} \tag{4}$$

Hence, $\mathbf{u}$ can be learned as the eigenvector corresponding to smallest eigenvalue of the following generalized eigenvalue problem:

$$\mathbf{X}_t\mathbf{Y}_t^T(\mathbf{Y}_t\mathbf{Y}_t^T)^{-1}\mathbf{Y}_t\mathbf{X}_t^T\mathbf{u}=\lambda\mathbf{X}_t\mathbf{X}_t^T\mathbf{u} \tag{5}$$

where λ is the eigenvalue corresponding to the eigenvector $\mathbf{u}$. $\mathbf{v}$ can be learned as $\mathbf{v}=\frac{(\mathbf{Y}_t\mathbf{Y}_t^T)^{-1}\mathbf{Y}_t\mathbf{X}_t^T\mathbf{u}}{\sqrt{\lambda}}$. By the pair of projection vectors $\mathbf{u}$ and $\mathbf{v}$, the discriminant features $\mathbf{x}_t$ and $\mathbf{y}_t$ can be obtained as $\mathbf{x}_t=\mathbf{u}^T\mathbf{X}_t$ and $\mathbf{y}_t=\mathbf{v}^T\mathbf{Y}_t$, respectively.

(2) Higher-order Correlation Coefficient Analysis

Multiple projection vectors under certain orthonormality constraints can be computed simultaneously by solving the following optimization problem:

$$\begin{aligned}&\{\mathbf{U},\mathbf{V}\}=\min\ \ tr(\mathbf{U}^T\mathbf{X}_t\mathbf{Y}_t^T\mathbf{V})\\ &s.\ t.\ \ \ \mathbf{U}^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{U}=\mathbf{I},\mathbf{V}^T\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{V}=\mathbf{I}\end{aligned} \tag{6}$$

where $\mathbf{U}\in R^{I_1\times\ell}$ and $\mathbf{V}\in R^{J_1\times\ell}$ are the projections matrices, ℓ is the number of projection vectors, and $\mathbf{I}$ is the identity matrix.

According to the objective functions defined in (6), we have:

$$\begin{aligned}\{\mathbf{U},\mathbf{V}\}&=\min\ \ tr(\mathbf{U}^T\mathbf{X}_t\mathbf{Y}_t^T\mathbf{V})\\ &=\min\ \ tr\left((\mathbf{U}^T\mathbf{X}_t)(\mathbf{V}^T\mathbf{Y}_t)^T\right)\\ &=\min\ \ [\![(\mathbf{X}_t\times_1\mathbf{U}^T)\otimes(\mathbf{Y}_t\times_1\mathbf{V}^T);(1:2)(1:2)]\!]\\ s.\ t.\ \ \mathbf{U}^T\mathbf{X}_t{\mathbf{X}_t}^T\mathbf{U}&=\mathbf{I},\mathbf{V}^T\mathbf{Y}_t{\mathbf{Y}_t}^T\mathbf{V}=\mathbf{I}\end{aligned} \tag{7}$$

The solution to the optimization problem consists of the ℓ smallest eigenvectors of the generalized eigenvalue problem in Eq. (5).

In this paper, in order to simultaneously learn the spatial and spectral patterns for bilateral hemispheres, two-way ($channel\times time$) raw EEG samples are constructed into high-order tensors. e.g., we use the complex Morlet wavelet to construct two-way ($channel\times time$) sample $\mathbf{X}^{channel\times time}$ into a third-order tensor $\mathbf{X}^{channel\times frequency\times time}$ in spatial-spectral-temporal domain. Let $\{\underline{\mathbf{X}}_m\}_{m=1}^M$, where $\underline{\mathbf{X}}_m\in R^{I_1\times\cdots\times I_N}$, denote M samples (tensor) in one class (e.g., left motor imagination), while $\{\underline{\mathbf{Y}}_m\}_{m=1}^M$, where $\underline{\mathbf{Y}}_m\in R^{J_1\times\cdots\times J_N}$, denote M samples in the other class (e.g., right motor imagination). For simplicity, we use $\underline{\mathbf{X}}^{I_1\times\cdots\times I_N\times M}$ and $\underline{\mathbf{Y}}^{J_1\times\cdots\times J_N\times M}$ to represent the samples in these two classes, respectively. For the convenience of formulation, the training samples are subtracted to be zero-mean.

The objective of the proposed HOCCA method is to find pairs of projections matrices $\mathbf{U}^{(n)}|_{n=1}^N$, where $\mathbf{U}^{(n)}\in R^{I_n\times L_n}$, and $\mathbf{V}^{(n)}|_{n=1}^N$, where $\mathbf{V}^{(n)}\in R^{J_n\times L_n}$, to project the training samples in these two classes into low dimensional spaces where each pair is minimally correlated. Based on an analogy with (7), we define HOCCA by replacing $\mathbf{U},\mathbf{V},\mathbf{X}_t,\mathbf{Y}_t$ with $\mathbf{U}^{(n)}|_{n=1}^N,\mathbf{V}^{(n)}|_{n=1}^N,\underline{\mathbf{X}},\underline{\mathbf{Y}}$, as:

$$\begin{aligned}\left\{\mathbf{U}^{(n)}|_{n=1}^N,\mathbf{V}^{(n)}|_{n=1}^N\right\}&=\min\ \ [\![(\underline{\mathbf{X}}\prod_{n=1}^N\times_n{\mathbf{U}^{(n)}}^T)\otimes\\ &(\underline{\mathbf{Y}}\prod_{n=1}^N\times_n{\mathbf{V}^{(n)}}^T);(1:N+1)(1:N+1)]\!]\\ s.\ t.\ (\underline{\mathbf{X}}\prod_{n=1}^N\times_n{\mathbf{U}^{(n)}}^T)_{(N+1)}^T&(\underline{\mathbf{X}}\prod_{n=1}^N\times_n{\mathbf{U}^{(n)}}^T)_{(N+1)}=\mathbf{I}\\ (\underline{\mathbf{Y}}\prod_{n=1}^N\times_n{\mathbf{V}^{(n)}}^T)_{(N+1)}^T&(\underline{\mathbf{Y}}\prod_{n=1}^N\times_n{\mathbf{V}^{(n)}}^T)_{(N+1)}=\mathbf{I}\end{aligned} \tag{8}$$

