Scientia Iranica D (2013) 20(6), 2061–2072

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions D: Computer Science & Engineering and Electrical Engineering
www.scientiairanica.com

Extracting common spatial patterns from EEG time segments for classifying motor imagery classes in a Brain Computer Interface (BCI)

H. Ghaheri* **and A.R. Ahmadyfard**

Department of Electrical Engineering and Robotic, Shahrood University of Technology, Shahrood, Iran.

Received 13 September 2012; accepted 11 May 2013

KEYWORDS
Electroencephalogram (EEG);
Brain Computer Interface (BCI);
Motor imagery;
Common Spatial Pattern (CSP);
Temporal segmentation;
One-Versus-the Rest (OVR) method.

Abstract. Brain Computer Interface (BCI) is a system which straightly converts the acquired brain signals such as Electroencephalogram (EEG) to commands for controlling external devices. One of the most successful methods in BCI applications based on Motor Imagery is Common Spatial Pattern (CSP). In the existing CSP methods, common spatial filters are applied on whole EEG signal as one time segment for feature extraction. The fact that ERD/ERS events are not steady over time motivated us to break down EEG signal into a number of sub-segments in this study. I combine this sentence with next one: "We believe the importance of EEG channels varies for different time segments in classification, therefore we extract features from each time segment using the analysis of CSP method. In order to classify Motor Imagery EEG signals, we apply a LDA classifier based on OVR (One-Versus-the Rest) scheme on the extracted CSP features. The considered Motor Imagery consists of four classes: left hand, right hand, foot and tongue. We used dataset 2a of BCI competition IV to evaluate our method. The result of experiment shows that this method outperforms both CSP and the best competitor of the BCI competition IV.

1. Introduction

A Brain Computer Interface (BCI) allows severely disabled people to control external environment. Electroencephalogram (EEG)-based BCI translates brain electric signals into operative control commands. The most commonly used mental control strategy in BCI researches is the Motor Imagery (MI), because it can be associated with an attenuation or increase of localized neural rhythmic activity called Event-Related Desynchronization (ERD) or Event-Related Synchronization (ERS) [1,2].

Scalp-recorded EEG signals, in addition to ERD and ERS activities, is contaminated by signals from natural brain activities and artifacts such as electromyogram (EMG) and electrooculogram (EOG) [3]. In presence of these distortions and other outliers, the signal to noise ratio of EEG signals is low; therefore recognizing ERD/ERS events from EEG signals is difficult. Human skull has bone tissue, so signals generated in brain cortex interfere with each other before getting to scalp [4]. This means that the signals of EEG channels are sum of several sources generated under human skull. Therefore, selecting useful channels for classifying classes and extracting important events from selected EEG channels is difficult.

In order to extract discriminative features from EEG signals, a large number of signal processing and pattern recognition algorithms have been employed [5]. One of the most successful algorithms for channel reduction in MI-based BCI is Common Spatial Pattern (CSP) [6]. CSP is a decomposition method which first

*. *Corresponding author. Tel.: +98-09196675739*
E-mail addresses: habibeh.ghaheri@gmail.com (H. Ghaheri), ahmadyfard@shahroodut.ac.ir (A.R. Ahmadfard)

was proposed by Koles (1991) [7] for detecting some sort of mental diseases.

In CSP method, EEG signals acquired on top of the scalp are decomposed to two matrices. One of the matrices is the estimation of source signals generated under skull and the other matrix is the weights that show the importance of source signals in creation of EEG signals acquired on the scalp. The columns of weights matrix are called spatial filters [6]. In a two-class MI problem, CSP method calculates the spatial filters for maximizing the variance of signals between two classes of Motor Imagery. This set of spatial filters transform EEG signals to a new space in which the covariance matrices of transformed data corresponding to each class is diagonal. The extracted spatial filters are the eigenvectors corresponding to the largest and smallest eigenvalues of common covariance matrix. The result of filtering EEG signals, using the selected spatial filters, are used for classification of two classes. For the classification of a two-class Motor Imagery using CSP, the accuracy of above 90% has been reported [8].

Before CSP being applied, an appropriate band pass filter is needed to properly extract ERD/ERS events from EEG signals, so that the selection of frequency band affects the result of CSP method [8]. Although a wideband of 8-30Hz was suggested [8], evidence showed that selecting subject-specific frequency bands could yield an improvement in the recognition rate of MI-based BCI [9]. A problem with this preprocessing is that the center frequency of a proper filter varies from one subject to another, Therefore, several approaches, such as Common Spatio-Spectral Pattern (CSSP) [10] and Common Sparse Spectral Spatial Pattern (CSSSP) [11] are proposed to find the appropriate frequency bands for filtering each subject.

Recently, an alternative approach called Subband Common Spatial Pattern (SBCSP) was proposed, and has been shown to yield superior classification accuracy compared against CSSP and CSSSP on a publicly available dataset [12]. Ang et al. in 2008 proposed another approach called Filter Bank Common Spatial Patterns (FBCSP), and showed the FBCSP yields superior classification accuracy compared against SBCSP method [13]. FBCSP [13], which won dataset IIa and IIb in BCI competition IV (2008) [14], uses CSP features from a set of nine fixed bandpass filters and feature selection algorithm based on mutual information to effectively choose the subject-specific features. This selection process selects features from the relevant frequency components.

Considering that EEG signals for a subject are represented as $K \times N$ matrix $X = [x_1, x_2, \cdots, x_N]$ where K is the number of EEG channels and N is the number of samples for each channel. CSP method maps the samples of EEG signals to a new space based on covariance matrix of the EEG signals. This method does not consider local structure of EEG signals, as for each channel, it uses the whole EEG signal as one time segment for computing the covariance matrix. CSP method considers each time point of all channels as a vector in the feature space, and maps it into another space based on the average covariance matrix of all EEG signals. As a result, the noise on one small time slot affects the covariance matrix of EEG signals. This introduces error when special filters are estimated.

Most of CSP-based method, as mentioned above, does not consider temporally local structure of the time EEG signals, too [15]. Time segment used in CSP method is very important, so that adding or removing one time sample varies covariance matrix and consequently spatial filters.

Wang and Zheng proposed a method [15] in which local structure of EEG signals is considered by defining time-dependent neighboring matrix, called LTCSP (Local Temporal Common Spatial Patten). With considering this adjacency matrix, the noise on one time point does not affect the extracting covariance matrix in the far time points. Therefore, LTCSP is less sensitive to potential outliers and artifacts. Similar to CSP method, LTCSP is an eigenvalue problem, so it is computationally simple, but it is more robust than CSP. Wang and Zheng [15] showed that LTCSP can discriminate better than CSP method in a two-class MI-based BCI problem.

