A Prior Neurophysiologic Knowledge Free Tensor-Based Scheme for Single Trial EEG Classification

Jie Li, Liqing Zhang, Dacheng Tao, Han Sun, and Qibin Zhao

Abstract—**Single trial electroencephalogram (EEG) classification is essential in developing brain–computer interfaces (BCIs). However, popular classification algorithms, e.g., common spatial patterns (CSP), usually highly depend on the prior neurophysiologic knowledge for noise removing, although this knowledge is not always known in practical applications. In this paper, a novel tensor-based scheme is proposed for single trial EEG classification, which performs well without the prior neurophysiologic knowledge. In this scheme, EEG signals are represented in the spatial-spectral-temporal domain by the wavelet transform, the multilinear discriminative subspace is reserved by the general tensor discriminant analysis (GTDA), redundant indiscriminative patterns are removed by Fisher score, and the classification is conducted by the support vector machine (SVM). Applications to three datasets confirm the effectiveness and the robustness of the proposed tensor scheme in analyzing EEG signals, especially in the case of lacking prior neurophysiologic knowledge.**

Index Terms—**Electroencephalogram (EEG), general tensor discriminant analysis (GTDA), single trial classification.**

I. INTRODUCTION

ELECTROENCEPHALOGRAM (EEG) is a measurement of electrical activity in the brain collected using noninvasive electrodes attached to the scalp. Over the years, EEG has been applied in a very wide variety of clinical and research contexts. Now, in computing, it is the most exploited sensory signal in brain computer interface (BCI) which aims to use brain activity to translate human intention and provide a new direct communication channel between brain and computer. The potential applications of BCI are vast and range from an interesting gadget in computer games to a useful tool for persons with severe motor disabilities. A number of EEG-based BCI systems have been developed recently [14], [21]–[23], [29] in which patterns of EEG in different mental states can be discriminated for information transmission by feature extraction and classification algorithms. Research [1], [8] has shown that their effectiveness and efficiency depend on the quality of EEG feature representation and the accuracy of pattern classification of the recorded single trial EEG.

In a BCI system, the subject is required to perform different tasks according to predefined mental control paradigms, which would induce the biofeedback based on specific responses to stimulus or event-related rhythm modulation, e.g., P300 speller paradigm [7], self-regulation of rhythm [22], and motor imagery [2], [26], and then the subject's intention is conveyed by the pattern changes of the recorded EEG. The most commonly used mental control paradigm in BCI is the motor imagery. This is because the motor imagery produces the attenuation of brain oscillatory activity within particular frequency bands over sensorimotor cortex [event-related desynchronizations (ERD)] [12], and depending on the part of the body imagined moving, the recorded EEG exhibits a distinctive pattern.

One of the most successful algorithms for single trial EEG classification, evidenced by the 2003 BCI Competition [9], is termed the common spatial patterns (CSP) [16]. CSP detects the spectral discriminations between two classes of tasks by calculating discriminative spatial patterns that maximize the variance of one class and at the same time minimize the variance of the other, wherein the variance of the band-pass filtered EEG signals directly reflects the spectral power of the band frequency. For the classification of two classes of motor imageries, CSP achieves the accuracy above 90% on single trials EEG samples [16]. Therefore, most existing BCI systems [2], [15] use the CSP algorithm to characterize EEG patterns achieving reasonable results in online discrimination of the motor imagery task. Extensions of CSP for the multiclass classification [11] and integrations with other forms information [10], [27], [35] have also received increasing attention recently.

Although the CSP algorithm proves to be highly successful, it is not optimized for the EEG classification problem. The performance of CSP severely depends on the prior neurophysiologic knowledge for noise removing, i.e., the preprocessing procedure of the temporal filtering, because the CSP algorithm calculates discriminative spatial patterns based on the temporal variances of signals. Only having the EEG signals band-pass filtered through the pre-identified frequency domain, high or low signal variances could reflect a strong (enhanced) or a weak (attenuate) rhythmic activity respectively [32]. For the improvements

Manuscript received March 13, 2008; revised July 02, 2008; accepted September 26, 2008. Current version published April 08, 2009. This work was supported in part by the Nanyang Technological University Start-Up Grant under Project M58020010, in part by the National High-Tech Research Program of China under Project 2006AA01Z125, in part by the National Natural Science Foundation of China under Project 60775007, and in part by the Shanghai Committee of Science and Technology under Project 08511501700.

