Journal of Neuroscience Methods 222 (2014) 238–249

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Computational Neuroscience

A tensor-based scheme for stroke patients' motor imagery EEG analysis in BCI-FES rehabilitation training

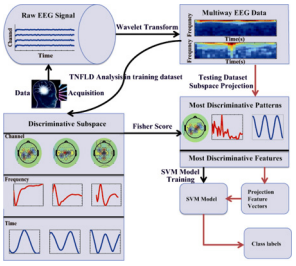
Ye Liu[a], Mingfen Li[b], Hao Zhang[a], Hang Wang[a], Junhua Li[a], Jie Jia[b], Yi Wu[b,**], Liqing Zhang[a,*]

[a] MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[b] Department of Rehabilitation, Huashan Hospital, Fudan University, Shanghai 200040, China

HIGHLIGHTS

- Motor imagery EEG patterns in spatial–spectral–temporal domain are simultaneously obtained.
- Prior knowledge of the most contributed channels and frequency bands is provided.
- The results are validated by online and offline simulations based on real EEG from strokes.
- Neural mechanism about motor cortex reorganization of stroke patients during rehabilitation is revealed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 10 October 2013
Received in revised form 8 November 2013
Accepted 13 November 2013

Keywords:
Electroencephalogram
Brain–computer interface
Tensor factorization
Rehabilitation

ABSTRACT

Background: Stroke is one of the most common disorders among the elderly. A practical problem in stroke rehabilitation systems is that how to separate motor imagery patterns from electroencephalographic (EEG) recordings. There is a sharp decline in performance of these systems when classical algorithms, such as Common Spatial Pattern (CSP), are directly applied on stroke patients.
New method: We propose a tensor-based scheme to detect motor imagery EEG patterns in spatial–spectral–temporal domain directly from multidimensional EEG constructed by wavelet transform method. Discriminative motor imagery EEG patterns are obtained by Fisher score strategy. Furthermore, the most contributed channel groups and frequency bands are selected from these patterns and utilized as prior knowledge for the following motor imagery tasks.
Results: We evaluate our scheme based on EEG datasets recorded from stroke patients. The results show that our method outperforms five other traditional methods in both online and offline recognition performance.
Comparison with existing methods: Unlike the existing methods, motor imagery EEG patterns in spatial–spectral–temporal domain are simultaneously obtained by our method, preserving the structural information of the multi-channel time-varying EEG.
Conclusions: Our scheme is encouraged to be transferred to some other practical rehabilitation applications for its better performance.

* Corresponding author. Tel.: +86 1300 317 6229.
** Corresponding author. Tel.: +86 021 34204423.
E-mail addresses: lyyx000@sjtu.edu.cn (Y. Liu), wuyi4000@163.com (Y. Wu), zhang-lq@cs.sjtu.edu.cn (L. Zhang).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.11.009

1. Introduction

Brain–computer interface (BCI), as a new approach for people to communicate actively with machine, is a promising technology that allows patients who have totally lost the ability of movement to communicate with the outside world. Recently, BCI has also been used as a rehabilitation tool, for patients with spinal cord injury and stroke (Pfurtscheller et al., 2005; Buch et al., 2008). Sufficient evidence has demonstrated that physical practice along with motor imagery practice can lead to improvements in wrist movements, isolated movements of the hands and fingers and object manipulation of the impaired upper limb (Crosbie et al., 2004; Page et al., 2007), making the rehabilitation training tool promising.

However, a pervasive problem faced by stroke rehabilitation systems is how to separate motor imagery EEG patterns from noisy EEG recordings. Some previous studies (Calautti et al., 2001; Loubinoux et al., 2003; Shahid et al., 2010) have proved that the motor imagery EEG patterns in stroke patients differ from those of healthy people in both spectral and spatial domains. There is a dynamic change in the activation pattern during recovery according to some fMRI and PET studies (Calautti et al., 2001; Loubinoux et al., 2003), and new pattern may deviate from that of healthy subjects. Therefore, cortex regions responded to motor imagery may vary over time. Moreover, Shahid et al. (2010) have reported that active rhythms may have been migrated and modulated during rehabilitation in affected hemisphere.

In this case, traditional feature extraction methods, like the widely used Common Spatial Pattern (CSP) (Ramoser et al., 2000), do not work well for extracting motor imagery EEG patterns. One issue is the overfitting problem for CSP when faced with large number of channels (Hill et al., 2006). Removing the redundant electrodes may help solve the overfitting problem. Another issue is the unavailable frequency band related to motor imagery of stroke patients. CSP calculates discriminative spatial patterns based on the temporal variances of signals, thus it is necessary to filter the EEG signals in a specific pre-identified frequency band related to motor imagery. However, motor imagery EEG patterns of stroke patients, which deviate from those of healthy subjects, are changing during rehabilitation. It is not available to define both channels configuration and frequency bands beforehand when directly applying CSP on stroke patients' signals.

Another practical problem is the small sample size (SSS) problem. The size of the training set is small but the dimensionality of the feature space is usually much larger, which leads to SSS problem. One solution is to reduce the dimensionality of the feature space using principal component analysis (PCA) (Han and Bhanu, 2004), but it also leads to another problem that useful discriminative information may not be preserved.

The problems mentioned above often lead to a poor performance of pattern recognition on stroke patients' EEG, which is a fundamental for implementing BCI-based systems. In order to tackle these problems, we proposed a tensor-based scheme for detecting motor imagery EEG patterns of stroke patients in a new rehabilitation training system combined BCI with Functional Electrical Stimulation (BCI-FES), which has four contributions as follows:

1. Motor imagery EEG patterns of stroke patients are detected in spatial–spectral–temporal domain from limited training datasets. By tracking the gradual changes of motor imagery EEG patterns in spectral and spatial domains during rehabilitation, some interesting phenomenon's about motor cortex recovery are revealed, providing physiological knowledge for biomedical rehabilitation engineering.
2. By obtaining the most discriminative patterns by Fisher score strategy in the training model course, prior knowledge of the most contributed channels group and frequency bands is provided for the following rehabilitation training course.
3. More variations of EEG can be captured as compared with the original datasets, so that it expands the capacity of the available database. Therefore, the scheme can solve the small sample size problem and has better generalization performance for the overfitting problem on small datasets.
4. Unlike other algorithms, such as power spectrum density (first order statistics) and CSP (second order statistics), our scheme handles data using higher-order statistics and thus can preserve the most discriminative patterns.

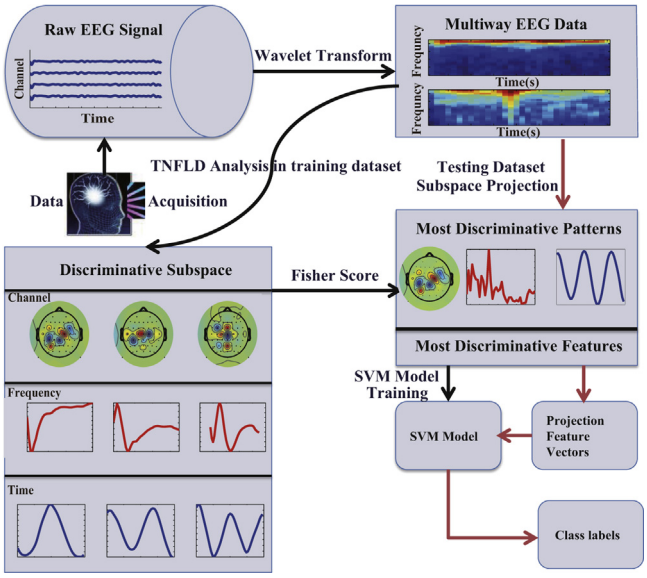


Fig. 1. A brief view of the tensor-based scheme. The black arrows stand for the procedure during the training stage, and the red ones stand for the testing procedure. The scheme is consisted of four components: pre-processing, tensor-based nearest feature line distance method, Fisher score for feature selection and SVM for classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The rest of the paper is organized as follows. Details of the tensor-based scheme is illuminated in Section 2. Section 3 briefly provides experiments configuration and data acquisition. Section 4 displays comparison results including the gradual changes of motor imagery EEG patterns and online and offline recognition performance. Finally, we give a brief discussion of the results in Section 5.

2. The tensor-based scheme

In this section, as shown in Fig. 1, we first present the tensor-based scheme, which is consisted of four components: pre-processing, tensor-based nearest feature line distance method, Fisher score for feature selection and support vector machine (SVM) (Chang and Lin, 2011) for classification. Afterwards, we detail all the components in the following subsections.

2.1. Data pre-processing

A step of pre-processing for EEG data is required to improve the low signal-to-noise ratio (SNR) before feature extraction. Firstly, we employ FastICA (Delorme and Makeig, 2004) method to remove artifacts arising from eye and muscle movements. We select the clean EEG-like components from three main criteria: (1) dipole-like scalp maps, (2) EEG-like spectra, (3) EEG-like temporal structures. EEG-like spectra is explained as spectra peak at (8–13 Hz) and (14–30 Hz) rhythm (Delorme and Makeig, 2004).

