Neurocomputing 108 (2013) 58–68

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification

Tae-Eui Kam [a], Heung-Il Suk [a], Seong-Whan Lee [b,*]

[a] Department of Computer Science and Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-713, Republic of Korea
[b] Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-713, Republic of Korea

ARTICLE INFO

Article history:
Received 9 December 2011
Received in revised form
28 September 2012
Accepted 5 December 2012
Communicated by Sanqing Hu
Available online 2 January 2013

Keywords:
Brain–Computer Interface (BCI)
Electroencephalogram (EEG)
Motor imagery classification
Spatial filter optimization

ABSTRACT

Neuronal power attenuation or enhancement in specific frequency bands over the sensorimotor cortex, called Event-Related Desynchronization (ERD) or Event-Related Synchronization (ERS), respectively, is a major phenomenon in brain activities involved in imaginary movement of body parts. However, it is known that the nature of motor imagery-related electroencephalogram (EEG) signals is non-stationary and highly variable over time and frequency. In this paper, we propose a novel method of finding a discriminative time- and frequency-dependent spatial filter, which we call 'non-homogeneous filter.' We adaptively select bases of spatial filters over time and frequency. By taking both temporal and spectral features of EEGs in finding a spatial filter into account it is beneficial to be able to consider non-stationarity of EEG signals. In order to consider changes of ERD/ERS patterns over the time–frequency domain, we devise a spectrally and temporally weighted classification method via statistical analysis. Our experimental results on the BCI Competition IV dataset II-a and BCI Competition II dataset IV clearly presented the effectiveness of the proposed method outperforming other competing methods in the literature.

1. Introduction

A Brain–Computer Interface (BCI), which translates human intentions into commands to control an external device by analyzing and recognizing brain signals, allows users to communicate with their environment without using the brains normal output pathways of peripheral nerves and muscles [1]. Therefore, it has been of great interest to many research groups for patients that suffer from amyotrophic lateral sclerosis, locked-in syndrome, etc.

Among various methods of eliciting the specific regulation of brain signals, imagined body-part movement, called motor imagery, has been most widely used in an Electroencephalogram (EEG)-based BCI due to its naturalness and spontaneousness. In terms of neurophysiology, motor imagery accompanies attenuation (ERD: Event-Related Desynchronization) or enhancement (ERS: Event-Related Synchronization) of rhythmical synchrony over the sensorimotor cortex in the specific frequency bands of μ-rhythm (8–13 Hz) and β-rhythm (14–30 Hz) [2,3]. However, it is also known that the motor imagery responsive frequency bands are highly variable for inter- and intra-subjects. Furthermore, due to the volume conduction effect in EEG [4–6] finding a class-discriminative spatial filter that certainly improves classification performance has been considered as a challenging problem in the BCI community [7–9].

While many previous methods showed satisfactory results in their experiments, they did not consider the non-stationary feature of human brain activity [10–12]. That is, they found an optimal spatial filter from samples in the predefined time segment and applied it to all the sample points within the time segment homogeneously in classification. We call this type of spatial filter 'homogeneous' spatial filter. On the while, in this paper, we propose a novel method of finding a bank of spatial filters, called 'non-homogeneous' spatial filters, each of which is learned for a different segment in time.

The main contributions of this paper are two-fold. First, we propose a novel method of finding non-homogeneous spatial filters. Unlike other Common Spatial Patterns (CSP)-based methods that find optimal spatial filters and apply them homogeneously over all samples, the proposed method optimizes spatial filters for each of the time–frequency segments to reflect the non-stationary feature of human brain activity. Second, we devise a spectrally and temporally weighted classification method. A weight determined by sum of Fisher ratios which are the measure of the discrimination level of selected bases of each spatial filter based on a ratio of between-class variance and within-class variance is assigned for each time–frequency segment. From our experiment on a public dataset, it was proved that the proposed weighting scheme improved the classification accuracy.

The rest of the paper is organized as follows. We review the previous work in Section 2. In Section 3, we describe a method

* Corresponding author. Tel.: +82 2 3290 3197; fax: +82 2 3290 3583.
E-mail addresses: kamte@korea.ac.kr (T.-E. Kam), hisuk@korea.ac.kr (H.-I. Suk), sw.lee@korea.ac.kr (S.-W. Lee).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.12.002

for non-homogeneous spatial filters optimization and for the construction of a time–frequency weighted classifier. Experimental results and analysis are presented in Section 4, and concluding remarks are drawn in Section 5.

2. Related work

One of the most effective and widely used algorithms to find an optimal spatial filter is the Common Spatial Pattern (CSP) [8,13,14], which maximizes the variance of EEG signals of one class while minimizing the variance of EEG signals of the other class. Here, a class refers to a motor imagery task. However, as the CSP algorithm considers only spatially discriminative patterns while ignoring the spectral characteristics of EEG signals, the performance is dependent on the preceding spectral filtering, in which the frequency band is commonly predetermined and fixed manually [15,16]. Hereafter, we use the terms of spectral filter and frequency band interchangeably throughout the paper.

For optimization of the spectral filter, variants of CSP methods, namely, Common Spatio-Spectral Pattern (CSSP) [17], Common Sparse Spectral Spatial Pattern (CSSSP) [18], Sub-Band Common Spatial Pattern (SBCSP) [15], Filter Bank Common Spatial Pattern (FBCSP) [19] have been proposed. The FBCSP, which presented the best performance on the BCI Competition IV dataset II-a, extracts CSP features with a filter bank, which is composed of multiple non-overlapping frequency bands with an equal bandwidth, to determine subject-specific optimal frequency bands. Later, adaptive FBCSP [20] and Discriminative Common Spatial Pattern (DCSP) [21] were proposed extending the FBCSP for the construction of a subject-specific filter bank. Wang and Zheng proposed a Local Temporal Common Spatial Patten (LTCSP) [22] that takes into account local temporal relations among signals in the estimation of covariance matrices of the CSP algorithm. This overcomes the limitation of the CSP that neglects temporal information in EEG signals.

Some research groups have devoted their efforts for time–frequency analysis of EEG signals. In order to reflect the temporal variation of neuronal signals, Qin et al. [23], Deng et al. [24], and Zhou et al. [25] devised, respectively, their own time–frequency weighting schemes. Ince et al. proposed a method using adaptive time–frequency segmentation to select discriminative time–frequency features [26,27]. More recently, other groups have tried classifying EEG signals with simultaneous analysis on the space–time–frequency domain based on machine learning techniques such as principal component analysis [28] or parallel factor analysis [29,30], etc. [31].

All these methods, however, assumed that the spatial filters of EEG signals were homogeneous during motor imagery by constructing spatial filters with fixed bases for whole time range of the signals. The bases can be most appropriate for whole time range of the signals, but it does not be guaranteed for segmented time range because of non-stationary features of the signals [10–12]. In this paper, we propose a novel method of finding non-homogeneous spatial filters over both the time and frequency domains. Unlike previous methods, it optimizes individual spatial filters to each decomposed time–frequency segment by selecting different sets of discriminative bases. By constructing spatial filters with an individual set of bases for each segmented time range, the non-stationary features of signals can be managed.

3. Proposed method

In this section, we describe a novel method to learn time–frequency varying spatial filters, which we call 'non-homogeneous spatial filters.' The main idea of our method is to decompose a single-trial EEG into space–time–frequency components and to find class-discriminative spatial filters, one for each pair of time and frequency components. The rationale for our approach is that brain activities are non-stationary in nature [10–12] and ERD/ERS patterns during motor imagery are not necessarily evoked in an identical or similar time segment [22,23,28]. Therefore, we need to consider those characteristics in finding the optimal spatial filters.