LTCSP method used wide band frequency filtering before extracting spatial filters. One defect of LTCSP can be that frequency filtering affects the spatial filter and extracted features that is consequent to wrong discriminating two classes.

One of the limitations for basic CSP method is the fact that it is only designed for a two-class problem. We used OVR (One-Versus-the Rest) algorithm to extend CSP to multiclass problem [16]. As CSP, LTCSP method, that is the extension of CSP method, is limited to binary-class problems, too. We can use all the extension approaches of CSP for LTCSP method to apply it for multiclass problem.

The goal of this paper is classifying four classes of Motor Imagery EEG signals. To conduct experiments, we utilized dataset 2a from the BCI competition, 2008 [14]. It contains EEG signals for 9 subjects. From each subject, it has been requested to imagine four motor imagery of left hand, right hand, both foots and tongue one at a time.

In this paper, we use the OVR approach to extend LTCSP method to a four-class problem. As we see in experimental results, LTCSP has better results than CSP in all subjects of our available data. The success of LTCSP method is considering the local structure of time signals.

Dynamic classifier such as HMM (Hidden Markov Model) was used for MI classification too [17]. Success of these classifiers is for considering temporal structure of EEG signals. This motivates us to use new method, considering time structure of EEG signals.

Observing time frequency distributions of EEG signals reveals that the energy of frequency bands during Motor Imagery varies with time. In other words, frequency patterns are not stable with time [18]. The sequence of this frequency patterns are important in classification. Gouy-Pailler et al. [18] considered this characteristic of MI signals, and proposed MS-JAD method for MI classification. MSJAD is the improved JAD method [19] that in contrary to CSP and JAD methods diagonalizes covariance matrices of four classes simultaneously. MSJAD uses time segmentation on JAD algorithm. In MSJAD method, EEG signals are divided into a number of time segments, then JAD method is applied on all time segments [18]. Pailler et al. [18] compared their proposed method with CSP. They applied CSP method on Motor Imagery time segment and then they divided filtered signals into some time windows. They extracted features from the time segments in frequency domain. They applied their method on our available data, and compared their results with results of the best competitor in the competition IV. Their proposed method does not have good results related to the best competitor.

In this paper, for considering temporal structure of EEG signals in CSP method, we use CSP method on EEG time segments to extract proper features for four-class Motor Imagery classification. Unlike reference [18], in our method, at first we divided EEG signals into some time segments, and then the CSP method was applied on each time segment. The features extracted from all time segments construct feature vector.

In fact, spatial filters are the weights that show the importance of signals under the skull for generating signals extracted on scalp. As mentioned above, the energy of frequency bands are not steady in the whole motor imagery time segment. It is possible that in some channels, some frequency bands activate in one time slot and deactivate in the others. In another time slot, it is possible that other frequency bands activate in other channels. Therefore, the importance of channels for classifying changes in different time slots. We divided time signals into several time segments, and applied CSP method in each segment. Through the experiment we found out the spatial filters vary for different time segments of EEG. This means that the importance of channels is different in each time segment for discriminating two classes.

We compared the result of our proposed method with the result obtained by the best competitor of the competition 2008. The best performance in this competition was related to FBCSP method [20]. In average, with simple features such as variance, we could obtain better results than the results obtained by the method proposed by Gouy-Pailler et al. [18] and the best competitor. In these two methods, features extracted from frequency domain are more complex than our method. Also, our proposed method had better results than LTCSP or the same results in 7 subject of 9 in a four-class problem.

Although our proposed method has good result, we used wide band filtering before applying it on the available data. We can easily apply CSP-based methods, mentioned before [10-13], to select subject-specific frequency bands in each time segment, and improve our method. With this, we can consider both time structure and proper frequency band.

In the current CSP method, noise on non-informative parts of EEG signal affects the estimation of covariance matrix. Thus the extraction of spatial filters which provide features for classification is affected by noise and artifacts. By breaking each EEG channel into equal size time windows, the noise for each time window only affects the features extracted from that window not all features. This reduces the effect of noise and outliers on feature extraction process.

The rest of the paper is organized as follows: We provide a short overview on the CSP and LTCSP methods in Section 2. In this section, we also explain our method and extension of it to a four-class problem by means of OVR method. Data acquisition, evaluation criterion and the result of experiments are explained in Section 3. Finally, the paper is drawn to conclusion in Section 4.

2. Theory

2.1. CSP method

The goal of CSP method is to design spatial filters that lead to new time series whose variances are optimal for the discrimination of two classes of EEG signals. We assume that we have two conditions of Motor Imaginary referred to as C_1 and C_2. Each condition contains some trials. The output of each trial in a two-class problem is EEG signals from the person when he/she is asked to imagine one of the two conditions C_1 or C_2 (e.g. move right hand or left hand). For a trial, signals are represented as $K \times N$ matrix $X = [x_1, x_2, \cdots, x_N]$ where K is the number of EEG channels and N is the number of time samples for each channel. Consequently, $x_i (i \in 1, 2 \cdots N)$ is a column vector with K dimension. It is worth to note that the mean value of the N samples for each channel of EEG signals must be zero. The estimation of the averaged normalized covariance matrix for each condition can be given by [8]:

$$\overline{C}_i = \frac{1}{T_i} \sum_{j \in T_i} \frac{X_j X_j^T}{\mathrm{trace}(X_j X_j^T)}, \qquad i \in \{1,2\}, \tag{1}$$

where T_i is the number of trials for ith condition of MI. The composite spatial covariance is determined by:

$$\overline{C} = \overline{C_1} + \overline{C_2}. \tag{2}$$

$\overline{C}$ can be factorized as:

$$\overline{C} = U^T \Lambda U, \tag{3}$$

where U and Λ are the matrix of eigenvectors and the diagonal matrix of eigenvalues for $\overline{C}$, respectively [8]. These matrices are arranged so that eigenvalues being sorted in descending order.