J. Li, L. Zhang, H. Sun, and Q. Zhao are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: lijie1216@sjtu.edu.cn; zhang-lq@cs.sjtu.edu.cn; jeremiah@sjtu.edu.cn; qbzhao@sjtu.edu.cn).

D. Tao is with the School of Computer Engineering, Nanyang Technological University, 639798, Singapore (e-mail: dctao@ntu.edu.sg).

Color versions of one or more figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNSRE.2008.2008394

of the information transfer rate and robustness of BCI, development of new paradigms in BCI is necessary. However, if the spectral characteristics in new paradigms are not well known as the motor imagery task (i.e., for some tasks, the prior neurophysiologic knowledge is not available), CSP fails to work. It is demanding to design a robust and effective algorithm for the single trial EEG classification.

In this paper, a tensor-based scheme is proposed to classify single trial EEG signals, which are represented by high order tensors (multiway arrays) i.e., multiple-modality patterns in the spatial-spectral-temporal domain. The multiple-modality patterns are constructed by decomposing EEG signals based on the wavelet transform. The general tensor discriminant analysis (GTDA) [4] is applied to reserve multilinear discriminative subspace from the training tensors and thus high-dimensional tensors are mapped to low-dimensional tensors. Fisher score [31] is then utilized to further remove the discriminatively redundant features from vectorized low dimensional tensors, and select most significant discriminative patterns from the multilinear subspace. Finally, support vector machine (SVM) [3] is conducted for classification in the feature space with the reduced dimension. Empirical studies demonstrate the proposed scheme outperforms power spectrum density (PSD), CSP, and nonnegative multiway factorization (NMWF), and works well without prior neurophysiologic knowledge.

Before we end the introduction section, it is worth reviewing briefly some popular tensor models developed in recent years, i.e., PARAFAC [30], Tucker model [25], nonnegative multiway/tensor factorization (NMWF/NTF) [28], streaming tensor analysis [37], tensor subspace analysis (TSA) [34], 2-D linear discriminant analysis (2DLDA) [24], GTDA [4], Bayesian tensor analysis [33], tensor locally linear embedding [39], and supervised tensor learning [36]. In this paper, GTDA is applied for discriminative multilinear subspace extraction because 1) the discriminative information reserved in the training tensors is preserved, whereas PARAFAC, Tucker model, NMWF/NTF, and TSA do not and 2) GTDA provides stable recognition accuracies due to the converged alternating projection training algorithm to obtain a solution of GTDA, whereas that of 2DLDA does not. GTDA decomposes tensors into core tensors and a series of discriminative matrices (linear subspaces) by maximizing the projected between-class variance and minimizing the projected within-classes variance over every modality. Therefore, GTDA can be employed here to construct multilinear discriminative subspace from high dimensional and high order tensors.

The remainder of the paper is organized as follows. Section II presents the proposed scheme. Section III describes three datasets from different types of paradigms. Section IV demonstrates the experiments and results to verify the effectiveness of the proposed scheme. Section V offers our conclusion.

II. Tensor-Based Scheme for the EEG Classification

In this section, we first present the proposed tensor scheme, and then detail descriptions of all components, i.e., wavelet transform based EEG signal representation, GTDA for multilinear subspace selection, Fisher score for feature redundancy elimination, and support vector machine (SVM) for classification.

A. Tensor-Based Scheme

This part presents the proposed tensor-based scheme for single trial EEG classification. Fig. 1 illustrates the proposed scheme, which consists of five components: first, 62 channels of EEG signals are collected by an ESI-128 Channel High-Resolution EEG/EP Systems (SynAmps2, Neuroscan, Charlotte,NC); second, the obtained EEG signals are decomposed by the wavelet transform and represented in the spatial-spectral-temporal domain as high-dimensional third-order tensors; third, GTDA is applied to reserve multilinear discriminative subspace from the training tensors and thus high-dimensional tensors are mapped to low-dimensional tensors; fourth, the vectorization operation is applied to these low-dimensional tensors and the Fisher score is utilized to eliminate redundant indiscriminative patterns from discriminative subspace; and finally, SVM is trained for precise classification.