3.2 Alternating solution

The problem defined in Eq. (8) does not have a closed form solution, so we derive a suboptimal solution instead by following the principle of the alternating projection method [33], where the complicated optimization problem is reduced into smaller conditional subproblems that can be solved through simple established methods. Therefore, Eq. (8) is decomposed into N different subproblems, as:

$$\begin{aligned}F&=\min\begin{pmatrix}[\![(\underline{\mathbf{X}}\prod_{n=1}^N\times_n{\mathbf{U}^{(n)}}^T)\otimes(\underline{\mathbf{Y}}\prod_{n=1}^N\times_n{\mathbf{V}^{(n)}}^T);\\ (1:N+1)(1:N+1)]\!]\end{pmatrix}\\ &=\min\begin{pmatrix}[\![(\underline{\mathbf{X}}^{(\bar{n})}\times_n{\mathbf{U}^{(n)}}^T)\otimes\\ (\underline{\mathbf{Y}}^{(\bar{n})}\times_n{\mathbf{V}^{(n)}}^T);(1:N+1)(1:N+1)]\!]\end{pmatrix}\\ &=\min tr\left\{{\mathbf{U}^{(n)}}^T\left([\![(\underline{\mathbf{X}}^{(\bar{n})})\otimes(\underline{\mathbf{Y}}^{(\bar{n})});(\bar{n})(\bar{n})]\!]\right)\mathbf{V}^{(n)}\right\}\\ s.\ t.\ \ &{\mathbf{U}^{(n)}}^T\left([\![(\underline{\mathbf{X}}^{(\bar{n})})\otimes(\underline{\mathbf{X}}^{(\bar{n})});(\bar{n})(\bar{n})]\!]\right)\mathbf{U}^{(n)}=\mathbf{I},\\ &{\mathbf{V}^{(n)}}^T\left([\![(\underline{\mathbf{Y}}^{(\bar{n})})\otimes(\underline{\mathbf{Y}}^{(\bar{n})});(\bar{n})(\bar{n})]\!]\right)\mathbf{V}^{(n)}=\mathbf{I}\end{aligned} \tag{9}$$

To simplify Eq.(9), we define $\mathbf{C}_{xy}^{(n)}$, $\mathbf{C}_{xx}^{(n)}$, $\mathbf{C}_{yy}^{(n)}$ and $\mathbf{C}_{yx}^{(n)}$ as

$$\begin{aligned}\mathbf{C}_{xy}^{(n)}&=([\![(\underline{\mathbf{X}}^{(\bar{n})})\otimes(\underline{\mathbf{Y}}^{(\bar{n})});(\bar{n})(\bar{n})]\!])\\ \mathbf{C}_{xx}^{(n)}&=([\![(\underline{\mathbf{X}}^{(\bar{n})})\otimes(\underline{\mathbf{X}}^{(\bar{n})});(\bar{n})(\bar{n})]\!])\\ \mathbf{C}_{yy}^{(n)}&=([\![(\underline{\mathbf{Y}}^{(\bar{n})})\otimes(\underline{\mathbf{Y}}^{(\bar{n})});(\bar{n})(\bar{n})]\!])\\ \mathbf{C}_{yx}^{(n)}&=([\![(\underline{\mathbf{Y}}^{(\bar{n})})\otimes(\underline{\mathbf{X}}^{(\bar{n})});(\bar{n})(\bar{n})]\!])\end{aligned} \tag{10}$$

Therefore, Eq. (9) is simplified as:

$$\begin{aligned}&\{\mathbf{U}^{(n)},\mathbf{V}^{(n)}\}=\min tr\{{\mathbf{U}^{(n)}}^T\mathbf{C}_{xy}^{(n)}\mathbf{V}^{(n)}\}\\ &s.\ t.\ \ \ {\mathbf{U}^{(n)}}^T\mathbf{C}_{xx}^{(n)}\mathbf{U}^{(n)}=\mathbf{I},\ {\mathbf{V}^{(n)}}^T\mathbf{C}_{yy}^{(n)}\mathbf{V}^{(n)}=\mathbf{I}\end{aligned} \tag{11}$$

We use the Lagrange multipliers to transform the objective function (11) to the following unconstrained multi-variate optimization problem, which is defined as:

$$\begin{aligned}&L(\lambda_x,\lambda_y,\mathbf{U}^{(n)},\mathbf{V}^{(n)})=tr\{{\mathbf{U}^{(n)}}^T\mathbf{C}_{xy}^{(n)}\mathbf{V}^{(n)}\}-\\ &\frac{\lambda_x}{2}\left\|{\mathbf{U}^{(n)}}^T\mathbf{C}_{xx}^{(n)}\mathbf{U}^{(n)}-\mathbf{I}\right\|_F-\frac{\lambda_y}{2}\left\|{\mathbf{V}^{(n)}}^T\mathbf{C}_{yy}^{(n)}\mathbf{V}^{(n)}-\mathbf{I}\right\|_F\end{aligned} \tag{12}$$

where λ_x and λ_y are the Lagrange multipliers.

The optimization is performed by setting the partial derivative of $L(\lambda_x, \lambda_y, \mathbf{U}^{(n)}, \mathbf{V}^{(n)})$ with respect to $\mathbf{U}^{(n)}$ and $\mathbf{V}^{(n)}$ to zero, respectively.

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{U}^{(n)}} &= \lambda_x \mathbf{C}_{xx}^{(n)} \mathbf{U}^{(n)} - \mathbf{C}_{xy}^{(n)} \mathbf{V}^{(n)} = 0 \\ \frac{\partial L}{\partial \mathbf{V}^{(n)}} &= \lambda_y \mathbf{C}_{yy}^{(n)} \mathbf{V}^{(n)} - \mathbf{C}_{yx}^{(n)} \mathbf{U}^{(n)} = 0 \end{aligned} \tag{13}$$

In (13), subtracting ${\mathbf{V}^{(n)}}^T$ times the second equation from ${\mathbf{U}^{(n)}}^T$ times the first equation we have:

$$\lambda_y {\mathbf{V}^{(n)}}^T \mathbf{C}_{yy}^{(n)} \mathbf{V}^{(n)} - \lambda_x {\mathbf{U}^{(n)}}^T \mathbf{C}_{xx}^{(n)} \mathbf{U}^{(n)} = 0 \tag{14}$$

Together with the constraints in the objective function (11), we have $\lambda_x = \lambda_y$. Let $\lambda = \lambda_x = \lambda_y$. Assuming $\mathbf{C}_{yy}^{(n)}$ is invertible we have

$$\mathbf{V}^{(n)} = \frac{{\mathbf{C}_{yy}^{(n)}}^{-1} \mathbf{C}_{yx}^{(n)} \mathbf{U}^{(n)}}{\lambda} \tag{15}$$

Substituting (15) in the first equation in (13) gives:

$$\mathbf{C}_{xy}^{(n)} {\mathbf{C}_{yy}^{(n)}}^{-1} \mathbf{C}_{yx}^{(n)} \mathbf{U}^{(n)} = \lambda^2 \mathbf{C}_{xx}^{(n)} \mathbf{U}^{(n)} \tag{16}$$

Hence, when $\mathbf{C}_{xx}^{(n)}$ and $\mathbf{C}_{yy}^{(n)}$ are nonsingular, $\mathbf{U}^{(n)}$ is achieved as the (unit) generalized eigenvectors corresponding to the smallest L_n generalized eigenvalues of Eq. (16). Then $\mathbf{V}^{(n)}$ is obtained by Eq. (15).