EEG-like components can be fit with very low residual variance (Delorme and Makeig, 2004). Afterwards, EEG signals are bandpass filtered within the specific frequency band 8–45 Hz, aiming at containing relatively broad rhythms. Finally, we use the wavelet transform method to transform EEG into a tensor in the spatial–spectral–temporal domain, and thus the structural information is contained in multi-channel time-varying EEG spectrums. In this paper, EEG is reconstructed using a complex Morlet wavelet (Mørup et al., 2006) with center frequency 1 and bandwidth parameter 2, i.e. $\phi(t) = \frac{1}{\sqrt{2\pi}} \exp(2i\pi t) \exp(-t^2/2)$ with frequencies from 8 to 45 Hz with 1 Hz interval.

2.2. The TNFLD algorithm

In this scheme, the tensor-based nearest feature line distance (TNFLD) algorithm is applied for extracting a multilinear discriminative subspace of stroke patients' signals and detecting motor imagery EEG patterns. Unlike other two-dimensional nearest feature line (NFL) methods (Chien and Wu, 2002; Li et al., 2000), TNFLD is used for obtaining a multilinear discriminative subspace for feature extraction rather than for classification. Moreover, it calculates the multilinear subspace directly from multidimensional data, aiming at preserving discriminative information from limited datasets.

2.2.1. Related work

Feature line (FL) is defined as a straight line passing through two prototypes in the same class. The feature line can be considered as a simpler version of the spline type manifold of the parametric appearance representation (Murase and Nayar, 1995). In Murase and Nayar (1995), the appearance manifold of an object is constructed from images of the object taken on a turnable (parameterized by a single parameter) under carefully controlled lighting (parameterized by another single parameter). However, it is difficult to ensure such strict conditions when acquiring images from diverse domains. The nearest feature line method can solve this modeling problem and provides a simple constructive solution without such conditions.

Suppose a variation in the image space from point y_1 to y_2 and the corresponding variation constructed by Gabor and wavelet from x_1 to x_2. $\delta y = ||y_2 - y_1||$ or $\delta x = ||x_2 - x_1||$ measures the degree of the change. We can use a straight line $\widetilde{x_1 x_2}$ to simulate the change of x, which can interpolate any variation. The feature line uses a linear model to provide infinite feature points and their information. Thus virtual points are generated by feature line and the available dataset is expanded by this feature line subspace.

2.2.2. Formulation

Definition 2.1. Tensor-based Feature Line

Tensor-based Feature Line (TFL) is described as a line passing through two multidimensional data $\underline{\mathbf{X}}_1, \underline{\mathbf{X}}_2 \in R^{I_1 \times I_2 \times \cdots \times I_M}$ with the same label, denoted as $\overline{\underline{\mathbf{X}}_1 \underline{\mathbf{X}}_2}$. As illustrated in Fig. 2(a), $\underline{\mathbf{X}}^*$ is the projection of $\underline{\mathbf{X}}$ onto TFL $\overline{\underline{\mathbf{X}}_1 \underline{\mathbf{X}}_2}$. Then, the FL distance between $\underline{\mathbf{X}}$ and $\overline{\underline{\mathbf{X}}_1 \underline{\mathbf{X}}_2}$ is defined as the Euclidean distance between $\underline{\mathbf{X}}$ and $\underline{\mathbf{X}}^*$, denoted as follows:

$$dis(\underline{\mathbf{X}}, \overline{\underline{\mathbf{X}}_1 \underline{\mathbf{X}}_2}) = dis(\underline{\mathbf{X}}, \underline{\mathbf{X}}^*) = \sqrt{\sum_{i_1=1}^{I_1} \cdots \sum_{i_M=1}^{I_M} \left((\underline{\mathbf{X}})_{i_1 \ldots i_m} - (\underline{\mathbf{X}})^*_{i_1 \ldots i_m} \right)^2} \tag{1}$$

where $\underline{\mathbf{X}}^* = \underline{\mathbf{X}}_1 + t(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1)$ and position parameter t is

$$t = \frac{[\![(\underline{\mathbf{X}} - \underline{\mathbf{X}}_1) \otimes (\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1); (1:M)(1:M)]\!]}{[\![(\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1) \otimes (\underline{\mathbf{X}}_2 - \underline{\mathbf{X}}_1); (1:M)(1:M)]\!]}$$

here $\underline{\mathbf{X}}^{I_1 \times \cdots \times I_M \times J_1 \times \cdots \times J_N} \otimes \underline{\mathbf{Y}}^{I_1 \times \cdots \times I_M \times K_1 \times \cdots \times K_P}$ represents the contracted product of two tensors $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ along the first M modes (Tao et al., 2007).

Unlike the traditional two-dimensional nearest feature line, tensor-based feature line takes the multidimensional (like spatial–spectral–temporal) information into consideration, preserving the discriminative structural information of the data which is useful for both feature extraction and data classification. Structural information of the data is usually available and considered as a priori information on the data nature. Hence TNL might provide advantages over the traditional feature line by enabling one to more effectively use the underlying structural properties.

Definition 2.2. Within-class FL distance and between-class FL distance

Assuming that $\Omega = \{\underline{\mathbf{X}}_n^{I_1 \times I_2 \times \cdots \times I_M}, c_n\}_{n=1}^N$ represents the training dataset, $\underline{\mathbf{X}}_n^{I_1 \times I_2 \times \cdots \times I_M}$ denotes the n-th sample of the training dataset and c_n represents the corresponding class label. Suppose the number of samples belonging to class c_n is $(l_n + 1)$, then for each point $\underline{\mathbf{X}}_n|_{n=1}^N$, we can calculate the number of TFLs, formed by the samples which share the same label with $\underline{\mathbf{X}}_n$, is $P_n = l_n(l_n - 1)/2$; Let $\{\underline{\mathbf{X}}^*_{n1}, \underline{\mathbf{X}}^*_{n2}, \ldots, \underline{\mathbf{X}}^*_{nP_n}\}$, which is also regarded as the generated virtual data, represents the projections of $\underline{\mathbf{X}}_n$ onto all the TFLs. The *within-class FL distance* between $\underline{\mathbf{X}}_n$ and all its projections is defined as $\sum_{p=1}^{P_n} dis^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^*_{np})$, and the pth projection of $\underline{\mathbf{X}}_n$ is denoted as $\underline{\mathbf{X}}^*_{np}$. Furthermore, the within-class FL distance between all the points in the training set and their projections is:

$$J_{within} = \sum_{n=1}^{N} \sum_{p=1}^{P_n} dis^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^*_{np}) \tag{2}$$

Similarly, the *between-class FL distance* between all the points and their projections is expressed as:

$$J_{between} = \sum_{n=1}^{N} \sum_{q=1}^{Q_n} dis^2(\underline{\mathbf{X}}_n, \underline{\mathbf{X}}^{\#}_{nq}) \tag{3}$$

where Q_n is the number of projections onto all the TFLs formed by the samples sharing different label with $\underline{\mathbf{X}}_n$, and $\underline{\mathbf{X}}^{\#}_{nq}$ is the qth projection of $\underline{\mathbf{X}}_n$. Fig. 2(b) gives a brief view of the concept of within-class FL distance and between-class FL distance.

The aim of TNFLD is to find a set of projection matrices $\{\mathbf{W}^{(d)} \in R^{I_d \times U_d},\ U_d \le I_d\}_{d=1}^M$, by which original EEG $\underline{\mathbf{X}}$ can be projected to $\underline{\mathbf{Y}} = \underline{\mathbf{X}} \prod_{d=1}^M \times_d \mathbf{W}^{(d)^T} \in R^{U_1 \times U_2 \times \cdots \times U_M}$. These matrices, which are optimal for distinguishing different classes, can project original data in a low dimensional space in which the within-class FL distance is minimized and the between-class FL distance is maximized. Therefore, TNFLD is formulated as an optimization problem:

$$f(\mathbf{W}^{(d)}) = \arg\min_{W_{d=1}^M} \frac{1}{NP_i} \sum_{i=1}^{N} \sum_{p=1}^{P_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip}) - \frac{1}{NQ_i} \sum_{i=1}^{N} \sum_{q=1}^{Q_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^{\#}_{iq}) \tag{4}$$

where $\underline{\mathbf{Y}}_i$ is the projection of $\underline{\mathbf{X}}_i$ by $\{\mathbf{W}^{(d)}\}_{d=1}^M$. $\underline{\mathbf{Y}}^*_{ip}$ is the p-th projection of $\underline{\mathbf{Y}}_i$ onto the corresponding TFL. Similarly, Q_i and $\mathbf{Y}^{\#}_{iq}$ represent corresponding parameters of the between-class FL distance. For the sake of simplicity, some notations are defined as:

$$J\{\underline{\mathbf{A}}, \underline{\mathbf{B}}; \mathbf{W}^{(d)}\} = \mathbf{W}^{(d)^T} [\![(\underline{\mathbf{A}} \bar{\times}_d \mathbf{W}^{(d)^T}) \otimes (\underline{\mathbf{B}} \bar{\times}_d \mathbf{W}^{(d)^T}); (\bar{d})(\bar{d})]\!] \mathbf{W}^{(d)}$$

$$L\{\underline{\mathbf{A}}, \underline{\mathbf{B}}; \mathbf{W}^{(d)}|_{d=1}^M\} = [\![\underline{\mathbf{A}} \Pi_{d=1}^M \times_d \mathbf{W}^{(d)^T} \otimes \underline{\mathbf{B}} \Pi_{d=1}^M \times_d \mathbf{W}^{(d)^T}; (1:M)(1:M)]\!]$$