B. EEG Signal Representation

In the proposed scheme, EEG signals are represented as high-order tensors in the spatial-spectral-temporal domain. For a two-way (channel$\times$time) EEG epoch sample $X_{(c,t)}$ at channel c and time t, the third-order (channel$\times$frequency$\times$time) tensor $X_{(c,f,t)}$ at channel c, frequency f and time t is given by the amplitude of the convolution with a wavelet function $w_{(t,f)}$

$$\mathbf{X}_{(c,f,t)} = \|w_{(t,f)} * X_{(c,t)}\|. \tag{1}$$

In this work, we select the complex Morlet wavelet, $\psi_{(t)} = (1)/(\sqrt{\pi\sigma})\exp(2i\pi\Omega t)\exp(-t^2/\sigma)$ (with the center frequency $\Omega = 1$, and the bandwidth parameter $\sigma = 2$) as the mother wavelet, since it has been well applied to the analysis of the temporal development of the frequency of EEG signals [28].

C. GTDA

In this scheme, GTDA is applied for discriminative multilinear subspace extraction. Let $\mathbf{X}_{i;j}$ denote the jth $(1 \le j \le N_i)$ training sample (tensor) in the ith $(1 \le i \le C)$ individual class. Totally, there are $N = \sum_{(i=1:C)} N_i$ training samples. $\mathbf{M}_i = (1/N_i)\sum_{j=1}^{N_i} \mathbf{X}_{i;j}$ is the mean tensor of the samples in the ith class, $\mathbf{M} = (1/N)\sum_{i=1}^{C} N_i \mathbf{M}_i$ is the mean tensor of all training samples, and U_k denotes the kth modality projection matrix for decomposition in the training procedure. Moreover, $\mathbf{X}_{i;j}|_{1\le i\le C}^{1\le j\le N_i}$, $\mathbf{M}_i|_{i=1}^{C}$, and $\mathbf{M}$ are all M-order tensors in $R^{L_1\times L_2\times\cdots\times L_M}$. In GTDA [4], the optimal projection matrices U_l^* is given by

$$U_l^*|_{l=1}^{M} = \underset{U_k^T U_k = I|_{k=1}^{M}}{\arg\max} \left(\begin{array}{l} \sum_{i=1}^{C} N_i \left\| (\mathbf{M}_i - \mathbf{M}) \prod_{k=1}^{M} \times_k U_k^T \right\|_{\mathrm{Fro}}^2 \\ -\zeta \sum_{i=1}^{C}\sum_{j=1}^{N_i} \left\| (\mathbf{X}_{i;j} - \mathbf{M}_i) \prod_{k=1}^{M} \times_k U_k^T \right\|_{\mathrm{Fro}}^2 \end{array} \right) \tag{2}$$

where ξ is a tuning parameter and is automatically selected during the training procedure according to [4].

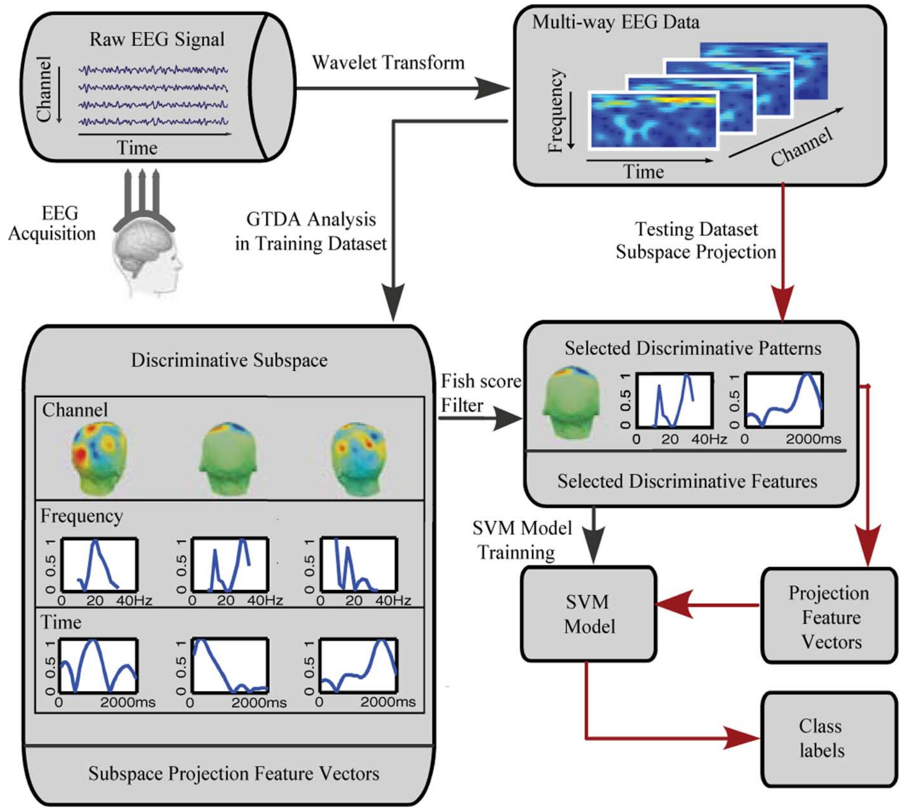


Fig. 1. The tensor-based scheme for single trial EEG classification. The black line stands for the procedure used in the training stage, and the red line stands for the testing procedure.