Then the whitening transformation matrix $P = U\lambda^{-1/2}$ can equalize the variances in the space spanned by U, i.e. all eigenvalues of $P^T\overline{C}P$ are equal to one [8]. This means that we have:

$$P^T\overline{C}P = I \qquad \text{or} \qquad P^T(\overline{C}_1 + \overline{C}_2)P = I, \tag{4}$$

where I is the identity matrix. Now if $\overline{C_1}$ and $\overline{C_2}$ are transformed as $S_1 = P^T\overline{C}_1P$ and $S_2 = P^T\overline{C}_2P$ and if $S_1$is factored by $S_1 = B^T\Lambda_1 B$, then $S_2 = B^T\Lambda_2 B$ and $\Lambda_1 + \Lambda_2 = I$.

This means that eigenvector with the largest eigenvalue for S_1 has the smallest eigenvalue for S_2 and vice versa. Thereupon the projection of whitened EEG signals into the space spanned by eigenvectors corresponding to the largest and smallest eigenvalues of S_1 will give new series of EEG signals that are optimal for the classification of two populations [8]. With the transformation matrix $W = PB$, the mapping of trial X is given by:

$$Z = W^T X. \tag{5}$$

The new signals are the rows of matrix Z. The $2m$ signals corresponding to the m first and last rows of Z are associated with the largest eigenvalues of each condition. Using the normalized variance of these signals, $2m$ features are extracted as follows:

$$f_P = \log\left(\frac{\mathrm{var}(Z_P)}{\sum_{i=1}^{2m} \mathrm{var}(Z_i)}\right), \tag{6}$$

where $P = 1, 2, \cdots, 2m$. Accordingly, from each trail of MI, a feature vector with $2m$ features are obtained and used for classification.

2.2. LTCSP method

As CSP method maximizes the variance of signals for one class and minimizes it for another class, the solution W can be determined by optimizing the following fraction:

$$\max \frac{\mathrm{trace}(W^T\overline{C}_1W)}{\mathrm{trace}(W^T\overline{C}_2W)}. \tag{7}$$

W must maximize the numerator against the denumerator subject to $W^T(\overline{C}_1 + \overline{C}_2)W = I$. This is an eigenvalue problem which can be converted as follows [15]:

$$\frac{\frac{1}{T_1}\sum_{i=1}^{T_1}\sum_{j=1}^{K}\left(w_j^T X_i X_i^T w_j\right)/\mathrm{trace}\left(X_i X_i^T\right)}{\frac{1}{T_2}\sum_{i=1}^{T_2}\sum_{j=1}^{K}\left(w_j^T X_i X_i^T w_j\right)/\mathrm{trace}\left(X_i X_i^T\right)}, \tag{8}$$

where T_1 and T_2 are the numbers of trials under conditions one and two, respectively. Index i corresponds to ith trial, and w_j is the jth column of matrix W. With deletion of indexes, we can expand the term $w^T XX^T w$ as follow:

$$w^T XX^T w = \frac{1}{2N}\sum_{l=1}^{N}\sum_{m=1}^{N}\left(w^T x_l - w^T x_m\right)^2. \tag{9}$$

This says that the sum of the squared pairwise distances between the projected data samples constructs the variance after filtering. In [15], a coefficient H_{lm}^X proposed that it specifies the adjacency value of samples x_l and x_m.

$$w^T XX^T w = \frac{1}{2N}\sum_{l=1}^{N}\sum_{m=1}^{N}\left(w^T x_l - w^T x_m\right)^2 H_{lm}^X. \tag{10}$$

The goal of adjoining this parameter is to reduce the noise effect of far samples in computation. So H_{lm}^X must decrease with increasing the distance between two temporally close data points. This value is between zero and one. Therefore, in Wang et al.'s method, a heat kernel was proposed for it:

$$H_{lm}^X = \begin{cases} \exp(-\|x_l - x_m\|^2/\sigma) & |1 - m| < \tau \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

H^X is the symmetric matrix that is called the adjacency matrix, and $\|.\|$ is the Euclidean norm of the two samples in R^K. In this function, the parameter τ is a positive number that defines the temporally local range out of which H^X is equal zero. The parameter σ is defined as $\sigma = c\sigma_0$, where σ_0 is the standard deviation of the squared norms of the training samples. Indeed, CSP is a special case of LTCSP when $\tau = N$ and $c = +\infty$ that means all samples are involved in optimization and have equal importance. Thus in CSP, the noise of one sample has the same influence on all other samples, whereas LTCSP eliminates this defect. It is worth to notice that adjacency matrix H must be constructed for each of the trials of both conditions.

After some manipulation, the average normalized temporally local covariance matrix is given by:

$$\widetilde{C}_i=\frac{1}{T_i}\sum_{j\in T_i}\frac{X_jL^XX_j^T}{\mathrm{trace}\left(X_jL^XX_j^T\right)},\qquad i\in\{1,2\},\tag{12}$$

where $L^X=D^X-H^X$ is the Laplacian Matrix that is a semi-positive definite matrix; therefore, it can be decomposed as:

$$L^X=L^{X^{1/2}}\times L^{X^{1/2}},$$

and D^X is a diagonal matrix that its diagonal elements are the row sums of H^X i.e. $D_{ll}^X=\sum_{m=1}^N H_{lm}^X$.

Then similar to CSP method, the problem is finding matrix W that maximizes fraction Eq. (13) subject to $\widetilde{W}^T(\widetilde{C}_1+\widetilde{C}_2)\widetilde{W}=I$.

$$\max\frac{\mathrm{trace}(W^T\widetilde{C}_1W)}{\mathrm{trace}(W^T\widetilde{C}_2W)}.\tag{13}$$

Thereupon optimal W is given by:

$$\widetilde{W}=\widetilde{U}\widetilde{D}^{-1/2}\widetilde{V},\tag{14}$$

where $\widetilde{U}$ is the matrix of eigenvectors, and $\widetilde{D}$ is the diagonal matrix of corresponding eigenvalues of $\widetilde{C}_1+\widetilde{C}_2$, and $\widetilde{V}$ is the eigenvectors matrix of $\widetilde{D}^{-1/2}\widetilde{U}^T\widetilde{C}_1\widetilde{U}\widetilde{D}^{-1/2}$. Then in LTCSP method the projected matrix of X is defined by:

$$\widetilde{Z}=\widetilde{W}^TXL^{X^{1/2}}.\tag{15}$$

The m row of the upside and m row of the bottom of Z are suitable signals for classification and the feature vector corresponding to these m signals is evaluated as follows:

$$\widetilde{f}_P=\log\left(\frac{\mathrm{var}(\widetilde{Z}_P)}{\sum_{i=1}^{2m}\mathrm{var}(\widetilde{Z}_i)}\right),\qquad P=1,2,\cdots,2m.\tag{16}$$

It has been reported that, compared with CSP, LTCSP significantly enlarges the difference of variances between the first and second classes. This implies that the variances obtained by LTCSP yield more discriminative information [15].