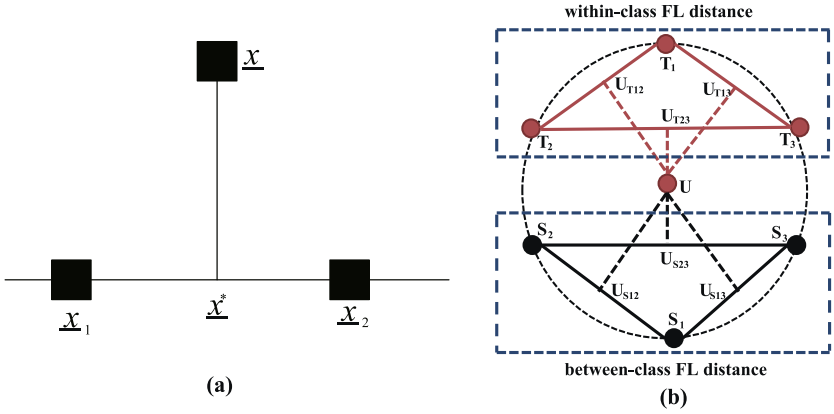


Fig. 2. (a) A brief view of tensor-based feature line (TFL). $\overline{\underline{\mathbf{X}}_1\underline{\mathbf{X}}_2}$ is the feature line and $\underline{\mathbf{X}}$ is a projection onto the feature line. (b) Concept of within-class FL distance and between-class FL distance. Different circles represent different training data. Red circles have the same label while the black ones have different label with red ones. Within-class FL distance is defined as the sum of the distance between data U and all its projections U_{T12}, U_{T13} and U_{T23}. Between-class FL distance is defined as the sum of the distance between data U and all its projections U_{S12}, U_{S13} and U_{S23}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where $\underline{\mathbf{A}}$ and $\underline{\mathbf{B}} \in R^{I_1 \times I_2 \times \ldots \times I_M}$ and $J(\underline{\mathbf{A}}, \underline{\mathbf{B}}; \mathbf{W}^{(d)}) \in R^{U_d \times U_d}$. Here the contracted product of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$ on all indices except the k-th index is denoted as $[\![\underline{\mathbf{X}} \otimes \underline{\mathbf{Y}}; (\bar{k})(\bar{k})]\!]$ (Tao et al., 2007).

The cost function defined in Eq. (4) is only determined by $dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip})$ and $dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^{\#}_{iq})$. Suppose $\underline{\mathbf{Y}}^*_{ip}$ is the projection of $\underline{\mathbf{Y}}_i$ onto the TFL formed by $\underline{\mathbf{Y}}_k$ and $\underline{\mathbf{Y}}_j$. Thus $\underline{\mathbf{Y}}^*_{ip}$ is expressed as

$$\underline{\mathbf{Y}}^*_{ip} = \underline{\mathbf{Y}}_j + t(\underline{\mathbf{Y}}_k - \underline{\mathbf{Y}}_j) = [\underline{\mathbf{X}}_j + t(\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j)] \prod_{d=1}^{M} \times_d \mathbf{W}^{(d)T} \tag{5}$$

where $\underline{\mathbf{Y}}_i$, $\underline{\mathbf{Y}}_j$ and $\underline{\mathbf{Y}}_k$ are the projections of $\underline{\mathbf{X}}_i$, $\underline{\mathbf{X}}_j$ and $\underline{\mathbf{X}}_k$ by the projection matrices. The position parameter t can be calculated by the following equation

$$\begin{aligned} t &= \frac{[\![(\underline{\mathbf{Y}}_i - \underline{\mathbf{Y}}_j) \otimes (\underline{\mathbf{Y}}_k - \underline{\mathbf{Y}}_j); (1:M)(1:M)]\!]}{[\![(\underline{\mathbf{Y}}_k - \underline{\mathbf{Y}}_j) \otimes (\underline{\mathbf{Y}}_k - \underline{\mathbf{Y}}_j); (1:M)(1:M)]\!]} \\ &= \frac{L\{(\underline{\mathbf{X}}_i - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}|_{d=1}^{M}\}}{L\{(\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}|_{d=1}^{M}\}} \end{aligned} \tag{6}$$

Then, we can get the FL distance $dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip})$ as follows:

$$\begin{aligned} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip}) &= [\![(\underline{\mathbf{Y}}_i - \underline{\mathbf{Y}}^*_{ip}) \otimes (\underline{\mathbf{Y}}_i - \underline{\mathbf{Y}}^*_{ip}); (1:M)(1:M)]\!] \\ &= f_1 + f_2 - 2f_3 + 2f_4 - 2f_5 + f_6 \end{aligned} \tag{7}$$

where $dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip})$ can be derived and decomposed into six different items of $f_n|_{n=1}^{6}$. Furthermore, f_1, f_2, f_3 can be expressed in a similar way and denoted as $f_{\langle 1;2;3\rangle}$. Similarly, f_4, f_5, f_6 can be written in $f_{\langle 4;5;6\rangle}$.[1]

$$f_{\langle 1;2;3\rangle} = L\{\langle \underline{\mathbf{X}}_i, \underline{\mathbf{X}}_i; \underline{\mathbf{X}}_j, \underline{\mathbf{X}}_j; \underline{\mathbf{X}}_i, \underline{\mathbf{X}}_j\rangle; \mathbf{W}^{(d)}|_{d=1}^{M}\}$$

$$\begin{aligned} f_{\langle 4;5;6\rangle} &= L\{\langle \underline{\mathbf{X}}_j; \underline{\mathbf{X}}_i; (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j)\rangle, (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}|_{d=1}^{M}\} \\ &\times \frac{L\{(\underline{\mathbf{X}}_i - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}|_{d=1}^{M}\}}{L\{(\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}|_{d=1}^{M}\}} \end{aligned}$$

However, from Eq. (7), it is clearly seen that M parameters need to be determined in one function, thus the problem defined in (4) does not have a closed form solution. Here the alternative least square (ALS) method is applied to to obtain a numerical solutions. Therefore, the optimization problem can be decomposed into M different optimization sub-problems. Here the six items $f_i, i = 1 \ldots 6$ can be derived to focus one variable when fixing the other ones [2]

$$f_{\langle 1;2;3\rangle}(\mathbf{W}^{(d)}) = trace(J\{\langle \underline{\mathbf{X}}_i, \underline{\mathbf{X}}_i; \underline{\mathbf{X}}_j, \underline{\mathbf{X}}_j; \underline{\mathbf{X}}_i, \underline{\mathbf{X}}_j\rangle; \mathbf{W}^{(d)}\})$$

$$\begin{aligned} f_{\langle 4;5;6\rangle}(\mathbf{W}^{(d)}) = &\frac{trace(J\{(\underline{\mathbf{X}}_i - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}\})}{trace(J\{(\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j), (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}\})} \\ &\times trace(J\{\langle \underline{\mathbf{X}}_j; \underline{\mathbf{X}}_i; \underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j\rangle, (\underline{\mathbf{X}}_k - \underline{\mathbf{X}}_j); \mathbf{W}^{(d)}\}) \end{aligned}$$

Now combing the above equations with (7), $dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip})$ can be derived as:

$$\begin{aligned} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip}) = &f_1(\mathbf{W}^{(d)}) + f_2(\mathbf{W}^{(d)}) - 2f_3(\mathbf{W}^{(d)}) + 2f_4(\mathbf{W}^{(d)}) \\ &- 2f_5(\mathbf{W}^{(d)}) + f_6(\mathbf{W}^{(d)}) \end{aligned} \tag{8}$$

Thus the optimal cost function (4) is decomposed into M different optimization sub-problems as follows. Here it is obvious that the optimal cost function is only dependent on one variable $\mathbf{W}^{(d)}$

$$\begin{aligned} f(\mathbf{W}^{(d)}) = &\underset{W^{(d)}}{argmin} \frac{1}{NP_i} \sum_{i=1}^{N} \sum_{p=1}^{P_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^*_{ip}; \mathbf{W}^{(d)}) \\ &- \frac{1}{NQ_i} \sum_{i=1}^{N} \sum_{q=1}^{Q_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}^{\#}_{iq}; \mathbf{W}^{(d)}) \end{aligned} \tag{9}$$

where $d = 1 \ldots M$.