The technical challenges of the proposed method are how to select discriminative bases of each filter and how to design classifiers based on the filters. Instead of constructing a homogeneous filter which has fixed bases for whole time range, the proposed method constructs non-homogenous filters which have different sets of bases for each segmented time range. The conventional CSP algorithm guarantees the fixed bases are discriminative for whole time range of signals, but does not guarantee it for segmented time range. The proposed method solves the problems by adopting the Fisher ratio which can determine the discrimination level of each basis. It selects discriminative bases of each filter based on a ratio of between-class variance and within-class variance. Also, the proposed method of Fisher ratio-based weighted classification manages the level of discrimination of each filter in the same framework.

A schematic diagram of the proposed method is illustrated in Fig. 1. We first decompose a multi-channel EEG into the tensor of space–time–frequency by applying filter bank-based spectral filtering, CSP-based spatial filtering, and time-domain segmentation in a row. Then, the proposed non-homogeneous spatial filter optimization is performed for each pair of time and frequency bands by means of a Fisher criterion. The features extracted from the spatially filtered signals are used to train classifiers. In the evaluation, the class label is determined based on a weighted sum of outputs from classifiers. We should note that the CSP algorithm and classifier training is based on the one-against-one approach.

3.1. Space–time–frequency representation

Motivated by Ang et al.'s work [19], we employ a filter bank, $S=\{s_1,s_2,\ldots,s_B\}$, that bandpass-filters a single-trial EEG of the c-th class $X_c \in \mathbb{R}^{N\times T}$, where N and T denote, respectively, the number of channels and the number of sample points. Similar to the previous work, the predefined and fixed multiple frequency bands are used as follows:

$$V_c^b = h^b \otimes X_c \tag{1}$$

where h^b denotes a spectral filter of the b-th frequency band in a filter bank, $\otimes$ denotes a convolution operator, $b \in \{1,2,\ldots,B\}$, B is the number of frequency bands of interest, and V_c^b represents the bandpass-filtered single-trial EEG of the c-th class.

We then spatially transform the bandpass-filtered signal V_c^b

$$Z_c^b = W^b V_c^b \tag{2}$$

where $W^b \in \mathbb{R}^{N\times N}$ denotes a spatial filter optimized from the b-th bandpass-filtered signals with the conventional CSP algorithm [8]. The rows of W^b denote stationary spatial filters. By simple manipulation, we can obtain

$$V_c^b = [W^b]^{-1} Z_c^b \tag{3}$$

In Eq. (3), the columns of $[W^b]^{-1}$ are time-invariant modes and these are called common spatial patterns [8]. Meanwhile, the columns of Z_c^b can be considered as the basis producing the measured single-trial EEG at the time-points.

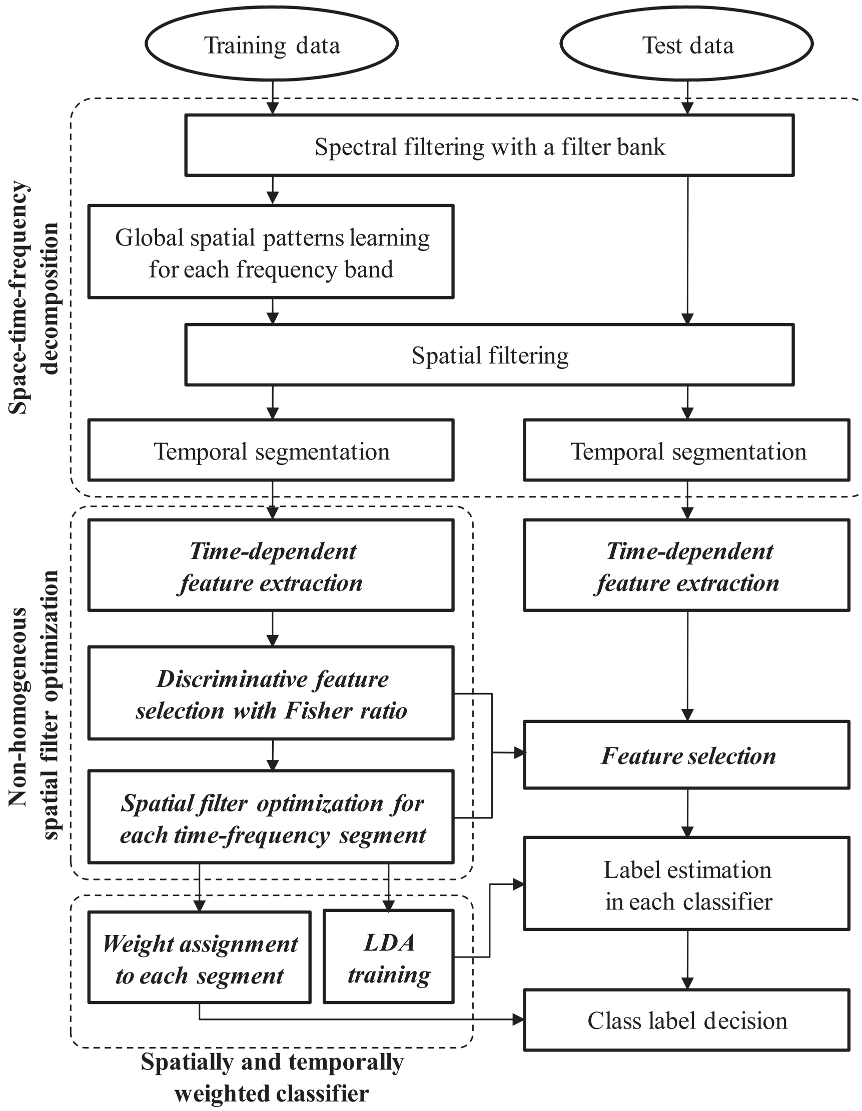

Fig. 1. A schematic diagram of the proposed method of non-homogeneous spatial filter optimization for multi-class motor imagery classification.

When it is considered for the nature of the non-stationary feature of human brain activity, it is believed that classifying a single-trial EEG with the features extracted from long-time samples is unreasonable. Thus, we dissect the spatially filtered single-trial EEG into overlapping multiple segments in time and assume that the electrical activity is stationary in a short-time window, *i.e.*, one time segment, but the activities of different windows may not be. We impose to overlap consecutive time segments, because it is unknown when the activity changes in time.

We then convert the single-trial EEG X_c into the tensor of space–time–frequency $\mathcal{Z}_c \in \mathbb{R}^{N\times K\times B}$, where N, K, and B denote, respectively, the number of channels, the number of time segments in a trial, and the number of frequency bands in a filter bank. Fig. 2 illustrates an example of transforming a single-trial EEG to a space–time–frequency tensor.