2.3. The Proposed method in a two-class problem

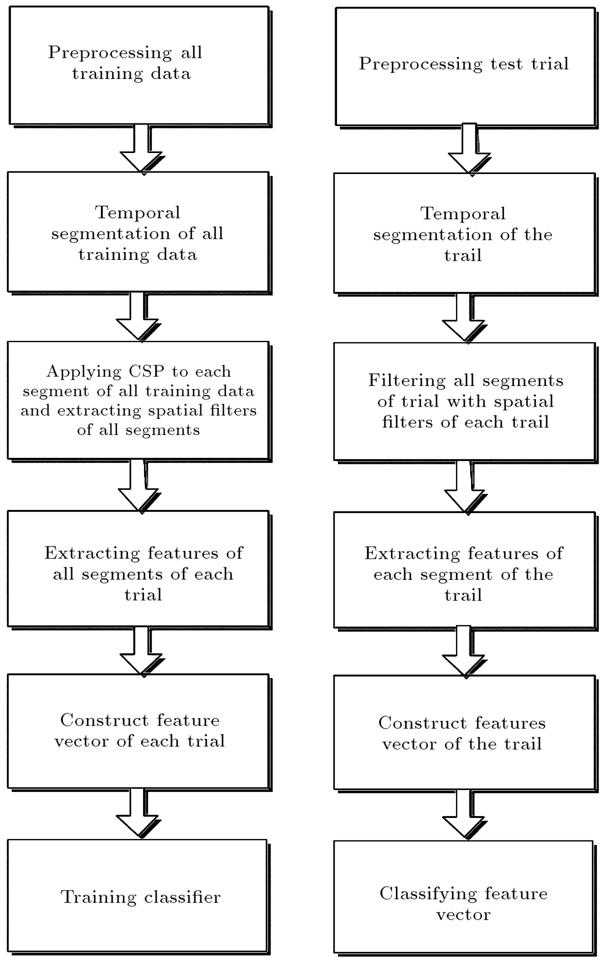
As mentioned above, in our method, EEG signals are divided into a number of time segments. Then CSP method is applied to each time segment for extracting proper features. We extend the proposed approach to a four-class Motor Imagery problem. This method has training and test phases. Flowchart of training and test process in a two-class problem is shown in Figure 1.

As EEG signals in our method are divided into a number of time segments, we refer to this method as SEG-CSP.

2.3.1. Training phase

The aim of training phase is to design a classifier for classification of EEG samples. This phase has these steps:

- **Preprocessing**: Prior to determining the spatial filters, the rate of EEG samples is reduced to 100 Hz, and then they filtered from 8 to 30 Hz, using zero-



(a) (b)

Figure 1. Flowchart for (a) training (b) test of EEG patterns in a two-class MI classification system in SEG-CSP method.

phase forward/backward FIR filter. This frequency band is chosen because it encompasses mu (8-13 Hz) and beta (14-30 Hz) rhythms, which have been shown to be most important for the movement classification [21]; by this filtering artifacts, such EOG and EMG signals removed because the range of frequency band of these artifacts is out of this range [3]. Since ERD/ERS can occur before and after motor imagery execution [1,2], for training our proposed method, we used the 3 s time interval that begins of 0.5 s before the onset of the visual cue and ends to 2.5s after cue.

- **Time segmentation**: After preprocessing, EEG signals are divided into a number of time segments with proper interference. As it is possible that the proper length of the time segments for feature extraction vary for each subject, the length of the time segments is considered a parameter in the proposed algorithm. In training phase, the accuracy of MI classification based on different lengths of time segments are evaluated. The length corresponding to the maximum accuracy is selected. The procedure for segmentation of EEG signals with segment length T second and time shift ΔT second is shown in Figure 2.
- **Extracting spatial filters**: In this step, CSP method is applied to each time segment in two classes, and spatial filters are extracted from each time segment. If we have S time segments, S matrices are obtained, which are denoted by $W_1, W_2, \cdots, W_S$.
- **Training features**: Features from each time segment are obtained by Eq. (6). Accordingly, from ith time segment, a feature vector is extracted, denoted by $\mathbf{F}^i$:

$$\mathbf{F}^i = \begin{bmatrix} f_1^i & f_2^i & \cdots & f_m^i & \cdots & f_{2m1}^i & \cdots & f_{2m}^i \end{bmatrix}, \quad i = 1, 2, \cdots, S. \tag{17}$$

The concatenation of feature vectors corresponding to all time segments results in a feature vector denoted by F:

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}^1 & \mathbf{F}^2 & \cdots & \mathbf{F}^S \end{bmatrix}. \tag{18}$$

Procedure of constructing feature vector is shown in Figure 3.

- **Learning classifier**: Using the extracted training features, a classifier is designed to determine the class label of a test MI.

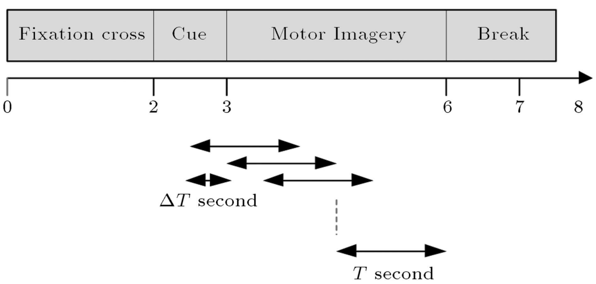


Figure 2. Time segmentation procedure with T second segment length and ΔT second time shift.

2.3.2. Evaluation phase

In evaluation phase, the following steps are taken for each test EEG data:

- Preprocessing of test EEG same as the training phase.
- Segmentation of EEG signals to a number of time segments with the optimal length obtained in the training phase.
- Applying mapping matrix W_i to ith time segment for $i = 1, 2, \cdots, S$ and extracting features of all segments.
- Concatenation of feature vectors extracted from all time segments and obtaining final feature vector.
- Classifying the test MI, using the extracted feature vector and the designed classifier in training phase.