In this paper, the gradient descent method is applied to solve the M minimization sub-problems as follows:

$$\mathbf{W}^{(d)}(t+1) = \mathbf{W}^{(d)}(t) - \alpha(t) \frac{\partial f(\mathbf{W}^{(d)})}{\partial \mathbf{W}^{(d)}} \tag{10}$$

where $\alpha(t)$ is the learning rate. From Eq. (8), we can find that $f(\mathbf{W}^{(d)})$ is composed of six items $f_i(\mathbf{W}^{(d)})|_{i=1}^{6}$. Therefore, we can compute

[1] For instance, $f_{\langle 1;2;3\rangle} = L\{\langle \underline{\mathbf{A}}_i, \underline{\mathbf{B}}_i; \underline{\mathbf{A}}_j, \underline{\mathbf{B}}_j; \underline{\mathbf{A}}_k, \underline{\mathbf{B}}_k\rangle; \mathbf{W}\}$ represents three functions $f_1 = L\{\underline{\mathbf{A}}_i, \underline{\mathbf{B}}_i; \mathbf{W}\}$; $f_2 = L\{\underline{\mathbf{A}}_j, \underline{\mathbf{B}}_j; \mathbf{W}\}$; $f_3 = L\{\underline{\mathbf{A}}_k, \underline{\mathbf{B}}_k; \mathbf{W}\}$. $f_{\langle 4;5;6\rangle} = L\{\langle \underline{\mathbf{A}}_i; \underline{\mathbf{A}}_j; \underline{\mathbf{A}}_k\rangle, \underline{\mathbf{B}}; \mathbf{W}\}$ represents three functions $f_4 = L\{\underline{\mathbf{A}}_i, \underline{\mathbf{B}}; \mathbf{W}\}$; $f_5 = L\{\underline{\mathbf{A}}_j, \underline{\mathbf{B}}; \mathbf{W}\}$; $f_6 = L\{\underline{\mathbf{A}}_k, \underline{\mathbf{B}}; \mathbf{W}\}$. Here $\langle . \rangle$ represents distributing element inside of $\langle . \rangle$ to element outside of $\langle . \rangle$

[2] Here, $f_{\langle 1;2;3\rangle} = trace(J\{\langle \underline{\mathbf{A}}_j, \underline{\mathbf{B}}_i; \underline{\mathbf{A}}_j, \underline{\mathbf{B}}_j; \underline{\mathbf{A}}_k, \underline{\mathbf{B}}_k\rangle; \mathbf{W}\})$ represents three functions $f_1 = trace(J\{\underline{\mathbf{A}}_i, \underline{\mathbf{B}}_i; \mathbf{W}\})$; $f_2 = trace(J\{\underline{\mathbf{A}}_j, \underline{\mathbf{B}}_j; \mathbf{W}\})$; $f_3 = trace(J\{\underline{\mathbf{A}}_k, \underline{\mathbf{B}}_k; \mathbf{W}\})$. $f_{\langle 4;5;6\rangle} = trace(J\{\langle \underline{\mathbf{A}}_i; \underline{\mathbf{A}}_j; \underline{\mathbf{A}}_k\rangle, \underline{\mathbf{B}}; \mathbf{W}\})$ represents three functions $f_4 = trace(J\{\underline{\mathbf{A}}_i, \underline{\mathbf{B}}; \mathbf{W}\})$; $f_5 = trace(J\{\underline{\mathbf{A}}_j, \underline{\mathbf{B}}; \mathbf{W}\})$; $f_6 = trace(J\{\underline{\mathbf{A}}_k, \underline{\mathbf{B}}; \mathbf{W}\})$. Here $\langle . \rangle$ represents distributing element inside of $\langle . \rangle$ to element outside of $\langle . \rangle$

each item $\partial f_i(\mathbf{W}^{(d)})/\partial \mathbf{W}^{(d)}|_{i=1}^{6}$ separately and then $\partial f(\mathbf{W}^{(d)})/\partial \mathbf{W}^{(d)}$ can be obtained by Eq. (9).

Algorithm 1 describes the alternating projection optimization procedure of TNFLD. The most discriminative patterns on each mode are obtained by steps 4 and 5. The algorithm will stop running if the convergence condition $\sum_{d=1}^{M} \|\mathbf{W}_{(t)}^{(d)} - \mathbf{W}_{(t-1)}^{(d)}\| < \varepsilon$ is satisfied.

Algorithm 1. Tensor-based nearest feature line distance algorithm

Input: The training dataset $\Omega = \{\underline{\mathbf{X}}_n^{I_1 \times I_2 \times \cdots \times I_M}, c_n\}_{n=1}^{N}$, $\underline{\mathbf{X}}_n^{I_1 \times I_2 \times \cdots \times I_M}$ denotes the n-th sample of the training dataset and c_n represents the corresponding class label, the learning rate α and the maximum number of iteration T.
Output: The set of optimal discriminative projection patterns on each mode $\{\mathbf{W}^{(d)}|_{d=1}^{M}\}$.
Initialization: Set $\{\mathbf{W}_0^{(d)} = 1|_{d=1}^{M}\}$
Method:
1: Obtain each combination pair of the training data $(\underline{\mathbf{X}}_i, \underline{\mathbf{X}}_j, \underline{\mathbf{X}}_k)$ with the same label and $(\underline{\mathbf{X}}_i, \underline{\mathbf{X}}_j, \underline{\mathbf{X}}_k)$ with different label between i and k, j.
2: For iteration $t = 1$ to T {
3: For mode $d = 1$ to M {
4: $f(\mathbf{W}_t^{(d)}) =$
$arg\ \min \frac{1}{NP_i}\sum_{i=1}^{N}\sum_{p=1}^{P_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}_{ip}^{*}; \mathbf{W}_t^{(d)}) - \frac{1}{NQ_i}\sum_{i=1}^{N}\sum_{q=1}^{Q_i} dis^2(\underline{\mathbf{Y}}_i, \underline{\mathbf{Y}}_{iq}^{\#}; \mathbf{W}_t^{(d)})$
5: Calculate the d-th optimal $\mathbf{W}_t^{(d)}$ in the t-th iteration using gradient descent method
6: }
7: break if convergence }

Then EEG $\underline{\mathbf{X}}$ is projected to $\underline{\mathbf{Y}}$ in a low dimensional space according to $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\prod_{d=1}^{M} \times_d \mathbf{W}^{(d)^T}$. Afterwards, $\underline{\mathbf{Y}}$ is concatenated into a feature vector y.

2.3. Fisher score

Fisher score is utilized to select the discriminative motor imagery EEG patterns from the multilinear subspace obtained by TNFLD in spatial–spectral–temporal domain. It is defined as

$$\text{Fisher score} = \frac{\|\mu_1 - \mu_2\|^2}{\sigma_1 + \sigma_2} \tag{11}$$

where μ_1, μ_2 denote means of the two classes over an individual feature, and σ_1, σ_2 represent corresponding variances.

Fisher score is computed for each individual feature in the vectorized features. Then features with n-largest Fisher scores are retained as the most discriminative features, and other features are discarded as redundant features. In this step, the most discriminative patterns are obtained from the corresponding projection matrices of the retained features.

2.4. SVM for classification

In this paper, SVM, which achieves high-level performance in many applications, is utilized for classification. A 5-fold cross-validation is used to choose suitable SVM parameters to predict the labels of the test datasets.

3. Experiments configuration

A promising BCI-FES rehabilitation system is developed for stroke patients to recover motor function. We have conducted clinical rehabilitation experiments using the BCI-FES system in cooperation with Shanghai Huashan Hospital. Eight stroke patients were recruited and randomized divided into BCI group ($n = 5$) and control group ($n = 3$) by the random number enclosed in the envelope. Patients in BCI group participated in BCI-FES rehabilitation training for 24 times in two months (three times per week), together with the regular therapies such as physical therapy, exercise therapy and acupuncture therapy. Patients in control group only received regular therapies. All the patients signed an informed consent form, and our study was approved by the Hushan Institutional Review Board. Our study fulfilled the clinical registration requirements, and the trial registration number was ChiCTR-TRC-12002541. Table 1 shows demographic information and motor functions scores for all patients in the two groups.

For BCI group, EEG is recorded by a 16-channel (FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier at a sampling rate of 256 Hz. During the whole experiment, subjects are required to take part in the motor imagery based tasks for three different courses: training model course, active rehabilitation training course and post-training model course. Mulder et al. (2004) reported that a familiar mental task often leaded to an improved motor imagery performance. Thus the motor imagery task assigned to the patients is to imagine drinking water. Some compound motions like palm flexion and arm stretch have been integrated into this task.

In the training model course, patients have to complete basic motor imagery related tasks for five sessions. Each session is consisted of forty trials and lasts for 240 s. At the beginning of each trial, a bold arrow with a command, randomly left or right, is shown on the screen, instructing patients to imagine left or right. The motor imagery procedure of one trial lasts for four seconds. After that, a two-second interval is given between trials during which the subjects can adjust their mental state and prepare for the following trials.

In the active rehabilitation training course, patients are asked to finish some virtual games covering different categories of daily missions, like balancing a beam or lifting balloons. For instance, in the lifting balloon game, patients are required to imagine left if the balloon appears in left part of the screen. When patients imagine correctly, visual and auditory feedback is generated by the interface; for example, the balloon is lifted by a virtual hand. Simultaneously, FES is triggered and delivered to patients' extensor carpus radialis muscles, which causes real movement of their

Table 1
Demographic information and motor functions scores of the upper limbs for all the patients in the BCI group and control group.