3.2. Non-homogeneous spatial filter optimization

Here, we propose a novel method of finding the discriminative spatial filters for each time–frequency segment individually by means of the Fisher ratio. We first compute the second-order statistics of the samples, which is a common approach of extracting features in the motor imagery-based BCI [8,13,14]. The variance of each space–time–frequency segment, $Y_c^{i,k,b}$, is computed as follows:

$$Y_c^{i,k,b} = V_c[Z_c^{i,k,b}] \tag{4}$$

where c is a class label, $V_c[\cdot]$ denotes a variance of the samples, and i, k, and b denote, respectively, an index of EEG channel, time segment, and frequency band. The spatially filtered signal $Z_c^{i,k,b}$ is obtained by Eq. (2) with a full rank matrix W^b, as explained above, deferring the selection of discriminative spatial filters. We compute the discriminative power of the features between two classes for each segment with the Fisher ratio as follows:

$$F^{i,k,b} = \frac{\sum_c D_c(\mu_c[\mathbf{Y}^{i,k,b}] - \mu[\mathbf{Y}^{i,k,b}])^2/(C-1)}{\sum_c V_c[\mathbf{Y}^{i,k,b}]/(D-C)} \tag{5}$$

where $\mathbf{Y}^{i,k,b}$ denotes the set of features extracted from the signals of the k-th time segment of the i-th channel in the b-th frequency band, $\mu_c[\mathbf{Y}^{i,k,b}]$ denotes the mean of the features of the c-th class trials, $\mu[\mathbf{Y}^{i,k,b}]$ denotes the mean of the features of the trials for all classes, D_c is the number of trials belonging to the class c,

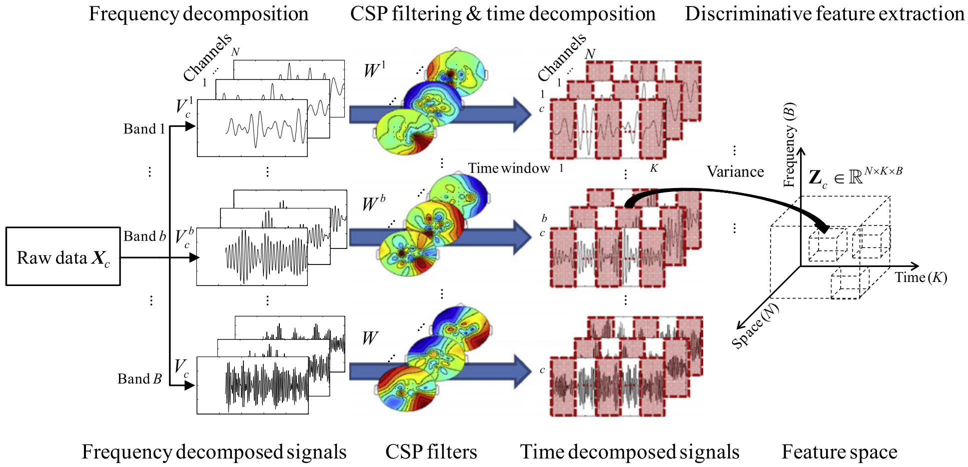

Fig. 2. An example of transforming a single trial EEG to a space–time–frequency tensor.

$D=\sum_c D_c$ is the total number of trials in the training dataset, and C represents the number of classes of interest.

We exploit the Fisher criterion defined in Eq. (5) for a metric that measures how strongly a feature is correlated with the corresponding class labels [32]. The features of a high Fisher ratio mean that they are class-discriminative, possibly improving classification performance. Therefore, it is natural to select the spatial filters of the high Fisher ratio in W^b and use them for feature extraction. In this paper, we select a set of discriminative spatial filters for each time–frequency segment with the following rule:

$$\mathbb{S}^{k,b}=\bigcup_i (F^{i,k,b}>\sigma) \tag{6}$$

where $\mathbb{S}$ denotes the set of indices of the selected row vectors in W^b, *i.e.*, spatial filters, k and b denote, respectively, an index of the time segment and the frequency band, i is the index of the row of W^b, and σ denotes a threshold that, in this paper, we determine based on the statistical significance with a F-test, *i.e.*, p-value ($p<0.01$). Based on $\mathbb{S}^{k,b}$, we compose different spatial filters $\hat{W}^{k,b}$ for a time(k)–frequency(b) segment. The process of non-homogeneous spatial filter selection across time–frequency segments is what distinguishes our method from other methods in the literature.

One may think of applying the CSP algorithm for each segment individually instead of applying it over all time-points and selecting discriminative spatial filters for each segment. But the assumption we used in our approach is that though evoked bases of common spatial patterns reflect the neurophysiological characteristics of changes in the human brain, the kinds of bases are coherent over time. Hence, our approach can be considered to select discriminative features from non-stationary brain activities, although we still do not know where the bases originate in the volumetric brain.

3.3. Feature extraction and classification

We compose a feature vector, following the common approach in the literature [8,13,14], with the logarithmic values of the normalized variances of the samples transformed by the selected spatial filters as follows:

$$\mathbf{f}_c=\left[\log\left(\frac{\hat{Y}_c^{i,k,b}}{\sum_{j=1}^{M}\hat{Y}_c^{j,k,b}}\right)\right]_{i\in\{1,2,\ldots,M\}} \tag{7}$$

where $\mathbf{f}_c$ is a feature vector of the single-trial EEG of the c-th class, $\hat{Y}_c^{i,k,b}$ denotes the variance of the samples filtered by the non-homogeneous spatial filter $\hat{W}^{k,b}$ in the segment of the tuple of the i-th channel, k-th time window, and b-th frequency band, and M is the number of row vectors in $\hat{W}^{k,b}$, which eventually determines the dimension of the feature vector $\mathbf{f}_c$.

We use Linear Discriminant Analysis (LDA) and Support Vector Machine (SVM), which are the most widely considered in BCI due to their simplicity. We apply them to each time–frequency segment individually. In order to overcome the problem of intra-subject variability of ERD/ERS patterns in a motor imagery task, we consider multiple frequency components and multiple time segments simultaneously in class label decision. We assign a weight to each time(k)–frequency(b) segment with the sum of Fisher scores of the selected non-homogeneous spatial filters $\omega(k,b)=\sum_{i\in\mathbb{S}^{k,b}}F^{i,k,b}$ and decide the class label l of the single-trial EEG with the following rule

$$l=\begin{cases}1 & \text{if } \sum_b^B\sum_k^K \omega(k,b)o(k,b)\geq 0\\ -1 & \text{otherwise}\end{cases} \tag{8}$$

where B is the number of frequency bands, K is the number of time segments, and $o(k,b)\in\{+1,-1\}$ denotes the output class label from the classifier of the (k, b)-th segment. This strategy of deciding the class label can be considered as ‘spectrally and temporally weighted classification’ and it is one of the contributions of this work.

4. Experimental results and analysis

In this section, we present feasibility and the effectiveness of the proposed method by comparing the classification performance with other competing methods in the literature and analyzing the experimental results on a publicly available dataset.

4.1. EEG dataset and preprocessing

We investigate the BCI Competition IV dataset II-a[1] that contains EEG signals recorded from 9 subjects performing 4 different motor imagery tasks: left-hand, right-hand, foot, and tongue. The dataset is comprised of two sessions conducted on

[1] Available at http://www.bbci.de/competition/iv/.

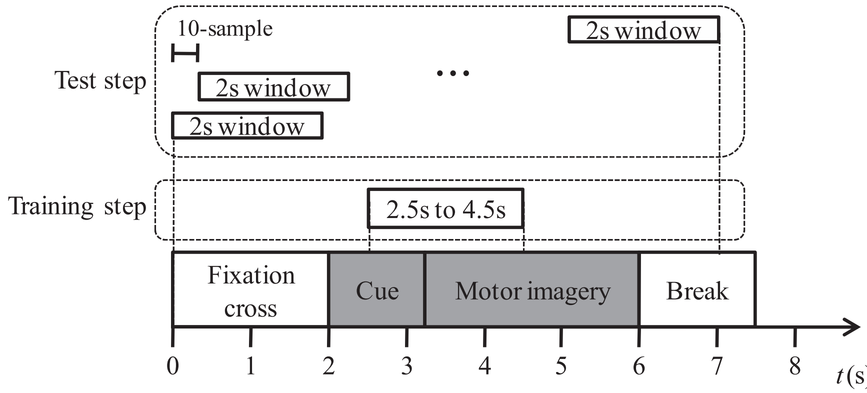

Fig. 3. Time scheme of the motor imagery EEG dataset and samples used for training and test.

different days. Each session includes 6 runs separated by short break. There are 12 trials for each of the 4 tasks in a run. Each run is further composed of 48 trials and one session consists of 288 trials totally. The EEG data were acquired using 22 electrodes and 3 EOG channels that were positioned according to the international 10–20 system [33]. The signals were sampled at 250 Hz and bandpass-filtered between 0.5 Hz and 100 Hz. An additional 50 Hz notch filter was enabled to suppress line noise. Refer to [34] for the details of the dataset.