In conclusion, the whole alternating projection optimization procedure of HOCCA is shown in Algorithm 1.

Algorithm 1 Higher-order Correlation Coefficient Analysis (HOCCA) Algorithm

Input: $\underline{\mathbf{X}}^{I_1 \times \cdots \times I_N \times M}$: zero-mean samples for one class. $\underline{\mathbf{Y}}^{J_1 \times \cdots \times J_N \times M}$: zero-mean samples for another class. σ: the threshold to test the convergence. T: the maximum number of iteration.

Output: Pairs of multilinear projections matrices $\left\{\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\right\}_{n=1}^{N}$, where $\mathbf{U}^{(n)} \in R^{I_n \times L_n}$ and $\mathbf{V}^{(n)} \in R^{J_n \times L_n}$.

Method:

1: Initialize $\left\{\mathbf{U}_{(0)}^{(n)}, \mathbf{V}_{(0)}^{(n)}\right\}_{n=1}^{N} = 1$.
2: **for** iteration $t = 1$ to T **do**
3: **for** iteration $n = 1$ to N. **do**
4: Calculate $\mathbf{C}_{xx(t-1)}^{(n)}, \mathbf{C}_{xy(t-1)}^{(n)}, \mathbf{C}_{yx(t-1)}^{(n)}, \mathbf{C}_{yy(t-1)}^{(n)}$ according to Eq. 10.
5: Learn $\mathbf{U}_{(t)}^{(n)}$ and $\mathbf{V}_{(t)}^{(n)}$ according to Eq. 16 and Eq. 15, respectively.
6: **end for**
7: Check: $\sum_{n=1}^{N} \|\mathbf{U}_{(t)}^{(n)} - \mathbf{U}_{(t-1)}^{(n)}\|_F^2 + \|\mathbf{V}_{(t)}^{(n)} - \mathbf{V}_{(t-1)}^{(n)}\|_F^2 \leq \sigma$.
8: **end for**
9: Return all $\mathbf{U}^{(n)} = \mathbf{U}_{(t)}^{(n)}$ and $\mathbf{V}^{(n)} = \mathbf{V}_{(t)}^{(n)}$ for all n.

3.3 Regularized HOCCA

In this section, we formulate a Regularizing HOCCA model (rHOCCA) to address the practical small sample size problem [7]. A regularization term $\eta\mathbf{I}$ is added to $\mathbf{C}_{xx}^{(n)}$ and $\mathbf{C}_{yy}^{(n)}$ defined in the objective function (11) to prevent overfitting and avoid the singularity of $\mathbf{C}_{xx}^{(n)}$ and $\mathbf{C}_{yy}^{(n)}$. Therefore, we have:

$$\begin{aligned} &\{\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\} = \min tr\{{\mathbf{U}^{(n)}}^T \mathbf{C}_{xy}^{(n)} \mathbf{V}^{(n)}\} \\ &s.\ t. \quad {\mathbf{U}^{(n)}}^T (\mathbf{C}_{xx}^{(n)} + \eta\mathbf{I}) \mathbf{U}^{(n)} = \mathbf{I}, \\ &\qquad\ \ {\mathbf{V}^{(n)}}^T (\mathbf{C}_{yy}^{(n)} + \eta\mathbf{I}) \mathbf{V}^{(n)} = \mathbf{I} \end{aligned} \tag{17}$$

Based on a similar derivation with HOCCA, rHOCCA solves the following generalized eigenvalue problem, in terms of $\mathbf{U}^{(n)}$.

$$\mathbf{C}_{xy}^{(n)} (\mathbf{C}_{yy}^{(n)} + \eta\mathbf{I})^{-1} \mathbf{C}_{yx}^{(n)} \mathbf{U}^{(n)} = \lambda^2 (\mathbf{C}_{xx}^{(n)} + \eta\mathbf{I}) \mathbf{U}^{(n)} \tag{18}$$

and $\mathbf{V}^{(n)}$ can be obtained as $\mathbf{V}^{(n)} = (\mathbf{C}_{yy}^{(n)} + \eta\mathbf{I})^{-1} \mathbf{C}_{yx}^{(n)} \mathbf{U}^{(n)} / \lambda$.

3.4 Complexity

The computational complexity and memory requirements of HOCCA are analyzed, in a similar way as in [24]. From a computational complexity point of view, the most demanding steps involve the calculations of the projection $\mathbf{C}_{xx(t-1)}^{(n)}, \mathbf{C}_{xy(t-1)}^{(n)}, \mathbf{C}_{yx(t-1)}^{(n)}, \mathbf{C}_{yy(t-1)}^{(n)}$. Therefore, the overall computational complexity is $TNt'_{\max} \sum_{n=1}^{N} I^n$.

With respect to the memory requirement, the respective computation can be done incrementally by reading sequentially. Hence, the memory needed for the HOCCA algorithm can be as low as I^n (ignoring lower-order terms).

4 Experimental Configuration

4.1 Data Acquisition

We conducted several experiments on two motor imagery datasets, which were one benchmark dataset recorded from 5 normal persons and one dataset recorded from 10 stroke patients, to validate the performance of HOCCA. The task was to classify the type of the imagination for each trial in an offline fashion.

(1) Dataset I: Dataset I was collected from five healthy subjects (labeled 'aa', 'al', 'av', 'aw' and 'ay' respectively) performing right hand and right foot motor imagery in a benchmark dataset of dataset IVa from BCI competition III [10]. EEG was recorded using 118 electrodes, and bandpass filtered between 8 and 30 Hz. We extracted a time segment from 500 to 2500 ms after the cue. A training set and a testing set were available for each subject. Their size was different for each subject. More precisely, 280 trials were available for each subject, among which 168, 224, 84, 56 and 28 composed the training set for the five subjects and the remaining trials were their test set.

(2) Dataset II: Dataset II was collected from 10 unilaterally paralyzed stroke patients diseased in two months who performed motor imagination of their disabled (left or right) upper limb in a BCI combined with Functional Electrical Stimulation (BCI-FES) rehabilitation training system for 24 times over two months (three times per week). EEG data was recorded by a 16-channel (FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier at a sampling rate of 256 Hz. Patients had to complete basic motor imagery related tasks for five sessions and each session comprised forty trials and lasted for 240 s. At the beginning of each trial, a bold arrow as the visual cue with a command, randomly left or right, was shown on the screen, instructing patients to imagine left or right. We extracted a time segment starting from 0.5s to 4.5s after the visual cue for analysis. All the trials in the training model course were divided into a training set with 120 trials and a testing set with 80 trials.