BCI group	S1	S2	S3	S4	S5	Overall	Control Group	S6	S7	S8	Overall
Age	65	65	67	74	67	66 ± 4.9	60	70	60	63 ± 5.8	
Gender	F	F	M	M	M	2F,	3M	M	F	M	1F,2M
Lesion side	R.C	R.BG	L.BG	R.BG	L.BG	4BG,1C	L.BG	R.BG	R.BG	All BG	
Hemiplegic	Limb	L	L	R	L	R	3L, 2R	R	L	L	2L, 1R
Handedness	R	R	R	R	R	All	R	R	R	R	All R
Stroke onset (Months)	2	2	1	1	2	1.6 ± 0.5	2	2	1	1.7 ± 0.6	
FMA (Full = 66)	Pre	22	16	16	11	12	15.4 ± 4.3	14	16	12	14.0 ± 2.0
	Post	31	29	25	18	22	25.0 ± 5.2	22	18	20	20.0 ± 2.0
ARAT (Full = 57)	Pre	0	0	2	0	0	0.4 ± 0.9	0	0	0	0 ± 0
	Post	10	34	19	11	23	19.4 ± 9.8	4	15	8	9.0 ± 5.6

Abbreviations: F = female; M = male; L = left; R = right; C = Cortex; BG = basal ganglia; FMA = Fugl-Meyer Motor Assessment; ARAT = Action Research Arm Test.

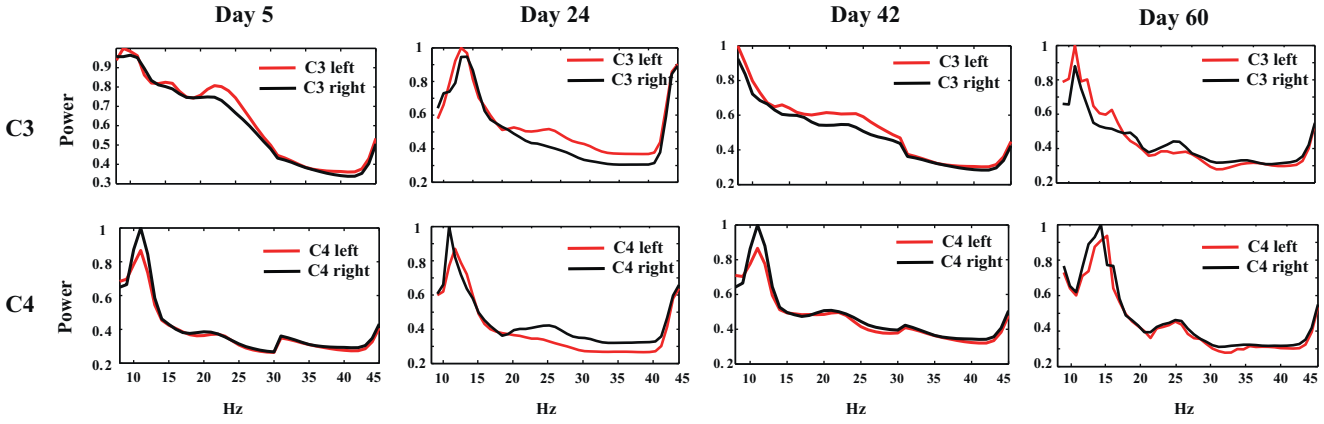


Fig. 3. Averaged power spectrum of channels C3 and C4 in bands 8–45 Hz for Subject 2 on day 5, 24, 42 and 60. The values on the *y* axis are normalized.

hands or arms. Detailed stimulation protocol of FES includes: Current with 25 mA, Pulse width with 150 μs, Frequency with 60 Hz, Rise time with 0.5 s, Fall time with 0.5 s and Stimulation time with 5 s.

In the post-training model course, subjects are required to finish the motor imagery related BCI tasks as described in the training model course. EEG in this course is used for rehabilitation efficacy assessment.

4. Results

For each patient, 100 left and 100 right trials are extracted from the training model course. The first 60 trials (30 trials for each class) are used as the training dataset and the other 140 trials (70 trials for each class) are taken as the test dataset. TNFLD is employed to obtain a multiway discriminative subspace. As the training dataset is not large, TNFLD can be assessed when faced with the small sample size problem. As the prior neurophysiology of stroke patients, especially the active cortex and frequency band related to motor imagery EEG, is not available, we filter the raw data into the given frequency band 8–45 Hz, which is expected to contain broad rhythms related to motor imagery. We evaluate our scheme in two aspects of motor imagery EEG patterns changes and online and offline recognition performance.

4.1. Motor imagery EEG patterns changes

In order to observe the spatial and spectral patterns of stroke patients, some traditional methods are first applied to obtain the motor imagery EEG patterns in spatial and spectral domains. In this paper, power spectrum density (PSD) is employed for spectral analysis. For each trial in the dataset (100 trials with left class and 100 trials with right class), PSD is calculated by a fast Fourier transform (length of data: 1024, sample rate: 256 Hz and frequency band: 8–45 Hz). After that, we average PSD of all the trials with the same class to get the mean PSD. In terms of spatial patterns, CSP tries to extract spatial-filtered matrix based on the 16-channel signals.

Lacking space for a detailed description of all the results of all subjects, we select EEG from day 5, 24, 42 and 60 to represent the different phases during rehabilitation. Different from healthy people, cortex regions responded to motor imagery might shift due to cortex reorganization. Without prior neurophysiology of the most important channels configuration, we observe the spectral information of patients' EEG in channels C3 and C4 which are the most important channels in healthy people (Pfurtscheller and Neuper, 1997). Fig. 3 illustrates the averaged power spectrum of C3 and C4 in the frequency bands 8–45 Hz for the second subject on the four chosen days. Subtle differences are found between the two tasks, especially around 24–25 Hz for C3 and 11–12 Hz for C4, but the difference is not so obvious. As illustrated in Fig. 4, the four spatial patterns extracted by CSP for the second subject on the four chosen days have no significance in discrimination.

Different from the traditional methods, spatial, spectral and temporal patterns are simultaneously obtained by TNFLD directly from the structural preserved multiway signals reconstructed by wavelet transform (16 channels; 8–45 Hz, step by 1 Hz; 1–4000 ms, step by 31.25 ms, that is the step by 8 points). By tracking the motor imagery EEG patterns during rehabilitation training, gradual changes of these patterns are observed over time. Fig. 5 presents the two discriminative spatial, spectral and temporal patterns with the two largest Fisher scores extracted by TNFLD for the second subject on the four chosen days. For the unaffected (left) hemisphere, the activated cortices are stable during the whole training, and mainly located in the left central lobe (C3). However, in the affected (right) hemisphere, larger cortical regions related to motor imagery are activated and shifted during rehabilitation. These regions gradually migrate from right central and parietal lobes (C4 + P4) to right central, frontal-central and parietal lobes (C4 + FC4 + P4), and finally to around central lobe (C4). Spectral characteristic are also observed over time. The most active frequency bands scatter at wide-ranged bands (10–25 Hz) at the beginning, but gradually concentrate on lower bands (10–18 Hz).

For comparison of performance, nonnegative multiway factorization (NMWF) (Mørup et al., 2006) is also utilized to extract motor

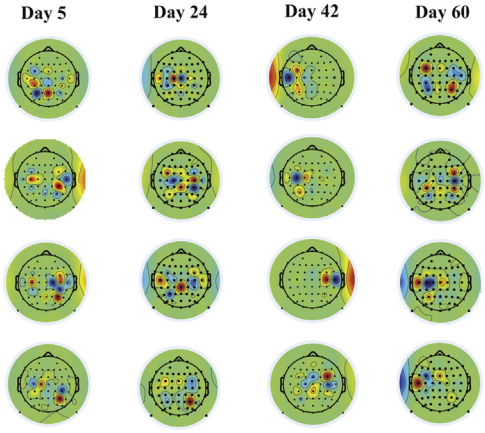


Fig. 4. The four most discriminative spatial patterns extracted by CSP for Subject 2 on day 5, 24, 42 and 60. Different columns represent different days while different rows stand for the four spatial patterns.