The selection of a time interval for the motor imagery classification is an important factor that has a great impact on the classification performance. In this paper, following Ang et al.'s work [19], we extract features from the signals of all electrodes except the electrooculograms (EOGs) between 0.5 s and 2.5 s after onset of the visual cue in the training step. As the evaluation is performed by using continuous classification output for each sample [34], the class label is evaluated on every 10-th sample by using a sliding window of length 2 s (called an 'outer-window') by overlapping 240 samples between two consecutive windows from onset of the visual cue. Fig. 3 shows the time scheme of the experimental paradigm and the EEG signals used for training and test, respectively.

We employ a filter bank approach covering 4–40 Hz with 4 Hz bandwidth for each frequency band, $B=9$ frequency bands in total. As the number of channels is 22, we use 22 bases of CSP filters W^b where $b \in \{1,2,\ldots,B\}$. Within the outer-window, we further segment the single-trial EEG in time with an 'inner-window' of size 100-, 200-, and 300-samples in order to construct a space–time–frequency tensor. Because consecutive inner-windows are 50% overlapping, we obtain 9, 4, and 2 time segments, i.e., $K=9$, 4, 2, respectively, in an outer-window.

We also investigate the BCI Competition II dataset IV[2] that contains EEG signals for self-paced left- and right-hand motor imagery tasks. It consists of 316 trials for training and 100 trials for test. The EEG data were acquired from a normal subject using 28 electrodes according to the international 10–20 system [33]. The EEG data were recorded at 1000 Hz with a band-pass filter between 0.05 and 200 Hz but they were downsampled at 100 Hz for comparison experiments with [22]. For the data set, we also employ a filter bank approach covering 4–40 Hz with 4 Hz bandwidth for each frequency band. We segment the single-trial EEG in time by a window of size 6-samples with a third overlapping in order to construct the space–time–frequency tensor.

[2] Available at http://www.bbci.de/competition/ii/.

4.2. Performance comparisons with other methods

Table 1 shows the performance of the proposed method with LDA and SVM classification models. We use a metric of Cohen's Kappa value that measures the agreement between two estimators [35]. The LDA model shows higher performance than the SVM model but there was no significant difference ($p=0.3990$) in the statistical analysis, paired T-test.

We compare the performance of the proposed method to that of two other competing methods, Ang et al.'s Filter Bank Common Spatial Pattern (FBCSP) [19] and Time Segmented FBCSP (TS-FBCSP) based on the LDA classification model with the BCI Competition IV dataset II-a. The FBCSP selects discriminative frequency bands from filter banks and constructs CSP filters for the frequency bands. The TS-FBCSP, the intermediate method between the FBCSP and the proposed method, is variation of the FBCSP applying our idea of dissecting the time domain into multiple segments and then optimizing homogeneous spatial filters for each time segment individually. Because the performance of the CSP is highly dependent on the number of basis pairs of CSP filters [8], we perform spatial filters optimization using the values of 2, 3, and 4 for the methods. The main difference between the proposed method and FBCSP/TS-FBCSP is the way of finding optimal spatial filters; homogeneity and nonhomogeneity. Instead of constructing a filter which has fixed bases for whole time range, the proposed method constructs filters which have different sets of bases for each segmented time range based on Fisher ratio.

The goal of the BCI Competition IV with dataset II-a was evaluation of classification algorithms based on the session-to-session transfer rate using the session 1 and session 2 dataset for training and evaluation respectively. We use a metric of Cohen's Kappa value that measures the agreement between two estimators [35]. As stated, there are multiple class labels, one for each outer-window. The classification accuracies obtained from every outer-window used for the three competing methods are illustrated in Fig. 4. The results were obtained with 2 pairs of bases in a CSP filter and LDA. The figure clearly shows that the maximum classification accuracies are obtained by the outer-windows positioned between 4 s and 5 s and there is a gradual decrease thereafter. In Table 2, we also summarize the best performance of the three competing methods from the time course of the Kappa values in Fig. 4. The proposed method outperforms the competing methods. The highest mean accuracy of 0.60 represents the best performances for 5 subjects. The statistical analysis shows significant difference between the previous FBCSP and the proposed method with $p=0.0167$ for $m=2$, $p=0.0066$ for $m=3$, and $p=0.0032$ for $m=4$ in paired T-test. For all the cases, the proposed method outperformed FBCSP in a 95% confidence level.

In order to see how the size of inner-window affects the classification performance of the proposed method, we summarize the best performance of 100-, 200-, and 300-samples inner-window in Fig. 5. The mean Kappa values are 0.60, 0.59, and 0.56, respectively. As the proposed method constructs non-homogeneous filters for each inner-window, the non-homogeneity of the proposed method is reduced when the size of inner-window becomes larger.

Table 3 shows the performance comparison of the proposed method with those of the winners of the competition. The proposed method represents the best performance in 6 subjects, marked in boldface with the highest mean accuracy. For reference, we briefly describe their methods. The 1st winner of the competition applied the FBCSP and selected discriminative frequency bands by means of the mutual information between class labels and feature vectors [19]. They applied homogeneous spatial filters in feature extraction. The 2nd winner applied the CSP on broadband-filtered (8–30 Hz) signals and used LDA and a naive Bayesian classifier with features composed of log variance of the spatially filtered signals. Unlike other competitors, the 3rd winner applied a channel selection method by recursively removing less discriminative one, and built a classifier with three Support Vector Machines (SVM). The statistical analysis shows significant difference between the proposed method and the 2nd and 3rd competition methods with $p = 0.0083$, and $p < 0.0001$, respectively in paired T-test. The proposed method outperformed 2nd and 3rd winners methods in a 99% confidence level.

Table 1
Classification performance of the proposed method with LDA and SVM classification models. The numbers represent Cohens Kappa values.

	Proposed method	
	LDA	SVM
Subject 1	**0.74**	**0.74**
Subject 2	**0.35**	0.34
Subject 3	**0.76**	**0.76**
Subject 4	**0.53**	**0.53**
Subject 5	**0.38**	0.34
Subject 6	**0.31**	**0.31**
Subject 7	0.84	**0.86**
Subject 8	**0.74**	0.72
Subject 9	0.74	**0.75**
Mean	**0.60**	0.59
(Std)	(0.21)	(0.22)
p-Value		0.3990

Table 2
Classification performances of the three competing methods, *i.e.*, Ang et al.'s Filter Bank Common Spatial Pattern (FBCSP), Time Segmented FBCSP (TS-FBCSP), and the proposed method. The numbers represent Cohen's Kappa values and m denotes the number of basis pairs.