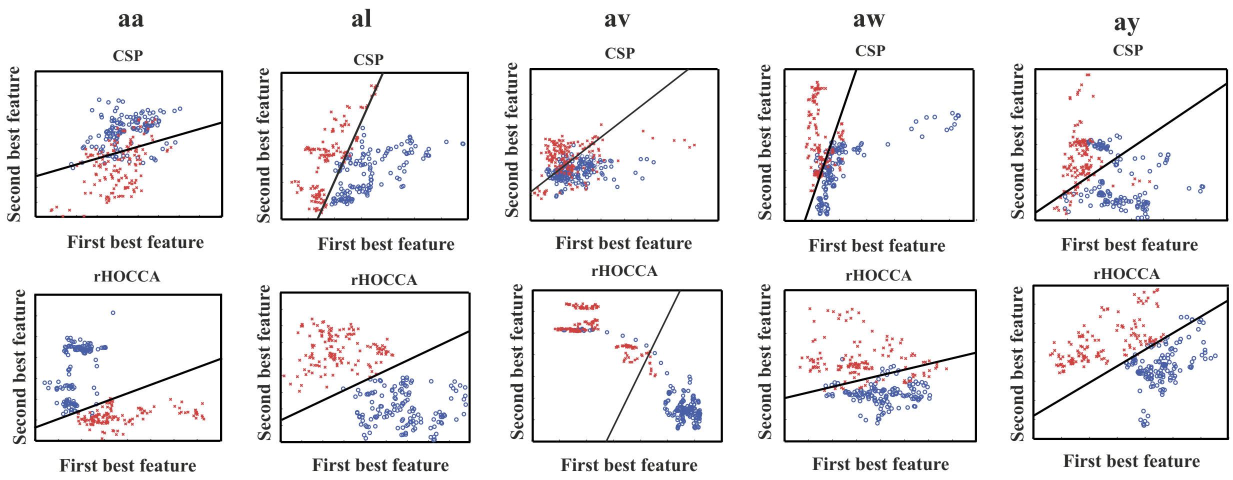

Figure 1. Distributions of two best features obtained by CSP and the proposed rHOCCA filters for Subjects aa, al, av, aw and ay. The best features were obtained by using the Fisher score strategy. The blue circles and red crosses denoted the first and second features of the two classes, respectively. The black line represented the LDA hyperplane. The features were plotted after normalization.

4.2 Data Preprocessing

EEG signals were bandpass filtered within a specific frequency band related to motor imagery. For healthy people, exemplary spectral characteristics of EEG in motor imagery tasks were α rhythm (8-13 Hz) and β rhythm (14-30 Hz) [27]. Thus Dataset I was bandpass filtered in the frequency range of 8-30 Hz. However, it was not available to obtain the spectral characteristics of stroke patients [17]. Thus raw data in Dataset II was filtered in a general band ranged from 5 to 40 Hz.

4.3 Feature Extraction and Classification

We employed five state-of-the-art approaches for comparison including Common Spatial Pattern (CSP) [28], Common Sparse Spectral Spatial Pattern (CSSSP) [14], Filter Bank Common Spatial Pattern (FBCSP) [1], Bayesian Spatio-spectral Filter Optimization (BSSFO) [30], and General Tensor Discriminant Analysis (GTDA) [32]. We employed each algorithm to extract discriminative features from different types of data to detect change of rhythmic activity in different subjects.

CSP, which has been evidenced as the most successful algorithms for EEG classification, discriminated two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class [28]. CSSSP method performed simultaneous optimization of an arbitrary (Finite Impulse Response) FIR filter within the CSP algorithm [14]. FBCSP, whose efficacy was demonstrated in the latest BCI Competition IV [26], selected discriminative frequency bands by means of the mutual information between class labels and feature vectors [1]. BSSFO utilized a Bayesian framework to simultaneously optimize spectral filters and spatial filters along with a modified factored-sampling method for probability density function (pdf) estimation [30]. GTDA, which was the extension of two-dimensional Linear Discriminant Analysis (LDA) in tensor space, was applied to reserve multilinear discriminative subspace from the training tensors [32].

For CSSSP, the regularization constant C was set to 0.3. FBCSP filtered the EEG signals into six nonoverlapping subband components. For BSSFO, particles were generated from a mixture of Gaussians as described in the original paper [30]. Finally, for GTDA and the proposed HOCCA and rHOCCA, we constructed the two-dimensional EEG data into third-order EEG tensors in spatial-spectral-temporal domain by a complex Morlet wavelet [34]. i.e. $\phi(t) = \exp(2i\pi t)\exp(-t^2/2)/\sqrt{2\pi}$ (center frequency: 1; bandwidth parameter: 2). Then we combined all the training samples into a fourth-order tensor as $channels \times frequency \times time \times trials$. Regularization parameter η in rHOCCA was set to 0.1.

A linear support vector machine (SVM) [12], which obtained top-level performance in many applications, was conducted for classification. A 5-fold cross-validation was used to choose suitable SVM parameters.

5 Experimental Results

In this section, we presented the experimental results on the two aforementioned datasets, where the first dataset was one widely-used benchmark dataset of Dataset IVa from BCI competition III and the last one referred to stroke patients. To begin with, we applied the proposed method on the first dataset where motor imagery related spatial and spectral characteristics were known, and evaluated its performance when compared with some traditional algorithms. Finally, the proposed method was applied on the second dataset where the discriminative spatial and spectral properties were not specifically identified.

5.1 Results on Dataset IVa from BCI competition III

The results were mainly given in two aspects of the classification accuracy and the merits of HOCCA.

(1) Classification accuracy. For HOCCA, rHOCCA, the training samples were divided into two parts: $\underline{\mathbf{X}}$ with all the samples in the right hand imagery, and $\underline{\mathbf{Y}}$ with all the samples in the right foot imagery. $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ were used as input for HOCCA and rHOCCA to find a pair of projection matrices $\{\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\}_{n=1}^{3}$ by which EEG samples from right hand imagery and right foot imagery were projected separately into their individual space. Each sample in the test dataset was projected into the two individual low-dimensional spaces

$\{\mathbf{U}^{(n)}\}_{n=1}^{3}$ and $\{\mathbf{V}^{(n)}\}_{n=1}^{3}$ separately, yielding two feature vectors representing the two classes. We finally combined these two feature vectors into one feature vector for classification. The feature dimensionality was set to 32. Table 1 showed the classification accuracies obtained by all the methods on the benchmark dataset. It is obvious that HOCCA and rHOCCA yielded superior classification accuracies against the other algorithms, and rHOCCA had the best performance. With a closer look at the HOCCA results, it was realized that bigger improvements were achieved by subject av with relatively poor performances.