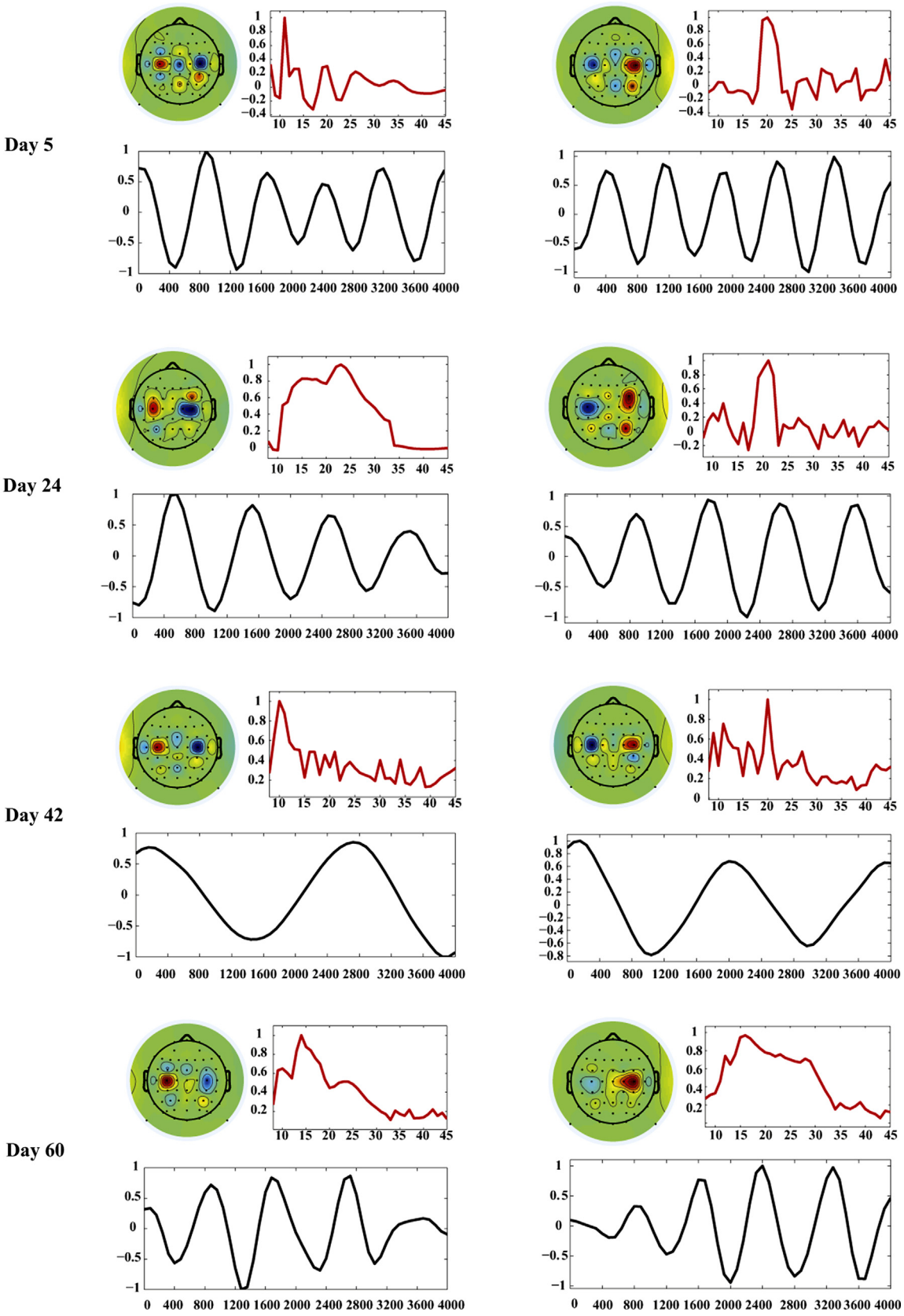


Fig. 5. The two most discriminative spatial, spectral and temporal patterns extracted by TNFLD using Fisher score strategy for Subject 2 on day 5, 24, 42 and 60. Left column represents the first discriminative patterns while right one shows the second discriminative patterns. Values in spectral and temporal patterns are normalized.

imagery EEG patterns in spatial–spectral–temporal domain. Fig. 6 shows two discriminative spatial, spectral and temporal patterns with two largest Fisher scores for Subject 2 on day 5, 24, 42 and 60. Different from the patterns obtained by TNFLD, the spatial patterns are chaotic and less distinctive on discrimination. Active frequency band scatters wide-ranged (10–30 Hz).

4.2. Recognition performance

4.2.1. Offline evaluation

We evaluate the classification performance of TNFLD with other five methods of CSP, regularized CSP (RCSP) (Lotte and Guan, 2011), wavelet transform method (WT), PSD and NMWF on the test dataset. For each trial, a fast Fourier transform is employed to calculate PSD features for the 16-channel one-way EEG data (*time*), as described in the subsection 'motor imagery EEG patterns changes'. With respect to CSP and RCSP, they are utilized to extract discriminative features for the 16-channel two-way EEG data (*channel* × *time*). The raw data for CSP and RCSP is filtered in a given frequency band 8–45 Hz, as the frequency band related to motor imagery for stroke patients is not available in advance. The other methods TNFLD, NMWF and WT are applied to extract features from multiway EEG data (*channel* × *frequency* × *time*). In our experiment, the third-order tensors are obtained by the wavelet transform method in the general spatial–spectra–temporal range (16 channels; 8–45 Hz, step by 1 Hz; 1–4000 ms, step by 31.25 ms, that is the step by 8 points). After that, the six methods are applied to extract features of EEG data. We calculate Fisher score for each feature as described in the Fisher score subsection, and select n ($n = 16$ in this paper) features corresponding to n largest Fisher scores. Then the n features are fed to a SVM classifier for parameters adjustment. The optimal numbers of features are determined and reduced to 2–10 according to the training performance, as more features cannot improve the training accuracy. Finally, the trained SVM classifier is utilized on the test dataset to evaluate their performance. Classification accuracies in the same week are averaged to represent the mean accuracy of the week. After that, those mean accuracies of six chosen weeks are averaged to represent the mean accuracy of the corresponding method.

Fig. 7 shows the mean accuracies of all the methods under different feature dimensionalities. For all the subjects, the accuracies of all the methods change greatly with the increase of feature dimensionality. Another important observation is that TNFLD and NMWF outperform other methods, and TNFLD has a better performance than NMWF.

Fig. 8 lists the classification results of all the algorithms on five subjects for the six weeks. According to the Mann–Whitney U test results between TNFLD and any other compared method, the better recognition accuracy for TNFLD is significant at the significance level of 0.05. The classification accuracies by TNFLD for almost all the subjects can reach $73.5 \pm 2.9\%$.

4.2.2. Simulated online evaluation

The online computer simulation is also studied in order to evaluate the performance of TNFLD when integrated in the BCI-FES system. All trials in the training model course are extracted for obtaining the most discriminative patterns by Fisher score in spatial–spectral–temporal domain. By searching the maximum absolute weights in the extracted spatial and spectral patterns, the most contributed channel groups and frequency bands are selected in the discriminative patterns and utilized as prior knowledge. For the sake of the instantaneity of online systems, we adopt the sliding window strategy to output recognition result every 0.125 s (Li and Zhang, 2010). Features extracted by CSP based on the prior knowledge in the training model course are fed to a SVM classifier for training. Afterwards, based on the prior knowledge, testing segments are generated by sliding window strategy in the rehabilitation course and used for evaluating the online performance of TNFLD. Accuracies in each experiment are calculated and the mean accuracies of the same six weeks chosen in offline evaluation are computed finally. Table 2 lists the online accuracies of the original CSP-SVM method and the CSP-SVM method combined with TNFLD (TNFLD-CSP-SVM) in the chosen six weeks for all the subjects. As expected, the recognition accuracy by TNFLD-CSP-SVM can reach $70.5 \pm 4.3\%$. According to a Mann–Whitney U test result between TNFLD-CSP-SVM and CSP-SVM, the better recognition accuracy for TNFLD-CSP-SVM is significant at the significance level of 0.05.

5. Discussion

This study presents a tensor-based scheme to obtain discriminative motor imagery EEG patterns in spatial–spectral–temporal domain and extract discriminative features with largest Fisher scores. We evaluate the performance and effect of the scheme on the BCI-FES system from two aspects of motor imagery EEG patterns changes and online and offline recognition performance. In-depth discussions about these results are conducted as follows.

5.1. Motor imagery EEG patterns changes

Due to the brain reorganization in motor areas of stroke patients, motor imagery EEG patterns may have gradually changed during rehabilitation and deviated from those of healthy people, leading to a poor performance of BCI-based systems. Most traditional methods detect motor imagery EEG patterns only in one-way (like PSD) or two-way (like CSP) EEG, drawing from prior knowledge of healthy subjects like frequency band and channels configuration which are different from stroke patients. Our results suggest that discriminative spatial–spectral–temporal patterns can be simultaneously decoded directly from multiway EEG by our proposed TNFLD method without prior knowledge. Compared with other multiway analysis methods like NMWF, TNFLD also takes class label information into consideration and can obtain more discriminative patterns.

By tracking the gradual changes of motor imagery EEG patterns during rehabilitation, a compensatory mechanism about the cortex reorganization is revealed. Our results suggest that larger regions in the contralateral hemisphere are activated. The importance of sensorimotor area (C4) and its neighboring areas like frontal premotor lobe and parietal lobe on the affected hemisphere increases over time. Detailed, at the beginning, parietal lobes (P4) and frontal premotor lobes (FC4) play a compensatory role but the importance decreases over time. Active areas gradually shift back to central lobes (C4). This interesting observation reveals rehabilitation mechanism to us: as a sign of cortex reorganization in the affected hemisphere, larger regions initially take parts of functionality but gradually given back during recovery. The same phenomenon are also reported in some other studies (Feydy et al., 2002). Spectral characteristics are also studied and explored over time. Active frequency bands decentralize at wide-ranged bands (10–25 Hz) at the beginning, but gradually concentrated to narrower bands (10–18 Hz). The dynamic migration on frequency bands implies that active rhythms are modulated during rehabilitation, as also reported in Shahid et al. (2010).