GTDA has the following advantages: 1) GTDA treats each modality of training tensors independently to tackle the small sample size problem in discriminative subspace selection, 2) GTDA preserves the discriminative information in the training tensors by taking the class label information into account, and 3) the optimization algorithm (the training stage) of GTDA is convergent. When the tensor representations of EEG data are processed by GTDA, the discriminative subspace constructed of a series of projection matrices U_l can be obtained. Big size tensors $\mathbf{X}_{(c,f,t)}$ are mapped to small size tensors $\mathbf{G}_{(c,f,t)}$ according to $\mathbf{G} = \mathbf{X}\prod_{d=1}^{M} \times_d U_d^T$, and $\mathbf{G}_{(c,f,t)}$ is then concatenated into a vector $\mathbf{g}_{(c,f,t)}$.

D. Fisher Score

After obtaining feature vectors $\mathbf{g}_{(c,f,t)}$ in the previous step, Fisher score is utilized to further remove redundant features, and select most significant discriminative patterns. Fisher score [31] is a discriminative measure of an individual feature for binary classification tasks. It is defined as

$$\text{Fisher score} = \frac{\|\mu_1 - \mu_2\|^2}{\sigma_1 + \sigma_2} \tag{3}$$

where μ_1 and μ_2 denote the means of class 1 and class 2 over an individual feature, and σ_1 and σ_2 denote corresponding variances.

For each individual feature, its fisher score is computed, and then features with the n-largest Fisher scores are retained as the most significant features, and other features are discarded as the discriminatively redundant features. In this step, the most significant discriminative patterns are obtained from the corresponding projection matrices of the retained features. It is worth emphasizing that other dimensionality reduction algorithms [3], [5], [17], [18], [40]–[42] can replace the Fisher score to achieve different advantages

E. SVM for Classification

SVM [38] obtains top-level performance in many applications, e.g., BCI [35], because it has good generalization ability in minimizing the VC dimension and achieving a minimal structural risk [38]. In this scheme, the Gaussian radial basis function (RBF) is used as the kernel function and a five-fold cross-validation is utilized to choose suitable SVM parameters to predict the labels of new samples in the testing phase.

III. Data Acquisition

To test the effectiveness and the robustness of the proposed tensor-based scheme, we assembled three datasets (Datasets 1–3) which collected EEG data from different subjects performing different mental tasks according to various mental control paradigms. Totally, nine healthy male subjects, aged from 21 to 30, took part in the experiments. 62 channels of EEG signals were recorded by an ESI-128 Channel High-Resolution EEG/EP Systems (SynAmps2, Neuroscan) at Laboratory for Brian-like Computing and Machine Intelligence, Shanghai Jiao Tong University, Shanghai, China, using the following channels located at standard positions of the 10–20 international system: FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT8, T7, C5,

C3,C1,CZ, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, P7, P5, P3, P1, PZ, P2, P4, P6, P8, PO7, PO5, PO3, POZ, PO4, PO6, PO8, CB1, O1, OZ, O2, and CB2. In the data collection stage, each subject was asked to seat in an armchair, keeping arms on the chair arms with two hands relaxing, and looks at a computer monitor approximately 1 m in front of the subject at eyes level.

Dataset 1 was collected from five subjects performing the motor imagery task in our online BCI system experiment [15], which has also been extensively used in existing BCI systems. The paradigm required the subject to control a cursor moving on the monitor by imagining the movements of his right or left hand for 2 s with a 4-s break between trials. For each subject, the data were collected over two sessions with a 15-min break in between. The first session was conducted without feedback, and 60 trials (30 trials for each class) obtained in this session were used for training and analysis. 140 trials (70 trials for each class) in the next session were taken as testing data to give online feedbacks. EEG signals were recorded, sampled at 500 Hz and bandpass filtered between 0.1 and 100 Hz.