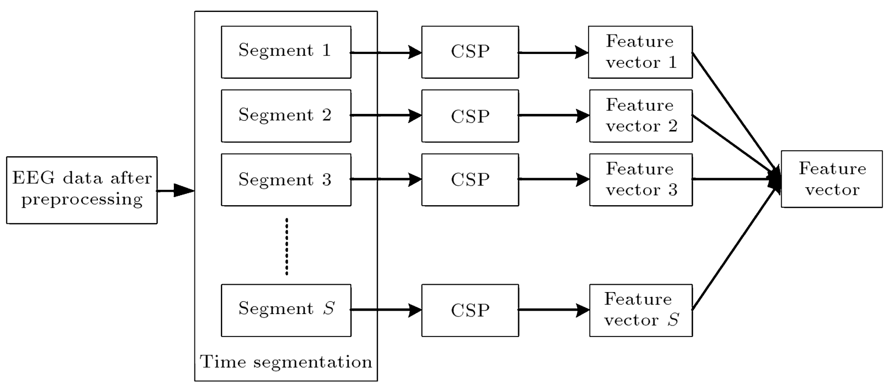


Figure 3. Procedure of constructing feature vector for an EEG trail.

2.4. Extension of CSP-based methods to a four-class problem

One of the limitations of basic CSP method is the fact that it is only designed for a two-class problem. Many studies have been conducted to extend this approach to a multiclass problem [16,19]. In these methods, a multiclass problem is broken into a number of two-class problems. One-Versus-the Rest (OVR) method computes the features of one condition versus the others. Hence, for solving a four-class problem, using OVR approach, we require four classifiers where each classifier states whether the test EEG belongs to that specific class or the rest. Therefore, in Eq. (1), the average covariance of Condition 2 is the sum of average covariance matrices of other 3 conditions [19]. If average covariance matrix of class i ($i = 1, 2, 3, 4$) is denoted by $\overline{R}_i$, then for the ith classifier that discriminates class i from other classes, in Eq. (1), we have:

$$\overline{C}_1 = \overline{R}_i \qquad \text{and} \qquad \overline{C}_2 = \sum_{k \neq i} \overline{R}_k. \tag{19}$$

Since LTCSP is limited to binary-class problems, we can use all these extension approaches of CSP for LTCSP method. In this paper, we use OVR approach to extend a two-class LTCSP method to a four-class problem. We name it OVR-LTCSP. In OVR-LTCSP method, we have four classifiers each of which is a two-class LTCSP method classifying each class versus the rest.

In this paper, we use the OVR approach to apply our SEG-CSP method to a four-class problem too. The extended method is referred to as OVR-SEG-CSP. Similar to basic CSP, for a four-class problem, we need four classifiers in which jth classifer discriminates jth class from the others ($j = 1, 2, 3, 4$). Each classifier is a two-class SEG-CSP method of which the training and test phase was explained in the previous section.

In OVR-SEG-CSP method, jth classifier has a set of matrix W named $W_1^j, W_2^j, \cdots, W_S^j$. The Procedure of training in OVR-SEG-CSP method is shown in Figure 4.

In evaluation phase, as shown in Figure 5, each test EEG data after preprocessing is divided into a number of time segments. Then from the prepared data, four feature vectors are extracted. Each feature vector is obtained using special filters for classifying one class versus the rest. Consequently, four results

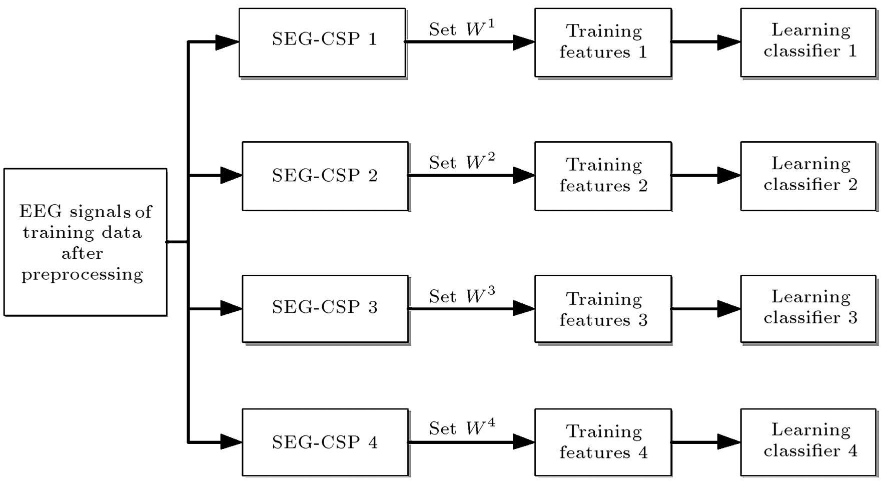


Figure 4. Procedure of training in OVR-SEG-CSP method.

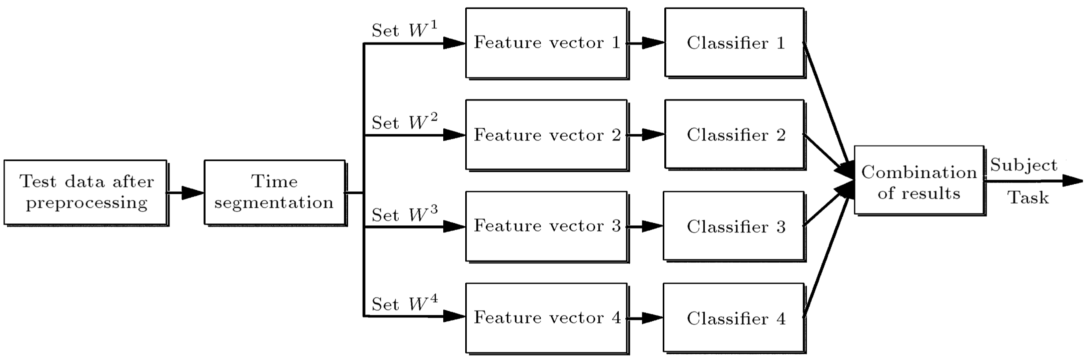


Figure 5. Evaluation phase in OVR-SEG-CSP method.

of classification by 4 classifiers are provided. The true label of the test EEG data is determined by the classifier that provides the highest class label probability [21]. In the problem with a few training data and noisy features, it is better to use steady and simple classifiers with a few parameters [17]. In this research, fisher's Linear Discriminant Analysis (LDAs) has been used for classification.