	FBCSP			TS-FBCSP			Proposed method with LDA
	$m=2$	$m=3$	$m=4$	$m=2$	$m=3$	$m=4$	
Subject 1	0.71	0.71	0.72	**0.81**	0.76	0.72	0.74
Subject 2	**0.38**	**0.38**	0.34	0.34	0.31	0.30	0.35
Subject 3	0.70	0.71	0.71	0.73	0.73	0.74	**0.76**
Subject 4	0.52	0.50	0.46	0.51	0.52	0.50	**0.53**
Subject 5	0.33	0.35	0.39	0.35	0.39	**0.41**	0.38
Subject 6	0.29	0.22	0.23	**0.34**	0.33	0.32	0.31
Subject 7	0.77	0.77	0.78	0.77	0.78	0.80	**0.84**
Subject 8	0.67	0.67	0.64	0.69	0.69	0.69	**0.74**
Subject 9	0.61	0.63	0.66	0.70	0.69	0.70	**0.74**
Mean	0.55	0.55	0.55	0.58	0.58	0.58	**0.60**
(Std)	(0.18)	(0.19)	(0.20)	(0.20)	(0.19)	(0.19)	(0.21)
p-Value	**0.0167**	**0.0066**	**0.0032**	0.2755	0.0816	**0.0339**	

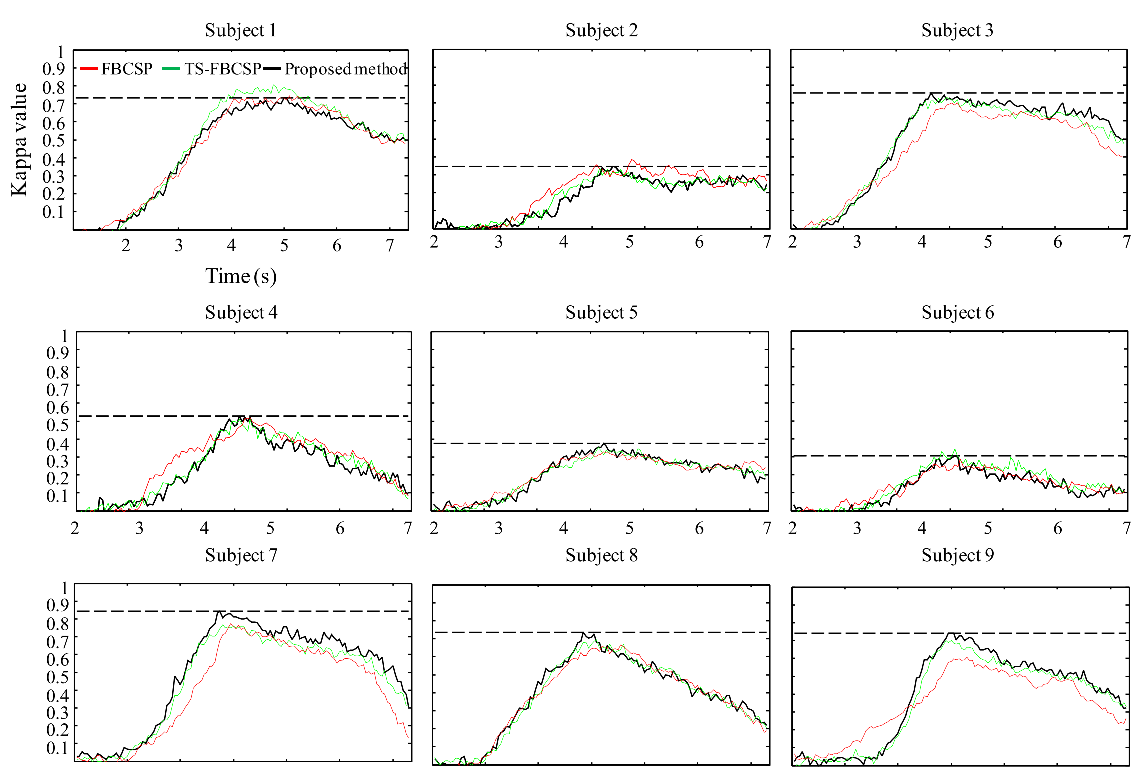

Fig. 4. Time course of the Kappa values for classification performance of the test data based on FBCSP, TS-FBCSP, and the proposed method. The dotted line denotes the maximum performance of the proposed method.

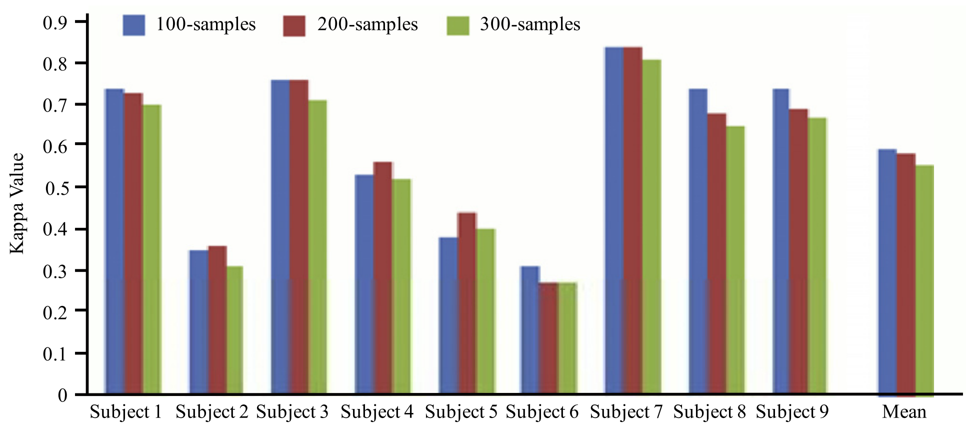

Fig. 5. Kappa values for classification performance of the test data with 100-, 200-, and 300-samples inner-window.

Table 3
Performance comparison of the proposed method with the winners of the BCI Competition IV Dataset II-a. The numbers represent Cohen's Kappa values.

	Proposed method	BCI competition VI dataset II-a		
	with LDA	1st winner	2nd winner	3rd winner
Subject 1	**0.74**	0.68	0.69	0.38
Subject 2	0.35	**0.42**	0.34	0.18
Subject 3	**0.76**	0.75	0.71	0.48
Subject 4	**0.53**	0.48	0.44	0.33
Subject 5	0.38	**0.40**	0.16	0.07
Subject 6	**0.31**	0.27	0.21	0.14
Subject 7	**0.84**	0.77	0.66	0.29
Subject 8	0.74	**0.75**	0.73	0.49
Subject 9	**0.74**	0.61	0.69	0.44
Mean (Std)	**0.60**	0.57	0.52	0.31
	(0.21)	(0.18)	(0.23)	(0.15)
p-Value		0.1775	**0.0083**	**0.0001**

Table 4
Classification performances of the four competing methods, Common Spatial Pattern (CSP), Discriminative CSP (DCSP), Local Temporal CSP (LTCSP), and the proposed method. The numbers represent percentage accuracies and m denotes the number of basis pairs.

	CSP	DCSP	LTCSP	Proposed method
$m=1$	71	73	88	88
$m=2$	71	70	84	
$m=3$	69	69	83	
$m=4$	63	66	82	
$m=5$	58	64	83	

With the BCI Competition II dataset IV, we compare the performance of the proposed method to that of other previous methods, Common Spatial Pattern (CSP), Discriminative CSP (DCSP) [21], and Local Temporal CSP (LTCSP) [22], in the literature. For the comparison, we use LDA in the experiments. Table 4 shows performance of the four competing methods with the percentage accuracy. In the experiment, the proposed method represents the best performance of 88% accuracy.

4.3. Analysis of the weight distribution in time–frequency segments

In this section, we analyze the weights distributed over time–frequency segments to present the effectiveness of the proposed method with the first dataset. Fig. 6 shows the normalized weights of the time–frequency segments for left- and right-hand motor imagery classification. In the figure, the darker the face color of the time–frequency segment, the higher its weight. Based on the figure, we can confirm the inter-subject variability of the motor imagery tasks. Another important feature is that the proposed method can represent the change of discriminative power between classes in time. This important fact is reflected by assigning a weight to each time–frequency segment and this approach eventually improves the performance.