Table 1. The classification accuracies of all the methods on Dataset IVa from BCI competition III (%)

Subject	CSP	CSSSP	FBCSP	BSSFO	GTDA	HOCCA	rHOCCA
aa	73.3	80.1	84.6	78.3	69.8	92.1	**93.7**
al	82.3	91.8	90.5	92.7	86.4	100	**100**
av	52.4	56.6	61.7	65.4	72.5	79.9	**80.2**
aw	78.6	89.2	86.3	89.2	81.2	98.7	**99.3**
ay	74.2	79.5	83.5	72.5	80.8	98.5	**98.9**
Mean	72.16	79.44	81.32	79.62	78.14	93.84	**94.42**

(2) Toward Understanding the Merits of HOCCA: In the previous section, we showed quantitative evidences indicating the proposed HOCCA-based methods can improve the classification accuracy in EEG-based BCIs. In this section, we provided more analysis and visualizations to better understand the nature and the impact of our proposed algorithm on nonstationary changes in the EEG signals and the feature space.

Fig. 1 visualized the features obtained by CSP and rHOCCA for Subjects aa, al, av, aw and ay, respectively. It is noted that for the ease in visualization only two features which had the highest Fisher scores in the dataset were plotted. The features were plotted after normalization. Red crosses represented one class while blue circles stood for the other. The black line represented the LDA hyperplane. Comparing the distributions of the two features with the highest Fisher scores extracted from CSP and the proposed rHOCCA algorithm clearly revealed that the rHOCCA features were more compact and thus more separable. It was observed that the features obtained by rHOCCA were more discriminative than CSP features, indicating learning individual spatial-spectral subspaces for each class contributed more to classification than only learning individual spatial subspaces.

To better explain the performance differences between the CSP and the rHOCCA algorithms, Fig. 2 compares some examples of the most relevant individual spatial-spectral filters obtained by rHOCCA for right hand class and right foot class, respectively, and Fig. 3 shows the spatial filters obtained by CSP for the five subjects. For rHOCCA, the most relevant subspaces on each mode of EEG tensor for each class were selected from the optimized paired projection matrices $\left\{\mathbf{U}^{(n)}, \mathbf{V}^{(n)}\right\}_{n=1}^{3}$ by a Fisher score strategy [7]. In general, these pictures showed that the important channels obtained by rHOCCA for four out of the five subjects (aa, al, aw and ay) were physiologically relevant, with strong weights over the motor cortex areas, as also expected from the study [27]. The most important channels obtained by rHOCCA were mainly centered on central cortical area (for right foot imagination) and left cortical area (for right hand imagination). The same phenomenon was also observed in the pictures given by CSP. Moreover, the results indicated that the channels and frequency information were not only subject-dependent but also class-dependent. e.g., for subject al, the largest variance of right foot mainly focused on central scalp map and a large peak around 20-25 Hz, while the largest variance of right hand mainly focused on left area and frequency band of 10-15 Hz.

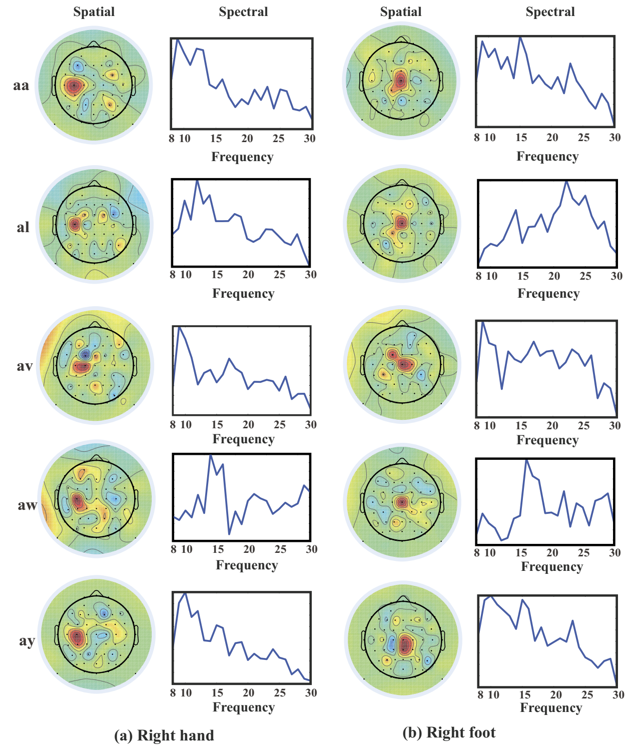

Figure 2. The most relevant subspaces in spatial-spectral domain for right hand class and right foot class, respectively. (a) Right hand class; (b) Right foot class. Left column represents projection matrices on spatial modality and right column represents projection matrices on frequency modality.

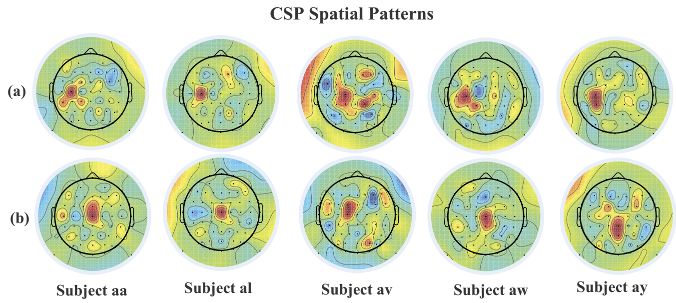

Figure 3. Spatial weights for the two most discriminative filters constructed by CSP for all the subjects in Dataset I. (a) Spatial filters for performing right hand motor imagery; (b) Spatial filters for performing right foot motor imagery.

5.2 Results on Dataset collected from stroke patients

In the above experiments, we showed quantitative evidences indicating the proposed HOCCA-based methods can achieve an improvement in classification accuracy for normal persons whose related spatial and spectral characteristics were available. In this experiment, without any prior knowledge like the active motor cortex regions and frequency bands, we tried to verify the feasibility and robustness of HOCCA when decoding the unknown information of stroke patients' motor imagery. The results were mainly given in three aspects: (1) Classification accuracy. (2) Merits of HOCCA. (3) Neurophysiologic rehabilitation mechanism in impaired cortex.

(1) Classification accuracy. For every patient, classification accuracy in each day was calculated, where feature dimensionality was set to 16. Then classification accuracies in the same month were averaged to represent the mean accuracy of the month. Table 2 gives the mean accuracies of the two months achieved by all the methods on the ten patients.

Note that HOCCA and rHOCCA methods outperformed all the other algorithms, and rHOCCA achieved the best classification performance. With a closer look at the results, it was realized that bigger improvements were achieved by almost all these patients with poor CSP performances. Comparisons using a Mann-Whitney U test between HOCCA-based methods and the other methods showed that the accuracies by HOCCA and rHOCCA were significantly higher.