3. Experiments

3.1. Evaluation criterion

In this paper, we used the accuracy criterion for presenting the result of classifying EEG signals of a two-class motor imagery data. We applied the proposed method to dataset $2a$ of the BCI competition 2008 [14]. In order to compare the performance of this method with the best competitor in the 2008 competition we used kappa score, as used for evaluation of methods in this competition, for a four-class problem. The kappa score is determined as follow [22]:

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{20}$$

where p_0 is the overall agreement on all test trails. This parameter is equal to the classification accuracy. p_e is defined as the chance agreement. These parameters are calculated as:

$$p_0 = \sum_{j=1}^{M} n_{jj}/N, \tag{21}$$

and

$$p_e = \sum_{j=1}^{M} n_{:i}n_{i:}/N^2. \tag{22}$$

n_{ij} is the element of confusion matrix on ith row and jth column which indicates how many trails of class i have been classified as class j. In an M-class problem, $n_{i:}$ and $n_{:i}$ refer to the sum of the elements of ith row and ith column of confusion matrix, respectively. The diagonal element n_{ii} represents the number of correctly classified test trials of ith class. The total number of trails is:

$$N = \sum_{i=1}^{M}\sum_{j=1}^{M} n_{ij}. \tag{23}$$

3.2. Dataset

The dataset 2a of the BCI Competition IV (2008) has been provided by the BCI research group at Graz University [14]. This dataset contains EEG signals recorded from nine subjects (persons) performing four different motor imagery tasks, i.e. left hand, right hand, both feet and tongue. The recording concluded 22 channels (with interelectrode distances of 3.5 cm), and was sampled with rate 250 Hz. Monopolar derivations were used throughout all recordings, where the left mastoid served as reference, and the right mastoid as ground. The dataset consists of 288 trials for training and 288 trials for test. This cue-based BCI includes three important periods. In the first 2 seconds a fixation cross is displayed on the screen; in the next second, cue is appeared; third period is from 3s to 6s in whitch subject is imagining the related action. The paradigm and montage are illustrated in Figure 6.

3.3. Results

Since BCI problems are subject-specific, in our proposed method, the length and overlap of time segments and m parameter of CSP method in these segments are different for each subject. For this reason, we must obtain optimal value of these parameters for each subject. If the overlap of time segments becomes large, the number of segments would be large. This causes the increase of dimensionality of feature vectors and the complexity of algorithm. Accordingly, the shift of time segments is chosen ($\delta T = 0.5$ s). The length of time segments varies from 0.5 to 3 s with step time 0.5 s (i.e. $T = 0.5, 1, 1.5, 2, 2.5, 3$ in Figure 2). The number of spatial filters ($2m$) is considered as a parameter which varies from 2 to 16 (m is 1 to 8). The parameters T and m are the same in all time segments. The result of a four-class problem for subject 1 with different parameters T and m has been shown in Table 1.

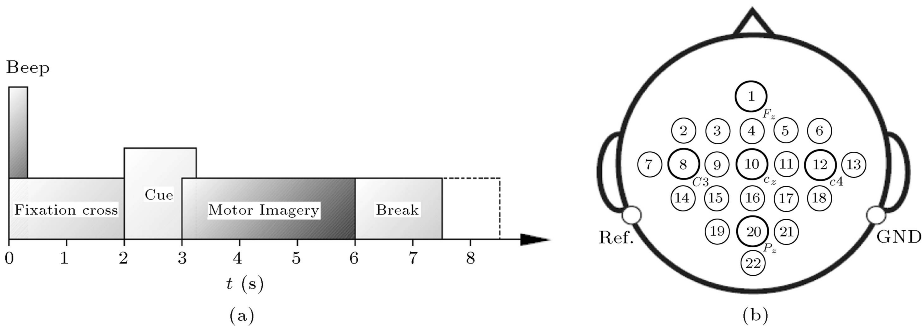


Figure 6. (a) Timing scheme of the BCI paradigm. (b) Electrode setup of the 22 channels [14].

Table 1. The classification rate for OVR-SEG-CSP method for subject 1 in classifying EEG signals of two class left and right hands with parameters T and m.

m	$T = 0.5$ s	$T = 1$ s	$T = 1.5$ s	$T = 2$ s	$T = 2.5$ s	$T = 3$ s
1	0.69	0.62	0.70	0.67	0.69	0.66
2	0.63	0.65	0.69	0.67	0.67	0.68
3	0.65	0.67	0.72	0.71	0.72	0.71
4	0.65	0.62	0.68	0.71	0.72	0.70
5	0.63	0.62	0.72	0.72	0.70	0.74
6	0.53	0.6	0.68	0.70	0.69	0.67
7	0.53	0.61	0.68	0.69	0.68	0.66
8	0.49	0.59	0.65	0.66	0.64	0.68

Table 2. The kappa score for evaluation of the proposed method OVR-SEG-CSP against the best competitor in the competition IV, three methods proposed in [18], OVR-LTCSP and OVR-CSP.

	The best competitor in BCI 2008	**The method proposed by Gouy-Pailler et al. [18]**			**OVR-CSP (m)**	**OVR-LTCSP (c, τ, m)**	**OVR-SEG-CSP (T, m)**
		CSP	**JAD**	**MSJAD**			
Subject 1	0.68	0.52	0.65	0.66	**0.74**(5)	**0.77(5,4,5)**	0.74 (3s,5)
Subject 2	0.42	0.39	0.40	0.42	0.34 (4)	0.35(9,5,4)	**0.43 (2.5s,6)**
Subject 3	0.75	0.67	0.77	0.77	0.76 (4)	**0.80 (3,4,2)**	**0.80**(2s,2)
Subject 4	0.48	0.50	0.50	0.51	0.51 (3)	**0.60 (10,2,3)**	**0.57**(2s,2)
Subject 5	0.40	**0.49**	0.44	0.50	0.22 (5)	0.31 (8,2,5)	0.31 (2.5s,2)
Subject 6	0.27	0.18	0.19	0.21	0.27 (3)	0.37 (2,5,3)	**0.38**(1s,1)
Subject 7	**0.77**	0.26	0.25	0.30	0.69 (3)	0.75 (2,5,3)	0.76 (2s,2)
Subject 8	**0.75**	0.57	0.72	0.69	0.62 (5)	0.74 (8,3,5)	**0.75** (1.5s,7)
Subject 9	0.61	0.40	0.50	0.46	0.79 (3)	**0.80 (3,5,7)**	**0.80** (2.5s,2)
Average	0.57	0.41	0.49	0.50	0.55	**0.61**	**0.61**

We repeated similar experiment for all subjects. The maximum performance and the corresponding parameters (m, T) for each subject have been shown in Table 2. For instance, for subject 1, the maximum performance is obtained when $T = 3$ s and $m = 5$ are selected.