4.4. Effects of spatial filter selection

In this section, we analyze the effect of spatial filter selection in time–frequency segments with the first dataset. In Figs. 7 and 8, we illustrate the weight distributions in terms of the Fisher ratios of the spatial filters over two dominant frequency bands for Subject 1 and Subject 8, respectively. Although many of the high Fisher ratios are distributed at the first and last m couples of spatial filters, this is the explanation for why the conventional approach of choosing the first and last m column vectors in spatial filters works. However, it is not always true across time–frequency segments and subjects. This indicates that the common approach of selecting pairs of first and last few spatial filters may not be optimal. Accordingly, the proposed method of selecting discriminative spatial filters adaptively in each time–frequency segment showed better classification performance. For example, many of the high weights for Subject 8 in Fig. 8 are distributed across the spatial filters. The proposed method captures this fact and selects the discriminative spatial filters accordingly and thus resulted in higher performances compared to TS-FBCSP, an improvement of as much as 0.05 for Subject 8, as shown in Table 2.

In contrast, the performance for Subject 1 is decreased by 0.08, 0.03, and 0.01 compared with the performances of TS-FBCSP with $m=2$, $m=3$, and $m=4$. We summarize that this is the result of selecting spatial filters of relatively low weights but still larger than the threshold σ in Eq. (6) (see Fig. 7). Hence, determining the optimal threshold in Eq. (6) is another problem that should be considered in our future work.

Fig. 9 shows the selected spatial filters and the change of their weights over time for the classification of left- and right-hand motor imagery tasks for Subject 8. The figure shows that the selected spatial filters are highly related to left- and right-hand motor imageries, because they are contralaterally well localized in the left- and right-sensorimotor cortex. When seeing change of weights of each basis over time, discriminative sets of bases and their weights are highly time-varying. As the proposed method constructs non-homogeneous spatial filters considering those non-stationary characteristics of ERD/ERS, it is superior to the other competing methods.

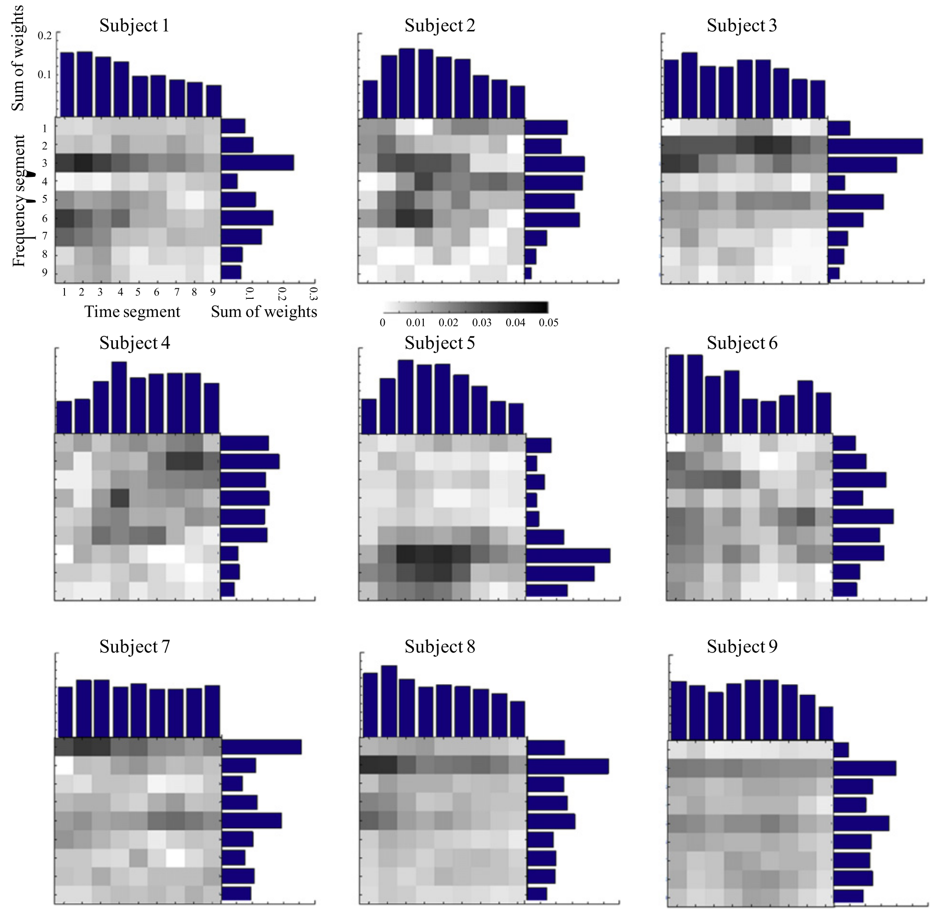

Fig. 6. Normalized weights of time–frequency segments and its distribution on time–frequency domain for left- and right-hand motor imagery tasks.

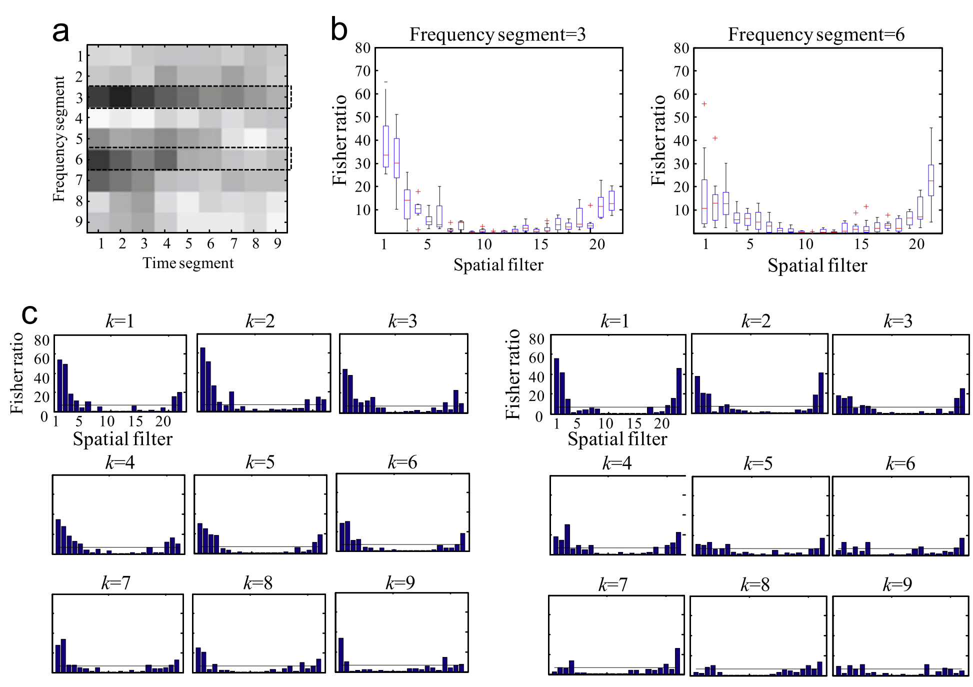

Fig. 7. Weight distributions of the spatial filters over time segments at the most discriminative frequency bands for Subject 1. (a) Weights of time–frequency segments. (b) Weight distribution of spatial filters. (c) Weight of components of spatial filters in each time segment k in the dominant frequency band.