Table 2. Mean classification accuracies of the two months obtained by all the methods on stroke patients (%)

Subject	Month	CSP	CSSSP	FBCSP	BSSFO	GTDA	HOCCA	rHOCCA
S1	M1	54.4	56.5	62.5	63.4	58.3	70.7	**71.4**
	M2	63.7	65.2	68.4	65.3	62.4	77.4	**77.6**
S2	M1	57.7	58.6	64.7	62.4	67.7	75.1	**75.3**
	M2	66.5	72.2	71.7	68.3	73.7	85.1	**85.9**
S3	M1	42.7	53.6	55.8	56.7	59.6	**72.7**	71.3
	M2	58.7	67.2	68.3	71.4	72.5	81.5	**82.3**
S4	M1	44.1	52.2	53.3	56.6	62.3	72.3	**73.5**
	M2	58.3	63.3	65.8	64.4	63.6	74.7	**74.9**
S5	M1	62.1	67.7	65.2	63.3	65.4	72.7	**74.7**
	M2	67.9	71.4	77.8	75.9	72.1	84.3	**84.7**
S6	M1	41.5	56.3	55.7	57.6	59.2	**67.3**	66.6
	M2	59.6	64.8	67.1	65.2	68.2	79.6	**79.8**
S7	M1	51.6	62.6	66.7	66.6	65.3	75.7	**76.6**
	M2	62.2	66.8	67.7	64.2	68.5	86.4	**86.6**
S8	M1	41.3	50.1	53.8	54.2	55.7	66.2	**67.4**
	M2	53.4	59.6	61.2	62.2	60.6	74.2	**74.4**
S9	M1	38.4	48.9	51.1	52.6	53.8	67.4	**68.2**
	M2	47.5	58.6	60.2	58.3	62.5	**74.3**	**74.3**
S10	M1	48.7	61.3	59.8	62.4	57.3	73.7	**74.1**
	M2	52.2	63.6	64.5	61.4	66.8	78.3	**79.2**
Mean	M1	48.2	56.7	58.6	59.5	60.4	71.3	**71.9**
	M2	59.0	65.2	67.2	65.6	67.1	79.5	**79.9**

(2) Toward Understanding the Merits of HOCCA. In this part, we provided more analysis and visualizations to better understand the nature and the impact of our proposed algorithm when applied on stroke patients. Fig. 4 visualized the features obtained by all the methods except FBCSP for Patient 5 on day 60. FBCSP employed a filter bank that bandpass filtered the EEG measurements into multiple bands, and then multiple pairs of CSP features were constructed based on each pair of bandpass and spatial filter. Thus, we did not visualize all these FBCSP features in each pair of bandpass and spatial filter. It is noted that for the ease in visualization, only two features which had the highest Fisher scores [7] were plotted. The features were plotted after normalization. Red crosses represented one class while blue circles stood for the other. The black line represented the Linear Discriminant Analysis (LDA) hyperplane. The results clearly revealed that the HOCCA and rHOCCA features were more compact and separable, indicating learning individual spatial-spectral subspaces for each class contributed more to classification than learning universal ones.

(3) Neurophysiologic plasticity mechanism in impaired cortex. To better understand the effectiveness and robustness of HOCCA-based methods when decoding the motor imagery patterns of stroke patients which were not available beforehand, we visualized the spatial-spectral filters obtained by rHOCCA method and the spatial filters learned by CSP in terms of Patient 5. In order to observe the gradual changes of EEG patterns in spatial and spectral domains over time, we chose the raw EEG of three days, day 1, 30 and 60, to represent the different phases during rehabilitation. rHOCCA was utilized to extract the spatial-spectral filters of Patient 5 with lesion in right side on these three days, as shown in Fig. 5. For comparison, Fig. 6 illustrated the chosen three days' spatial filters learned by CSP fo Patient 5. From these pictures, we can clearly see that the channels and frequency information were class-dependent. e.g., for spectral information on Day 1, the largest variance of right movement imagination mainly focused on 6-12 Hz, while the largest variance of left movement imagination mainly focused on the frequency bands of 5-12 Hz and 20-30 Hz. Another important observation was that during motor recovery motor imagery EEG patterns of stroke patient were changing and quite different from the ones by motor imagery of normal persons. The differences attributed to active cortex regions and frequency bands. In detail, the spatial filters obtained by CSP appeared as messy, with large weights in several unexpected locations from a neurophysiological point of view. On the contrary, rHOCCA filters were physiologically more relevant. In detail, the most significant channels for right movement imagination were focused at around left central areas (like C3), as also expected from the study [27]; however, the channels contributed to left movement imagination were with strong weights over not only the right central areas (like C4) but also the frontal-central and parietal areas (like FC4 and P4). Similar phenomena were also reported in some other studies [11, 15]. As for spectral characteristics, active frequency band was updated from a wide-ranged band (8-30 Hz) at the beginning to a lower band (8-13 Hz) after two months. This dynamic band accentuation implied that active rhythms might be modulated during rehabilitation. Similar observation was also reported from the literature [29].

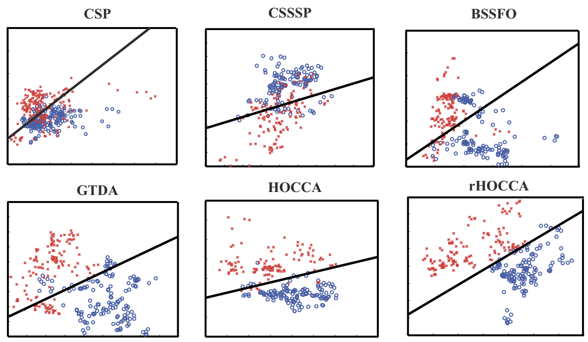

Figure 4. Distributions of the best two features obtained by all the methods except FBCSP for Patient 5 on day 60. The blue circles and red crosses denoted the features of the two classes, respectively. The black line represented the LDA hyperplane. The features were plotted after normalization.

6 Discussion

This study presented a tensor-based algorithm, namely HOCCA, for EEG classification which aimed to seek individual spatial and spectral subspaces for each class so that each class was projected into its own subspace separately. First, EEG was represented by high order tensors (multiway arrays) i.e., multiple-modality patterns in the spatial-spectral-temporal domain. Next, HOCCA was applied on

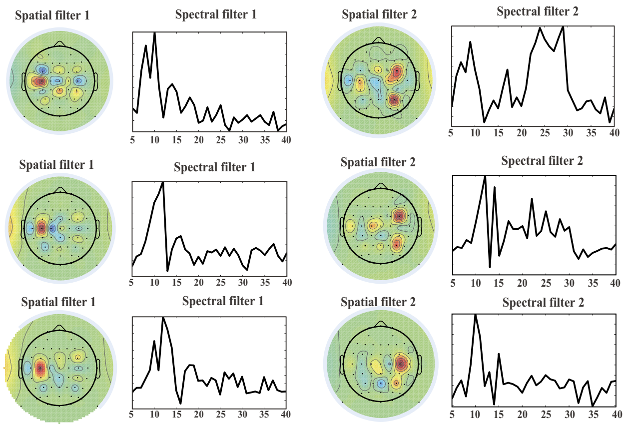

Figure 5. rHOCCA spatial-spectral filters of day 1, 30 and 60 for Patient 5, respectively. Spatial-spectral filters 1 and 2 correspond to right and left upper limbs imagery, respectively. From top row to bottom row: day1, 30 and 60. x-axis in the spectral filters represents frequency while y-axis is power and values are normalized. In spatial filter, red dots show higher power while blue ones represent lower power.

high order EEG tensors to obtain individual multilinear discriminative subspace for each class by minimizing the mutual correlation between classes, and thus high-dimensional tensors were mapped to low-dimensional tensors. Finally, some classical classifiers were conducted for classification in the feature space with the reduced dimension.