Some studies also observed the pattern changes on the stroke patients after six months of stroke onset. Different from our findings of the patients diseased in two months, Tam et al. (2011) reported that the motor imagery EEG patterns were relatively stable after six months of stroke onset. Moreover, contralateral activation were revealed in some patients. Detailed, active patterns were produced in premotor area (like FC4) and the parietal area (like P4) instead of sensorimotor area, which is also observed by our

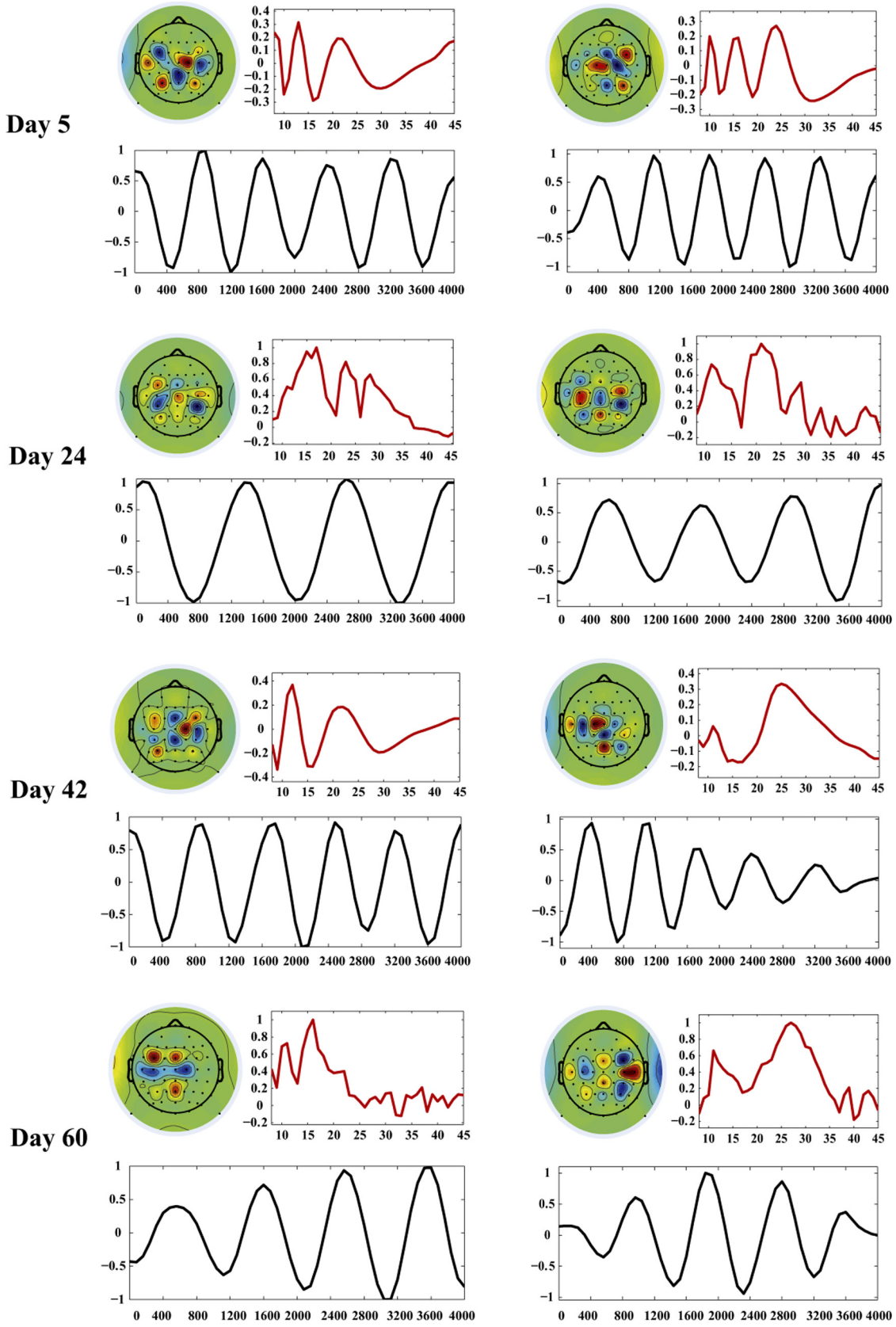


Fig. 6. The two most discriminative spatial, spectral and temporal patterns extracted by NMWF using Fisher score strategy for Subject 2 on day 5, 24, 42 and 60. Left column represents the first discriminative patterns while right one shows the second discriminative patterns. Values in spectral and temporal patterns are normalized.

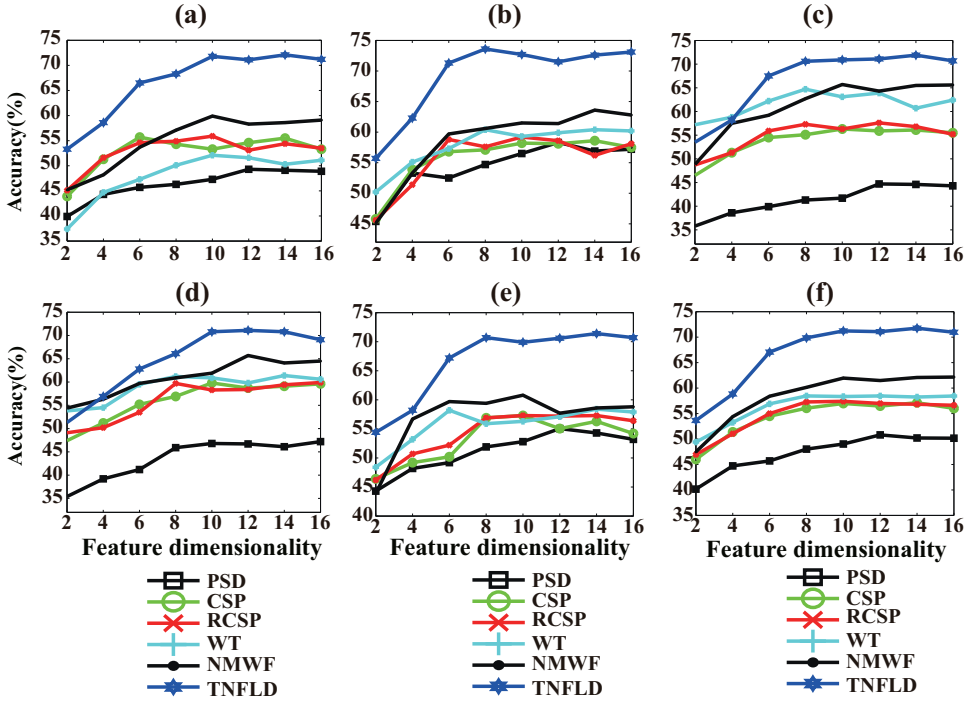


Fig. 7. A comparison in classification accuracies of all the the methods under different feature dimensionalities. The horizontal axis represents the varying feature dimensionality while the vertical axis shows the mean accuracies. (a) Subject 1. (b) Subject 2. (c) Subject 3. (d) Subject 4. (e) Subject 5. (f) Group mean.

study. Differently, ipsilesional activation were also demonstrated in most of their patients, and three of the patients showed activation on the contralesional side, which may be due to the unmasking of the ipsilateral pathway. These findings about cortex reorganization may supply additional rehabilitation mechanism and give an important contribution to biomedical rehabilitation engineering.

5.2. Classification performance

The results prove that TNFLD outperforms some traditional methods, like CSP, PSD, RCSP, WT and NMWF, in offline recognition performance. Furthermore, the most contributed channel groups and frequency bands are produced in the discriminative patterns

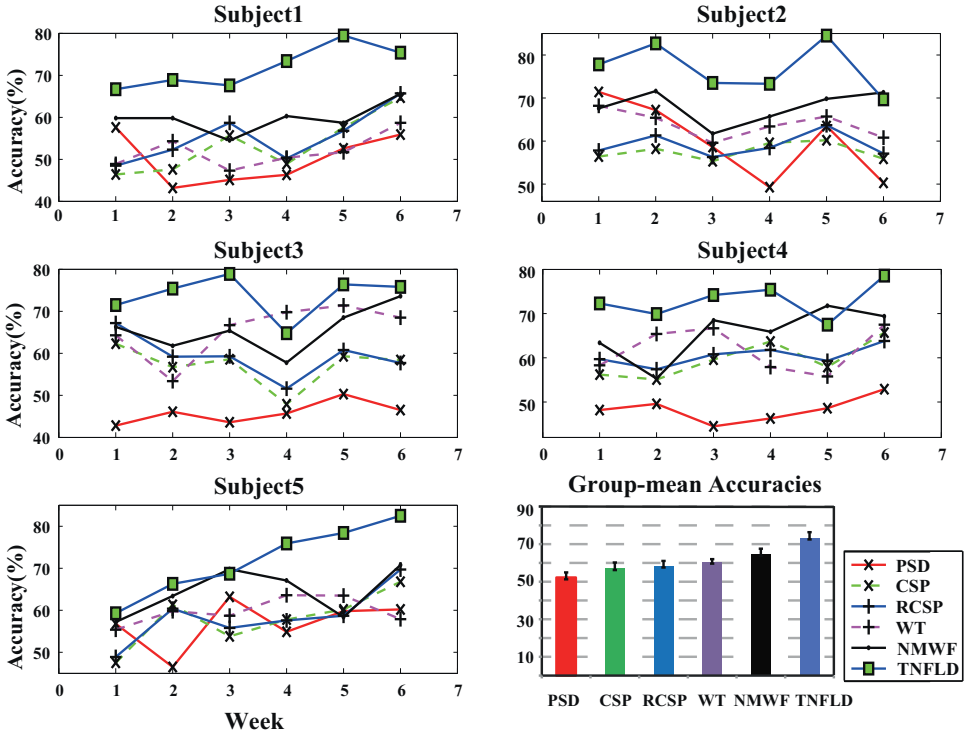


Fig. 8. A comparison in classification accuracies of TNFLD and some other algorithms. The horizontal axis represents the chosen six weeks of two months while the vertical axis shows the mean accuracy of every week on the test dataset. The first five subfigures represent accuracies of the five stroke subjects and the last subfigure represents the group-mean accuracies of the five subjects.