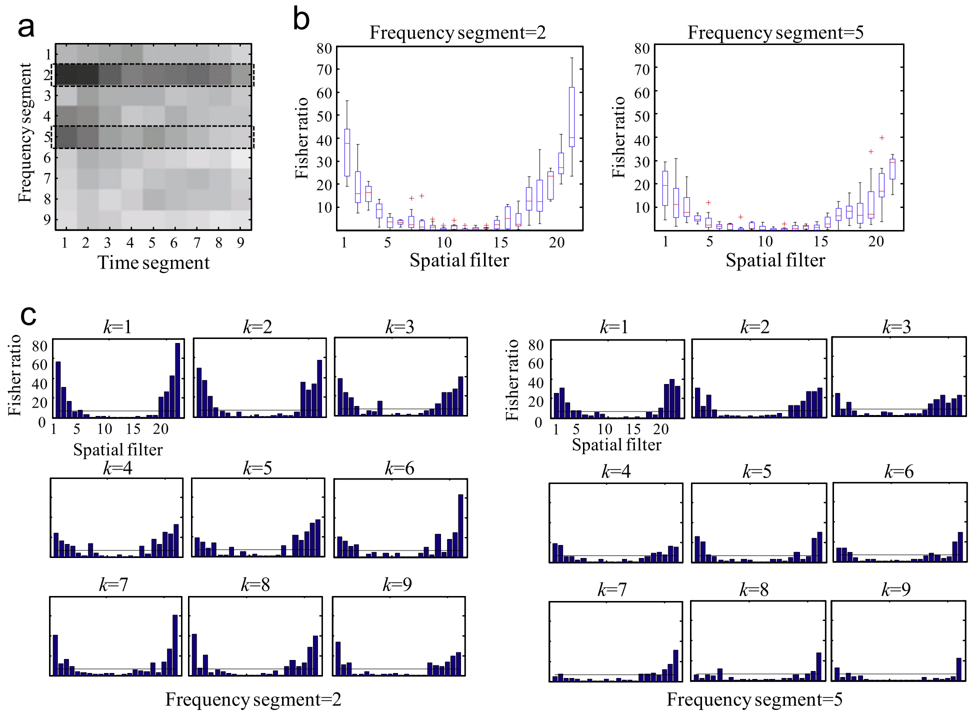

Fig. 8. Weight distributions of the spatial filters over time segments at the most discriminative frequency bands for Subject 8. (a) Weights of time–frequency segments. (b) Weight distribution of spatial filters. (c) Weight of components of spatial filters in each time segment k in the dominant frequency band.

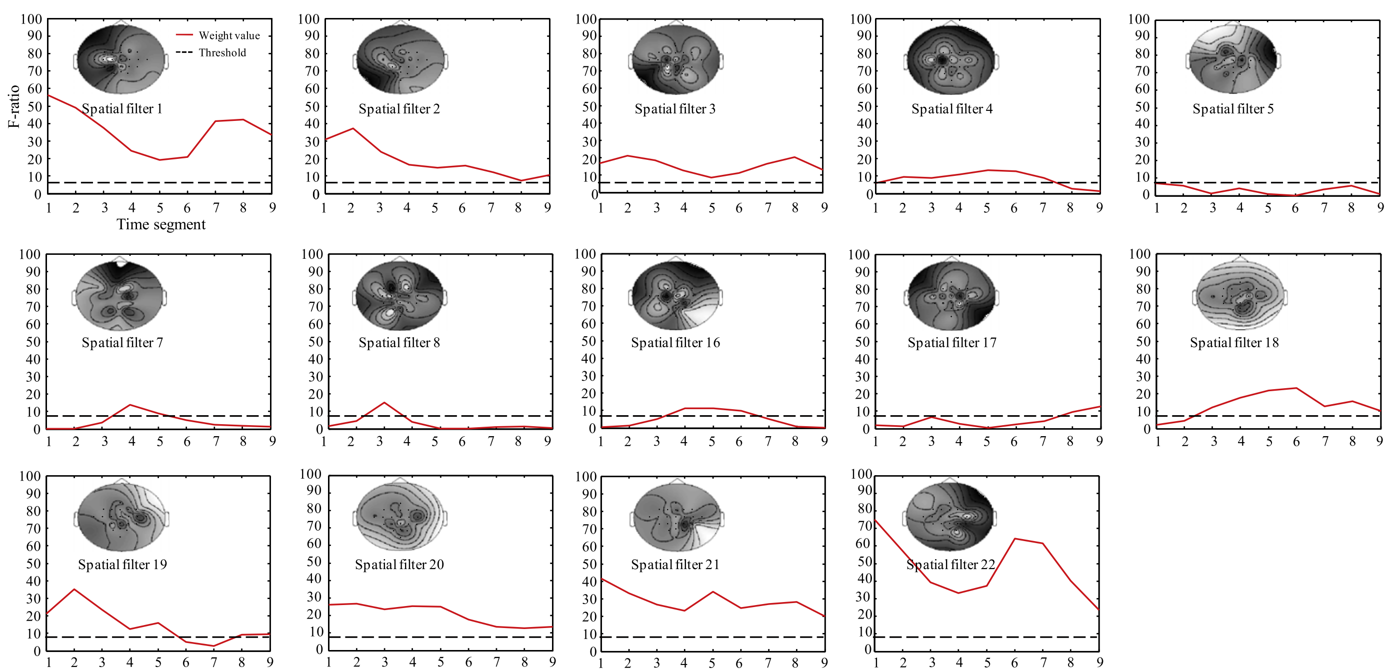

Fig. 9. The selected spatial filters and the change of their weights over time for left- and right-hand motor imagery tasks for Subject 8. The dotted line is a threshold used for selecting spatial filters.

5. Conclusions and further research

The ERD/ERS patterns generated by motor imagery have high intra- and inter-subject variability in the space, time, and frequency domains. Many research groups have assumed that the spatial distribution of the single-trial EEG was homogeneous and applied spatial filters homogeneously across all the time. Unlike previous methods in the literature, we proposed a novel method of finding non-homogeneous spatial filters in time–frequency segments. The method optimized individual spatial filters over time with different sets of discriminative bases. We designed the strategy of the spectrally and temporally weighted classification rule to decide a class label. By the weighted classifiers based on our non-homogeneous spatial filters, we considered non-stationarity of EEG signals and achieved performance improvement for motor imagery classification. Based on our experimental results on the BCI Competition IV dataset II-a and BCI Competition II dataset IV, the effectiveness of the proposed method is clear and it outperforms the other competing methods.

While the proposed method enhanced the performance, it used a filter bank with a fixed bandwidth dissecting a wide band of interest into smaller ones across subjects without considering spectral variation in task relevant brain responses. Therefore, it will be our forthcoming research issue.

Acknowledgments

This treatise was by National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) under Grant 2012-005741 and the project of Global Ph.D. Fellowship.

References

[1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, T. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.

[2] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain–computer communication, Proc. IEEE 89 (7) (2001) 1123–1134.

[3] G. Pfurtscheller, C. Neuper, D. Flotzinger, M. Pregenzer, EEG-based discrimination between imagination of right and left hand movement, Electroencephalogr. Clin. Neurophysiol. 103 (6) (1997) 642–651.

[4] G. Pfurtscheller, A. Berghold, Patterns of cortical activation during planning of voluntary movement, Electroencephalogr. Clin. Neurophysiol. 72 (3) (1989) 250–258.

[5] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta, M. Hallett, Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG, Electroencephalogr. Clin. Neurophysiol. 93 (5) (1994) 380–389.

[6] P. Derambure, L. Defebvre, K. Dujardin, J. Bourriez, J. Jacquesson, A. Destee, J. Guieu, Effect of aging on the spatio-temporal pattern of event-related desynchronization during a voluntary movement, Electroencephalogr. Clin. Neurophysiol. 89 (3) (1993) 197–203.

[7] D. McFarland, L. McCane, S. David, J. Wolpaw, Spatial filter selection for EEG-based communication, Electroencephalogr. Clin. Neurophysiol. 103 (3) (1997) 386–394.

[8] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, Clin. Neurophysiol. 110 (5) (1999) 787–798.

[9] C. Brunner, M. Naeem, R. Leeb, B. Graimann, G. Pfurtscheller, Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis, Pattern Recognition Lett. 28 (8) (2007) 957–964.

[10] W. Truccolo, M. Ding, K. Knuth, R. Nakamura, S. Bressler, Trial-to-trial variability of cortical evoked responses: implications for analysis of functional connectivity, Clin. Neurophysiol. 113 (2) (2002) 206–226.