Extensive experiment comparisons were performed on two datasets containing one dataset acquired from normal persons and one dataset related to unilaterally paralyzed stroke patients. Experimental results showed that our algorithm yielded relatively higher classification accuracies compared with five state-of-the-art approaches. In particular, considering the classification results in the two datasets, one interesting phenomenon was observed: for the subjects with poor CSP performances (i.e., CSP error rate more than 30% like av in Dataset I, and almost all the subjects in Dataset II), the proposed HOCCA-based algorithms significantly outperformed the CSP algorithm. Moreover, compared to Dataset I, the classification performance difference between CSP and HOCCA-based methods in the neuro-rehabilitation Dataset II was more salient.

To explore the reasons, firstly, it could be the fact that HOCCA-based methods, which are different from the other methods which directly use one common subspace for projection of all the classes, seeks a set of individual multilinear subspaces for each class and projects the samples in each class into its own subspace. This makes sense, since the spatial and spectral information are not only subject-dependent but also class-dependent, which can be found in the evident difference of both spatial and frequency patterns in two classes' EEG shown in Fig. 5. Secondly, the performance of the CSP-based methods heavily depends on their operational frequency band and channels, which, however, can not be obtained for stroke patients beforehand. As for stroke patents, their injured brain is under recovery. There should be a higher chance that the most contributed channels group and frequency bands will change [17, 5, 23]. To the best of our knowledge, there is no agreed clinical conclusion about motor imagery patterns of stroke patients. Our HOCCA-based methods are more robust than CSP-based methods by performing autonomous selection of key channels and frequency band. The robustness of our proposed methods is very useful because it can be applied to extract the discriminative features and patterns when the spatial and spectral characteristics in some paradigms are not available. Thirdly, different from CSP-based methods, the proposed algorithm is performed in the tensor space rather than the vector space, yielding both the spatial and frequency filters containing the maximum discriminative information. The best performance achieved by HOCCA-based methods in the classification results evidence that channel selection and frequency selection extract extra effective information of EEG, and they complement each other, both contributing to the classification ability. All these results show that HOCCA-based methods could be more successful and robust in capturing the spatial-spectral filters.

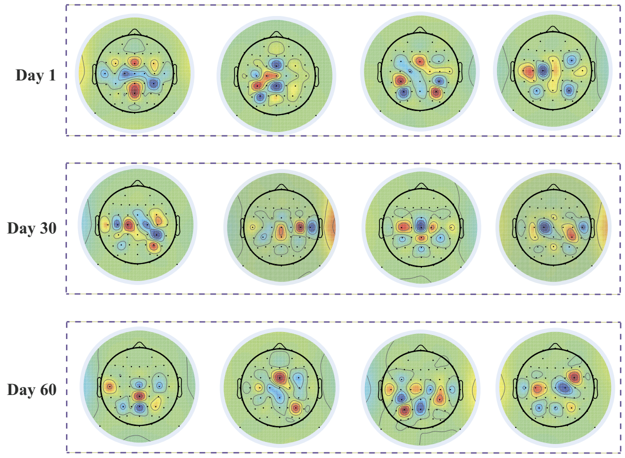

Figure 6. CSP spatial patterns for Patient 5 on day 1, 30 and 60 (from top to bottom: day 1, 30 and 60). Red dots show higher power while blue ones represent lower power.

7 Conclusions

In this work, a tensor-based method, called Higher-order Correlation Coefficient Analysis (HOCCA), was proposed for motor imagery EEG classification. Since spatial and spectral information of EEG data are not only subject-dependent but also class-dependent, HOCCA attempted to seek a set of individual multilinear subspaces for each class rather than a single common subspace for all the samples, by which the samples belonging to different classes were projected into their own subspace separately such that they were easily to be classified. Experimental analysis for classification of motor imagery EEG in two different datasets recorded from the healthy people and stroke patients demonstrated the superior performance of our algorithm when compared with some state-of-the-art methods, indicating that learning individual multi-modal spaces for each class contributed more to classification than learning only one common space.

ACKNOWLEDGEMENTS

The work was supported by the Key Basic Research Program of Shanghai(15JC1400103), the National Basic Research Program of China (Grant No. 2015CB856004), the National Natural Science Foundation of China (Grant No. 61272251) and JSPS KAKENHI (Grant No. 15K15955).

REFERENCES

[1] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang, 'Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b', *Frontiers in neuroscience*, **6**, (2012).

[2] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan, 'Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs', *Pattern Recognition*, **45**(6), 2137–2144, (2012).

[3] Kai Keng Ang, Karen Sui Geok Chua, Kok Soon Phua, Chuanchu Wang, Zheng Yang Chin, Christopher Wee Keong Kuah, Wilson Low, and Cuntai Guan, 'A randomized controlled trial of EEG-based motor imagery Brain-Computer Interface robotic rehabilitation for stroke', *Clinical EEG and neuroscience*, 1550059414522229, (2014).

[4] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek, 'Optimizing the channel selection and classification accuracy in EEG-based BCI', *IEEE Transactions on Biomedical Engineering*, **58**(6), 1865–1873, (2011).

[5] Sergi Bermudez i Badia, Hani Samaha, Andres Garcia Morgade, and Paul FMJ Verschure, 'Exploring the synergies of a hybrid BCI-VR neurorehabilitation system', in *Virtual Rehabilitation (ICVR), 2011 International Conference on*, pp. 1–8. IEEE, (2011).

[6] N Birbaumer, N Ghanayim, T Hinterberger, and B Iversen, 'A spelling device for the paralysed', *Nature*, **398**(6725), 297–298, (1999).

[7] Christopher M Bishop and Nasser M Nasrabadi, *Pattern recognition and machine learning*, volume 1, springer New York, 2006.

[8] Gilles Blanchard and Benjamin Blankertz, 'BCI competition 2003-data set IIa: spatial patterns of self-controlled brain rhythm modulations', *IEEE Transactions on Biomedical Engineering*, **51**(6), 1062–1066, (2004).

[9] B. Blankertz, G. Curio, and K. Muller, 'Classifying single trial EEG: Towards brain computer interfacing', in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 157–164, (2002).