In order to compare the proposed method with the existing CSP method, we applied the OVR-CSP method [8] to the BCI dataset, i.e. OVR-CSP method is applied on whole considered 3 s (from 2.5 s to 5.5 s). The number of spatial filters $(2m)$ is considered as a parameter which varies from 2 to 16 (m is 1 to 8). We obtained the classification rate for different values of m. The maximum classification rate for each subject and the optimal m for obtaining it are shown in column 5 of Table 2.

As we see in Table 2, our proposed method has a better result than OVR-CSP in all subjects except subject 1. In the worst case, i.e. subject 1, two methods have the same results. In subject 1, we obtained maximum kappa in $T = 3$ and $m = 5$. With $T = 3$, we only have one time segment as OVR-CSP, that two methods are the same and have the same results. Thereupon, current CSP method is the state of our proposed method with $T = 3$.

In subject 8, the maximum kappa in our proposed method has been obtained in $T = 1.5$ s and $m = 7$. For this case, four time segments with $T = 1.5$ s and $\Delta T = 0.5$ s exist in 2.5 to 5.5 time interval. First and second important spatial filters of each time segment in classifier 1 (left hand versus the rest) are shown in Figure 7.

As we see in this figure, spatial filters are different in time segments. This means that the importance

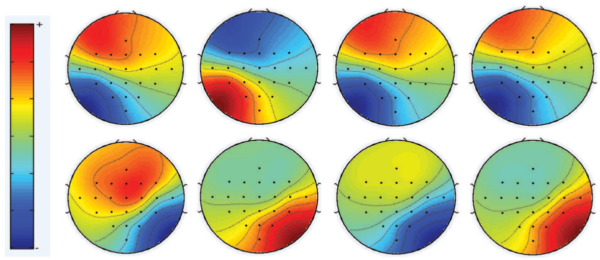


Figure 7. The selected two spatial filters in 4 time segment method for classifying left and right hand motor imagery classes for subject 2 by SEG-CSP.

of channels is different for classifying two classes in time segments. In current CSP method, only one time segment and consequently one spatial filter matrix exist, that means the importance of channels considered similar in whole time segment. It is possible that specific events happen in one time segment that do not happen in the others. It is possible that one frequency band activates in one channel in one time segment that deactivate in other segment. For this reason, the importance of channels is different in segments. With extracting the events happening in different time intervals and considering all of them in classification process, we can get better results.

We apply OVR-LTCSP method on our available data. Similar to work done in [15], we had set $\sigma = c\sigma_0$, where $m \in \{1, \cdots, 8\}$, $\tau \in \{2, \cdots, 5\}$ and $c \in \{1, 2, \cdots, 10\} \cup +\infty$ for LTCSP method. As mentioned above, $\sigma 0$ is the standard deviation of the squared norms of the training samples that we considered the training data of four classes as training data, where this and other parameters (i.e. m, τ and c) are fixed during the four classifiers of OVR method in one experiment. The maximal kappa and the corresponding parameters (m, τ, c) for each subject have been shown in Table 2.

As we see, both OVR-CSP and our proposed methods have the same results in 3 subjects of the nine. In four of the subjects, our proposed method has better results than OVR-LTCSP and in 2 subjects, vice versa. Anyway, both of them have the same results in average.

Our proposed method has two parameters (m and T) and used CSP method in each time segment. LTCSP has 3 parameter and is more complex than CSP. thereupon, LTCSP is more complex than our proposed method which gives more time for execution. In our proposed method, time segmenting is done on EEG signals; CSP is applied in all time segments and features are extracted from all time segments. Therefore, the same process is applied in all segments. We can execute our proposed method in parallel process on the time segments. Thereupon, our proposed method is faster than LTCSP method in training and test phase.

In Table 2, we also compare the performance of three addressed methods proposed by Gouy-Pailler et al. (2010) [18]. In CSP-based method proposed by them, in each classifier of OVR approach, at first they applied CSP on whole 3 s motor imagery time segments. After filtering by spatial filters, filtered signals divided into five time segments and features in frequency domain are extracted from each segment, and all features classified by logistic regression method. We emphasize that in the proposed method, EEG signals are divided into a number of equal lengths time segments. Then the CSP method is applied on each time segment. We could obtain better results than the CSP-based method. This means that segmentation before CSP can extract more effective events for classification. Our proposed method outperformed MSJAD method proposed by Pailler et al. [18] that is more complex than our method. They used equal length for time segments for all subjects, but we considered it variable and selected optimal value for each subject.

The results of the method proposed by the best competitor of competition 2008 are shown in Table 2, too. The best competitor used OVR approach for expanding FBCSP method [13] to a four-class problem. They used 2 s time segment from 2.5 s to 4.5 s. In FBCSP, nine band-pass filter is applied in the time segment of EEG signals, and CSP is applied on the filtered signals in each frequency bands. After filtering the signals by spatial filters in each frequency band, features are extracted. Features are reduced by a feature selection algorithm based on mutual information classified by SVM (Support Vector Machine) classifier. Our proposed method outperformed the method proposed by the best competitor in six subjects of the nine by the simple features such as variance, the simple classifier such as fisher LDA and without feature reduction.

We used wide band filtering before doing our proposed method on EEG signals. We applied CSP method in each time segments. We can apply CSP-base method such as SBCSP, FBCSP and other stead CSP methods in each time segment, and improve our method. With this manner subject-specific frequency patterns are extracted from each time window and features with more information for classification are obtained from EEG signals. With parallel processing in all segments we can save time.

4. Conclusions

In this paper we addressed the problem of multiclass BCI Motor imagery recognition. We used dataset 2a of BCI competition IV to evaluate the performance of the proposed method. This database contains EEG signals related to four different motor imagery tasks of nine subjects.

One of the most successful methods in MI-based BCI is CSP method. This method obtains spatial filters from EEG data; such filters reveal the importance of EEG channels in discrimination of MI data. In the existing CSP method, spatial filters are extracted using EEG channels as one time segment. The fact that the generated frequency patterns in Motor Imagery are different from one time slot to another, motivated us to apply CSP on EEG time segment. The augment of features from EEG time segments are used as the feature vector for discriminating EEG signals for two MI classes. We utilized the One-Versus-the-Rest (OVR) approach [16] to extend the proposed method to the four-class MI problem. We used LDA algorithm for this classification.

CSP method is based on estimating covariance matrix, and does not consider temporal segmentation of EEG signals. LTCSP [15] is the extension of CSP method that considers time structure of the signals.