[11] A. Kaplan, A. Fingelkurts, A. Fingelkurts, S. Borisov, B. Darkhovsky, Nonstationary nature of the brain activity as revealed by EEG/MEG: methodological, practical and conceptual challenges, Signal Process. 85 (11) (2005) 2190–2212.

[12] M. Tarvainen, J. Hiltunen, P. Ranta-aho, P. Karjalainen, Estimation of nonstationary EEG with Kalman smoother approach: an application to event-related synchronization (ERS), IEEE Trans. Biomed. Eng. 51 (3) (2004) 516–524.

[13] C. Guger, H. Ramoser, G. Pfurtscheller, Real time EEG analysis with subject specific spatial patterns for a brain–computer interface, IEEE Trans. Rehabil. Eng. 8 (4) (2000) 447–456.

[14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, J. Neural Eng. 4 (2) (2007) R1–R13.

[15] Q. Novi, C. Guan, T. Dat, P. Xue, Sub-band common spatial pattern (SBCSP) for brain–computer interface, in: Proceedings of the Third International IEEE EMBS Conference on Neural Engineering, Kohala Coast, Hawaii, USA, May 2–5, 2007, pp. 517–207.

[16] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. Müller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Trans. Signal Process. Mag. 25 (1) (2008) 44–56.

[17] S. Lemm, B. Blankertz, G. Curio, K. Müller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE Trans. Biomed. Eng. 52 (9) (2005) 1541–1548.

[18] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K. Müller, Combined optimization of spatial and temporal filters for improving brain–computer interface, IEEE Trans. Biomed. Eng. 53 (1) (2006) 2274–2281.

[19] K. Ang, Z. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain–computer interface, in: Proceedings of the IEEE International Joint Conference on Neural Networks 2008, Hong Kong, June 1–6, 2008, pp. 2390–2397.

[20] P. Kavitha, C. Guan, C. Lau, A. Vinod, An adaptive filter bank for MI based brain–computer interface, in: Proceedings of the 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada, August 20–24, 2008, pp. 1104–1107.

[21] K. Thomas, C. Guan, C. Lau, A. Vinod, K. Ang, A new discriminative common spatial pattern method for motor imagery brain–computer interfaces, IEEE Trans. Biomed. Eng. 56 (11) (2009) 2730–2733.

[22] H. Wang, W. Zheng, Local temporal common spatial patterns for robust single-trial EEG classification, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (2) (2008) 131–139.

[23] L. Qin, B. He, A wavelet-based time–frequency analysis approach for classification of motor imagery for brain–computer interface applications, J. Neural Eng. 2 (4) (2005) 65–72.

[24] J. Deng, J. Yao, J. Dewald, Classification of the intention to generate a shoulder versus elbow torque by means of a time–frequency synthesized spatial patterns BCI algorithm, J. Neural Eng. 2 (4) (2005) 131–138.

[25] J. Zhou, J. Yao, J. Deng, J. Dewald, EEG-based classification for elbow versus shoulder torque intentions involving stroke subjects, Comput. Biol. Med. 39 (5) (2009) 443–452.

[26] N. Ince, S. Arica, A. Tewfik, Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary time–frequency tilings, J. Neural Eng. 3 (3) (2006) 235–244.

[27] N. Ince, A. Tewfik, S. Arica, Extraction subject-specific motor imagery time–frequency patterns for single trial EEG classification, Comput. Biol. Med. 37 (4) (2007) 499–508.

[28] T. Ferree, M. Brier, J. Hart, M. Kraut, Space–time–frequency analysis of EEG data using within-subject statistical tests followed by sequential PCA, NeuroImage 45 (1) (2009) 109–121.

[29] M. Weis, F. Roemer, M. Haardt, D. Jannek, P. Husar, Multi-dimensional space–time–frequency component analysis of event related EEG data using closed-form PARAFAC, in: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing 2009, Taipei, Taiwan, April 19–24, 2009, pp. 349–352.

[30] M. Weis, D. Jannek, F. Roemer, M. Haardt, T. Guenther, M. Haardt, P. Husar, Multi-dimensional PARAFAC2 component analysis of multi-channel EEG data including temporal tracking, in: Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, August 31–September 4, 2010, pp. 5375–5378.

[31] N. Ince, F. Goksu, A. Tewfik, S. Arica, Adapting subject specific motor imagery EEG patterns in space–time–frequency for a brain computer interface, Biomed. Signal Process. Control 4 (3) (2009) 236–246.

[32] C. Bishop, Neural Networks for Pattern Recognition, Oxford UP, Oxford, UK, 1995.

[33] H. Jasper, The ten-twenty electrode system of the international federation, Electroencephalogr. Clin. Neurophysiol. 10 (2) (1958) 371–375.

[34] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, G. Pfurtscheller, BCI Competition 2008—Graz Data Set A, provided by the Institute for Knowledge Discovery (Laboratory of Brain–Computer Interfaces), Graz University of Technology.

[35] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46.

Tae-Eui Kam received the B.S. and M.S. degrees in computer science and engineering from Korea University, Seoul, Korea, in 2009 and 2011, respectively. He is currently a Ph.D. candidate in the Department of Computer Science and Engineering in Korea University. His research interests include brain–computer interface and machine learning.

Heung-Il Suk received the B.S. and M.S. degrees in computer engineering from Pukyong National University, Busan, Korea, in 2004 and 2007, respectively, and the Ph.D. degree in computer science and engineering, Korea University, Seoul, Korea, in 2012. From 2004 to 2005, he was a Visiting Researcher at the Computer and Vision Research Center, University of Texas at Austin, USA. He received the Outstanding Paper Award at the Korea Computer Congress in 2007. He also received the Silver Award at the 18th Samsung Human-Tech Thesis Prize in 2012. His current research interests include machine learning, brain–computer interfaces, and neuroimaging analysis.

Seong-Whan Lee received the B.S. degree in computer science and statistics from Seoul National University, Seoul, Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology (KAIST), Seoul, Korea, in 1986 and 1989, respectively. He is currently the Hyundai Motor Chair Professor at Korea University, Seoul, where he is the Head of the Department of Brain and Cognitive Engineering and the Director of the Institute for Brain and Cognitive Engineering. He is the Principal Investigator of the World Class University (WCU) project on ''Brain and Cognitive Engineering'' research, which is funded by the Ministry of Education, Science and Technology of Korea. His current research interests include pattern recognition, computer vision, and brain informatics. He has more than 250 publications in international journals and conference proceedings, and authored ten books.

He was the winner of the Annual Best Student Paper Award of the Korea Information Science Society in 1986. He received the First Outstanding Young Researcher Award at the Second International Conference on Document Analysis and Recognition in 1993, and the First Distinguished Research Award from Chungbuk National University in 1994. He received the Outstanding Research Award from the Korea Information Science Society in 1996. He received the Lotfi Zadeh Best Paper Award at the IEEE International Conference on Machine Learning and Cybernetics in 2011. He also received the Scientist of the Month Award from the Ministry of Education, Science and Technology of Korea in 2012.

A Fellow of the IEEE, IAPR, and Korean Academy of Science and Technology, he has served several professional societies as a chairman or governing board member. He was the founding Co-Editor-in-Chief of the International Journal of Document Analysis and Recognition. He has been an Associate Editor of several international journals including Pattern Recognition, ACM Transactions on Applied Perception, IEEE Transactions on Affective Computing, Image and Vision Computing, International Journal of Pattern Recognition and Artificial Intelligence, and International Journal of Image and Graphics. He was a General or Program Chair of many international conferences and workshops and was also on the program committees of numerous conferences and workshops.