[10] Benjamin Blankertz, K-R Muller, Dean J Krusienski, et al., 'The BCI competition III: Validating alternative approaches to actual BCI problems', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**(2), 153–159, (2006).

[11] Cinzia Calautti, François Leroy, Jean-Yves Guincestre, and Jean-Claude Baron, 'Dynamics of Motor Network Overactivation After Striatocapsular Stroke A longitudinal PET Study Using a Fixed-Performance Paradigm', *Stroke*, **32**(11), 2534–2542, (2001).

[12] Chih-Chung Chang and Chih-Jen Lin, 'LIBSVM: a library for support vector machines', *ACM Transactions on Intelligent Systems and Technology (TIST)*, **2**(3), 27, (2011).

[13] Evariste Demandt, Carsten Mehring, Klaus Vogt, Andreas Schulze-Bonhage, Ad Aertsen, and Tonio Ball, 'Reaching movement onset-and end-related characteristics of eeg spectral power modulations', *Frontiers in neuroscience*, **6**, (2012).

[14] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and K-R Muller, 'Combined optimization of spatial and temporal filters for improving brain-computer interfacing', *IEEE Transactions on Biomedical Engineering*, **53**(11), 2274–2281, (2006).

[15] A Feydy, R Carlier, A Roby-Brami, B Bussel, F Cazalis, L Pierot, Y Burnod, and MA Maier, 'Longitudinal study of motor recovery after stroke recruitment and focusing of brain activation', *Stroke*, **33**(6), 1610–1617, (2002).

[16] Tamara G Kolda and Brett W Bader, 'Tensor decompositions and applications', *SIAM review*, **51**(3), 455–500, (2009).

[17] Darren J Leamy, Juš Kocijan, Katarina Domijan, Joseph Duffin, Richard AP Roche, Sean Commins, Rónán Collins, and Tomas E Ward, 'An exploration of EEG features during recovery following stroke–implications for BCI-mediated neurorehabilitation therapy', *Journal of neuroengineering and rehabilitation*, **11**(1), 9, (2014).

[18] Steven Lemm, Benjamin Blankertz, Gabriel Curio, and K-R Muller, 'Spatio-spectral filters for improving the classification of single trial EEG', *IEEE Transactions on Biomedical Engineering*, **52**(9), 1541–1548, (2005).

[19] Ye Liu, Mingfen Li, Hao Zhang, Hang Wang, Junhua Li, Jie Jia, Yi Wu, and Liqing Zhang, 'A tensor-based scheme for stroke patients' motor imagery EEG analysis in BCI-FES rehabilitation training', *Journal of neuroscience methods*, **222**, 238–249, (2014).

[20] Ye Liu, Hao Zhang, Min Chen, and Liqing Zhang, 'A boosting-based spatial-spectral model for stroke patients' EEG analysis in rehabilitation training', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **24**(1), 169–179, (2016).

[21] Ye Liu, Qibin Zhao, and Liqing Zhang, 'Uncorrelated multiway discriminant analysis for motor imagery EEG classification', *International journal of neural systems*, **25**(04), 1550013, (2015).

[22] Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, Bruno Arnaldi, et al., 'A review of classification algorithms for EEG-based brain–computer interfaces', *Journal of neural engineering*, **4**, (2007).

[23] Isabelle Loubinoux, Christophe Carel, Jérémie Pariente, Sophie Dechaumont, Jean-François Albucher, Philippe Marque, Claude Manelfe, and François Chollet, 'Correlation between cerebral reorganization and motor recovery after subcortical infarcts.', *Neuroimage*, **20**(4), 2166, (2003).

[24] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos, 'Uncorrelated multilinear discriminant analysis with regularization and aggregation for tensor object recognition', *IEEE Transactions on Neural Networks*, **20**(1), 103–123, (2009).

[25] J. R. Millan, F. Renkens, J. Mourino, and W. Gerstner, 'Non-invasive brain-actuated control of a mobile robot by human EEG', *IEEE Transactions on Biomedical Engineering*, **51**(6), 1026–1033, (2004).

[26] M Naeem, C Brunner, R Leeb, B Graimann, and G Pfurtscheller, 'Seperability of four-class motor imagery data using independent components analysis', *Journal of neural engineering*, **3**(3), 208, (2006).

[27] Gert Pfurtscheller and Christa Neuper, 'Motor imagery and direct brain-computer communication', *Proceedings of the IEEE*, **89**, 1123–1134, (2001).

[28] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, 'Optimal spatial filtering of single trial EEG during imagined hand movement', *IEEE Transactions on Rehabilitation Engineering*, **8**, 441–446, (2000).

[29] S. Shahid, R.K. Sinha, and G. Prasad, 'Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation', *BMC Neuroscience*, **11**, 1–2, (2010).

[30] Heung-Il Suk and Seong-Whan Lee, 'A novel bayesian framework for discriminative feature extraction in brain-computer interfaces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(2), 286–299, (2013).

[31] Wing-Kin Tam, Kai-yu Tong, Fei Meng, and Shangkai Gao, 'A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **19**(6), 617–627, (2011).

[32] D. Tao, X. Li, X. Wu, and S.J. Maybank, 'General tensor discriminant analysis and gabor features for gait recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **29**(10), 1700–1715, (2007).

[33] Dacheng Tao, Xuelong Li, Xindong Wu, and Steve Maybank, 'Tensor rank one discriminant analysis-a convergent method for discriminative multilinear subspace selection', *Neurocomputing*, **71**(10), 1866–1882, (2008).

[34] Anthony Teolis, *Computational signal processing with wavelets*, Springer Science & Business Media, 2012.

[35] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, Theresa M Vaughan, et al., 'Brain-computer interfaces for communication and control', *Clinical neurophysiology*, **113**(6), 767–791, (2002).

[36] Wei Wu, Xiaorong Gao, Bo Hong, and Shangkai Gao, 'Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL)', *IEEE Transactions on Biomedical Engineering*, **55**(6), 1733–1743, (2008).

[37] Qibin Zhao, Cesar F Caiafa, Danilo P Mandic, Zenas C Chao, Yasuo Nagasaka, Naotaka Fujii, Liqing Zhang, and Andrzej Cichocki, 'Higher order partial least squares (HOPLS): a generalized multilinear regression method', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**(7), 1660–1673, (2013).

[38] Qibin Zhao, Tomasz M Rutkowski, Andrzej Cichocki, and Liqing Zhang, 'High resolution common spatial frequency filters for classifying multi-class EEG', in *Advances in Cognitive Neurodynamics (II)*, 683–688, Springer, (2011).

[39] Qibin Zhao, Liqing Zhang, and Andrzej Cichocki, 'Multilinear generalization of common spatial pattern', in *ICASSP 2009*, pp. 525–528. IEEE, (2009).