We used time segmentation for considering the information of consecutive time segments. Our proposed method and LTCSP have the same result in average, but our method is simpler than the other.

The best competitor used FBCSP method [13] for classifying a four-class motor imagery. Our method outperforms in many subject and in average of the best competitor's method. Our proposed method conquest MSJAD and CSP and JAD-based method proposed by Gouy-Pailler et al. [18]. In their CSP-based method they apply CSP method on whole motor imagery time segment, and then they divided filtered signal, by extracted spatial filters, into several time segments. In our method, we used time segmentation and then CSP applied in each time segment. More success of our method shows that time segmentation before CSP has a significant effect on the improvement of results.

Obtaining the good results with our proposed method and using the simple features such as variance and simple classifier such as fisher LDA shows that information of consecutive time segments are important for classification. Our good achievement is due to the fact that the effect of noise and artifacts are reduced by segmenting EEG signals into a number of time slots.

However, the performance of this method depends on an appropriate pre-filtering in frequency domain. Since MI-based BCI is a subject-specific problem, as a future work, the performance of BCI system can be improved by selecting proper frequency band for each subject individually. For instance, we can apply CSP-based method such as SBCSP and FBCSP methods instead of CSP method in each time interval, and extract frequency bands, activate or deactivate, in each time segment.

Our proposed method applies the same process on all time segments; therefore, we can have faster method with parallel processing.

References

1. Pfurtscheller, G. and Neuper, C. "Motor imagery activates primary sensorimotor area in humans", *Neurosci. Lett.*, **239**(2-3), pp. 65-68 (1997).

2. Pfurtscheller, G. and Lopes da Silva, F.H. "Event-related EEG/MEG synchronization and desynchronization: basic principles", *Clin. Neurophysiol.*, **110**(11), pp. 1842-1857 (1999).

3. Fatourechi, M., Bashashati, A., Ward, R.K. and Birch, G.E. "EMG and EOG artifacts in brain computer interface systems: A survey", *Clinical Neurophysiology*, **118**, pp. 480-494 (2007).

4. Sanei, S. and Chambers, J.A., *EEG Signal Processing*, John Wiley & Sons, Centre of Digital Signal Processing, Cardiff University, UK (2007).

5. Bashashati, A., Fatourechi, M., Ward, R.K. and Birch, G.E. "A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals", *J. Neural En.*, **4**, pp. 32-57 (2007).

6. Müller-Gerking, J., Pfurtscheller, G. and Flyvbjerg, H. "Designing optimal spatial filters for single-trial EEG classification in a movement task", *Electroenc. Clin. Neurophysiol.*, **110**, pp. 787-798 (1999).

7. Koles, Z.J. "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. Electroencephalogr", *Clin. Neurophysiol*, **79**(6), pp. 440-447 (1991).

8. Ramoser, H., Muller-Gerking, J. and Pfurtscheller, G. "Optimal spatial filtering of single trial EEG during imagined hand movement", *IEEE Trans. Rehabil. Eng.*, **8**(4), pp. 441-446 (2000).

9. Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G. and Müller, K.-R. "Combined optimization of spatial and temporal filters for improving brain-computer tnterfacing", *IEEE Trans. Biomed. Eng.*, **53**(11), pp. 2274-2281 (2006).

10. Lemm, S., Blankertz, B., Curio, G. and Müller, K.-R. "Spatio-spectral filters for improved classification of single trial EEG", *IEEE Trans. Biomed. Eng.*, **52**(9), pp. 1541-1548 (2005).

11. Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G. and Müller, K.-R. "Combined optimization of spatial and temporal filters for improving brain-computer interfacing", *IEEE Trans. Biomed. Eng.*, **53**(11), pp. 2274-2281 (2006).

12. Novi, Q., Guan, C., Dat, T.H. and Xue, P. "Sub-band common spatial pattern (SBCSP) for brain-computer interface", *3rd International IEEE/EMBS Conference on Neural Engineering CNE*, **07**, pp. 204-207 (2007).

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

14. http://bbci.de/competition/iv/index.html

15. Wang, H. and Zheng, W. "Local temporal common spatial patterns for robust single-trial EEG classification", *IEEE Trans. Neural Systems and Rehabilitation Engineering*, **16**(2), pp. 1534-432 (2008).

16. Dornhege, G., Blankertz, B., Curio, G. and Müller, K.-R. "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms", *IEEE Transactions on Biomedical Engineering*, **51**(6), pp. 993-1002 (2004).

17. Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F. and Arnaldi1, B. "A review of classification algorithms for EEG-based brain-computer interfaces", *J. Neural Eng.*, **4**, pp. 1-13 (2007).

18. Gouy-Pailler, C., Congedo, M., Brunner, C., Jutten, C. and Pfurtscheller, G. "Nonstationary brain source separation for multiclass motor imagery", *IEEE Transactions on Biomedical Engineering*, **57**(2), pp. 469-478 (2010).

19. Dornhege, G., Blankertz, B., Curio, G. and Müller, K.-R. "Increase information transfer rates in BCI by CSP extension to multi-class", *Advances in Neural Information Processing Systems*, pp. 733-740 (2004).

20. Chin, Z.Y., Ang, K.K., Wang, C., Guan, C. and Zhang, H. "Multi-class filter bank common spatial pattern for four-class motor imagery BCI", *31st Annual International Conference of the IEEE EMBS Minneapolis Minnesota*, USA, pp. 2-6 (2009).

21. Brunner, C., Naeem, M., Leeb, R., Graimann, B. and Pfurtscheller, G. "Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis", *Pattern Recognition Letters*, **28**, pp. 957-964 (2007).

22. Schlögl, A., Kronegg, J., Huggins, J.E. and Mason, S.G. "Evaluation criteria in BCI research", In *Toward Brain-Computer Interfacing*, MIT Press, pp. 327-342 (2007).

Biographies

Habibeh Ghaheri received her BSc and MS in Electrical Engineering from Shahrood University, Iran, in 2009 and 2012, respectively. Her current research interests are in pattern recognition, signal and image processing. She has worked also on brain computer interface topic recently.

Alireza Ahmadyfard graduated from Isfahan University of Technology in Electrical Engineering in 1989. He obtained his MSc degree in Communication Engineering from Amkabir University of Technology in 1994 and his PhD in Object Recognition in Computer Vision from University of Surrey in 2002. Now he is a member of Shahrood University of Technology. His research interests include signal processing, pattern recognition and computer vision.