Table 2
Online mean accuracies of the six chosen weeks for all the subjects by CSP-SVM and TNFLD-CSP-SVM (Overall: Group-mean accuracy of all the patients for CSP-SVM is 61.2 ± 4.8%, group-mean accuracy of all the patients for TNFLD-CSP-SVM is 70.5 ± 4.3%. Selected channels: 3 for each hemisphere. Frequency band: frequencies with the 20-largest absolute weights. Feature dimensionality: 6.

		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Subject1	CSP-SVM	47.8	52.5	53.6	58.7	58.9	67.3
	TNFLD-CSP-SVM	58.2	61.8	65.6	67.5	71.3	69.4
Subject2	CSP-SVM	58.9	60.5	63.7	65.8	66.6	67.5
	TNFLD-CSP-SVM	66.3	72.4	69.8	73.6	79.5	75.5
Subject3	CSP-SVM	62.4	57.4	60.3	59.6	65.7	63.5
	TNFLD-CSP-SVM	67.2	72.9	77.4	68.8	76.3	75.1
Subject4	CSP-SVM	53.5	55.4	58.3	63.6	66.8	72.4
	TNFLD-CSP-SVM	64.4	66.2	63.5	70.1	69.6	79.7
Subject5	CSP-SVM	52.4	63.6	59.3	62.6	65.5	70.6
	TNFLD-CSP-SVM	63.4	71.9	68.7	72.7	75.3	80.8

chosen by Fisher score and utilized as prior knowledge for the following rehabilitation training course. Combined with these prior knowledge, the online recognition performance of TNFLD-CSP-SVM is obviously higher than that of CSP-SVM. It guarantees instantaneity and effectiveness of the BCI-FES system and makes the most innovative contribution to the BCI-based systems.

To explore the reasons, we exploit some interesting advantages of TNFLD: first, superior to one-way and two-way EEG analysis methods (PSD and CSP, respectively), TNFLD extracts multilinear discriminative features for classification by multiway discriminative subspace, which greatly improves significant interpretations of EEG classification in the spatial–spectral–temporal domain. Second, TNFLD expands the capacity of the available database by generating reasonable virtual data and then solves the small sample size problem. Third, TNFLD never depends on some unavailable neurophysiological knowledge of patients, like pre-identified frequency band and channels configuration, leading to a higher performance than CSP. Finally, different from WT method, TNFLD preserves structural information of EEG data in the process of projection of the original data, and simultaneously takes class label information into consideration which makes TNFLD a better classification method than the multiway decomposition method NMWF. One interesting issue has to be discussed that some methods can also obtain the prior knowledge like frequency bands (Higashi and Tanaka, 2012) and channels configuration (Tam et al., 2011) instead of our scheme. One point is that these methods cannot obtain these prior knowledge simultaneously. We simultaneously integrate the 'best' channel group with active frequency bands to improve the recognition performance of BCI-based rehabilitation system. Another problem is that these methods require collecting more EEG signals for model training, resulting in a time-consuming task. Differently, our scheme explores the prior knowledge only from a small datasets.

5.3. Motor function recovery

In order to evaluate the effect of our BCI-FES rehabilitation system on motor function recovery of stroke patients, motor function of upper limb for each patient in the BCI group and control group is evaluated with Action Research Arm Test (ARAT) (Lyle, 1981) and Fugl-Meyer Motor Assessment (FMA) (Fugl-Meyer et al., 1975). Table 1 lists the motor functions scores before and after the training. The Wilcoxon signed-rank test for two related samples is used to evaluate differences in the rehabilitation index scores. Significance level is set at 0.05. Wilcoxon tests reveal there is no significant difference between the groups for the FMA and ARAT scores at the pre-test ($p < 0.05$), indicating that both groups have similar motor function to begin with. An independent Wilcoxon test for the respective change of each mean FMA and ARAT score at the post-test reveal that there is a significant difference between the groups ($p < 0.05$). The results show that motor function in BCI group has greater improvement than that in control group.

The results of motor function evaluated by FMA and ARAT scores suggest that a significant improvement of motor functions of upper limbs has been achieved in the BCI group, revealing an active effect of the BCI-FES system on stroke patients' rehabilitation.

6. Conclusion

In this paper, a tensor-based scheme is proposed to detect the motor imagery EEG patterns of stroke patients in spatial–spectral–temporal domain, and reveal a compensatory mechanism about cortex reorganization through tracking gradual changes of these patterns over time. The scheme attempts to produce the most contributed channel group and active frequency bands as the prior knowledge for the following rehabilitation training course. We evaluate our proposed scheme based on EEG datasets recorded from stroke patients. The results show that our method outperforms other traditional methods in both online and offline recognition.

Acknowledgments

This work was supported by the NSFC-JSPS International Cooperation Program (Grant No. 61111140019), the National Natural Science Foundation of China (Grant No. 91120305), the 12th Five-year Plan supporting 480 project of Ministry of Science and Technology of the People's Republic of China (No. 2013BAI10B03), and the Key Projects of Shanghai Science and Technology on Biomedicine (No. 10DZ1950800).

References

Buch E, Weber C, Cohen LG, Braun C, Dimyan MA, Ard T, et al. Think to move: a neuromagnetic brain–computer interface (BCI) system for chronic stroke. Stroke 2008;39(3):910–7.
Calautti C, Leroy F, Guincestre JY, Baron JC. Dynamics of motor network overactivation after striatocapsular stroke a longitudinal PET study using a fixed-performance paradigm. Stroke 2001;32(11):2534–42.
Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2011;2(3):27.
Chien J, Wu C. Discriminant waveletfaces and nearest feature classifiers for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2002;24(12):1644–9.
Crosbie J, McDonough S, Gilmore D, Wiggam M. The adjunctive role of mental practice in the rehabilitation of the upper limb after hemiplegic stroke: a pilot study. Clinical Rehabilitation 2004;18(1):60–8.
Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 2004;134(1):9–21.
Feydy A, Carlier R, Roby-Brami A, Bussel B, Cazalis F, Pierot L, et al. Longitudinal study of motor recovery after stroke recruitment and focusing of brain activation. Stroke 2002;33(6):1610–7.

Fugl-Meyer A, Jääskö L, Norlin V. The post-stroke hemiplegic patient. II. Incidence, mortality, and vocational return in göteborg, sweden with a review of the literature. Scandinavian Journal of Rehabilitation Medicine 1975;7(2):73.

Han J, Bhanu B. Statistical feature fusion for gait-based human recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2; 2004. p. p. II-842.

Higashi H, Tanaka T. Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification. IEEE Transactions on Bio-Medical Engineering 2012;60:1100–10.

Hill NJ, Lal TN, Schröder M, Hinterberger T, Widman G, Elger CE, et al. Classifying event-related desynchronization in EEG, ECoG and MEG signals. In: Pattern Recognition. Springer; 2006. p. 404–13.

Li J, Zhang L. Bilateral adaptation and neurofeedback for brain computer interface system. Journal of Neuroscience Methods 2010;193(2):373–9.

Li S, Chan K, Wang C. Performance evaluation of the nearest feature line method in image classification and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22(11):1335–9.

Lotte F, Guan C. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. IEEE Transactions on Biomedical Engineering 2011;58(2):355–62.

Loubinoux I, Carel C, Pariente J, Dechaumont S, Albucher JF, Marque P, et al. Correlation between cerebral reorganization and motor recovery after subcortical infarcts. Neuroimage 2003;20(4):2166.

Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. International Journal of Rehabilitation Research 1981;4(4):483–92.

Mørup M, Hansen L, Parnas J, Arnfred S. Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization. Technical University of Denmark Technical Report; 2006.

Mulder T, Zijlstra S, Zijlstra W, Hochstenbach J. The role of motor imagery in learning a totally novel movement. Experimental Brain Research 2004;154(2):211–7.

Murase H, Nayar SK. Visual learning and recognition of 3-D objects from appearance. International Journal of Computer Vision 1995;14(1):5–24.

Page S, Levine P, Leonard A. Mental practice in chronic stroke results of a randomized, placebo-controlled trial. Stroke 2007;38(4):1293–7.

Pfurtscheller G, Müller-Putz GR, Pfurtscheller J, Rupp R. EEG-based asynchronous BCI controls functional electrical stimulation in a tetraplegic patient. EURASIP Journal on Applied Signal Processing 2005 2005:3152–5.

Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in humans. Neuroscience Letters 1997;239(2–3):65.

Ramoser H, Muller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Transactions on Rehabilitation Engineering 2000;8:441–6.

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

Tam WK, Tong Ky, Meng F, Gao S. A minimal set of electrodes for motor imagery BCI to control an assistive device in chronic stroke subjects: a multi-session study. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2011;19(6):617–27.

Tao D, Li X, Wu X, Maybank S. General tensor discriminant analysis and gabor features for gait recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007;29(10):1700–15.