Dataset 2 was collected from another two subjects performing two cognitive tasks, i.e., the figure perception and the mental arithmetic. In the figure perception task, the monitor displayed two classes of geometric figures (squares and triangles), and the subject must identify the class of the figure. In the mental arithmetic task, the monitor displayed an arithmetic formula containing three integers and the subject must perform the relevant mental calculation. The two tasks were carried out turn-by-turn, with the subject performing a task for 2 s, relaxing for 4 s, and then switching to the other task. Each subject took part in four data collection sessions with 60 trials (30 trials for each task) in each and 10-min break in between. Each session was conducted in the same fashion, and we took each session data as a training dataset in turn, and other sessions were used as testing data. EEG signals were recorded, sampled at 1000 Hz and bandpass filtered between 0.1 and 100 Hz.

Dataset 3 was collected from further two subjects performing a memory task in which the monitor briefly displayed either a plain white background or an English word on the white background for one second with an inter-trial interval of 4 s. When an English word was presented, the subject was required to recall its meaning and pronunciation. When the empty background was presented, the subject needed not to make any explicit responses. The recorded EEG data were segmented into two classes of epochs according to whether it corresponded to displaying the English word (the memory task) or the empty background. The sessions of data collection and the subdivision of testing and training data were the same as Dataset 2.

IV. Experiments and Results

In this section, we describe our experiments and results on the three aforementioned datasets, where dataset 1 contains data acquired from the motor imagery task and Datasets 2 and 3 refer to cognitive and memory tasks. To begin with, we applied the proposed scheme in Dataset 1, whose related spectral characteristic was known, and evaluated its performance in comparison with that of three popular algorithms. Then, the proposed scheme was applied to Datasets 2 and 3, where the discriminative spectral properties for them were not specifically identified.

A. Results on Dataset 1

Dataset 1 contains data about two classes of motor imagery tasks. The performance of the proposed tensor-based scheme in this dataset was compared with three popular algorithms, which are power spectrum density (PSD) [19], common spatial patterns (CSP) [16], and nonnegative multiway factorization (NMWF) [28]. We utilized each algorithm to extract discriminative features from different types of data to detect change of rhythmic activity in different mental states. For PSD, power spectral density values were computed as discriminative features from one-way EEG data (time) at given channels and frequency bands. CSP was applied to extract discriminative features from two-way EEG data (channel$\times$time), and NMWF was utilized to extract features from multiway EEG data (channel$\times$frequency $\times$ time) according to [20]. The corresponding discriminative patterns were illustrated, respectively, and comparisons of them among different algorithms were made (in order to give more clear comparisons, the spatial patterns were illustrated by focusing in the centro-parietal region and the spectral patterns were all showed in absolute weights).

According to [12] and [16], exemplary spectral characteristics of EEG in motor imagery tasks were involved in α rhythm (8–13 Hz) and β rhythm (14–30 Hz). In details, imagining left or right hand movement causes ERD within 8–30 Hz frequency band on the contralateral hemisphere, and ERS on the ipsilateral hemisphere. Those phenomena happened in the centro-parietal region, especially evident at channels C3 and C4.

To evaluated performances of four algorithms when the prior neurophysiologic knowledge was available, we preprocessed data by given frequency band filter (8–30 Hz, which contains all rhythms related to motor imagery according to the available spectral characteristic). Visual inspection showed that artifacts had been filtered out. The filtered signals were segmented into epochs (1–2000 ms).

For each trial, PSD features (8–30 Hz) at C3, C4, and CZ channels were calculated based on a temporal Fourier transform. Significant differences were observed in the averaged PSD for the two performed tasks. Imagining left hand movement led to the decrease of α and β rhythms' power at C4 channel and the increase at C3, whereas the contrary phenomena occur during imagining right hand movement. Fig. 2 illustrates the average power spectrum at channels C3 and C4 for the second subject, evident differences are presented between the two tasks, especially around 12–13 Hz.

The two most significant spatial patterns extracted by CSP for the second subject are illustrated in Fig. 3, showing highest discriminative weights at C3 and C4 in the centro-parietal region, respectively. The most significant pattern for the left movement imagination is focused at channel C3 [2] as illustrated in the left part of Fig. 3. Similarly, for the right-hand movement imagination, the focus is at C4 [2] as the other most significant pattern shown in the right part of Fig. 3.

For each epoch sample, the tensor was constructed in the given spatial-spectral-temporal range (62 channel; 8–30 Hz; 1–2000 ms, step by 20 ms). Fig. 4 shows the spatial, spectral

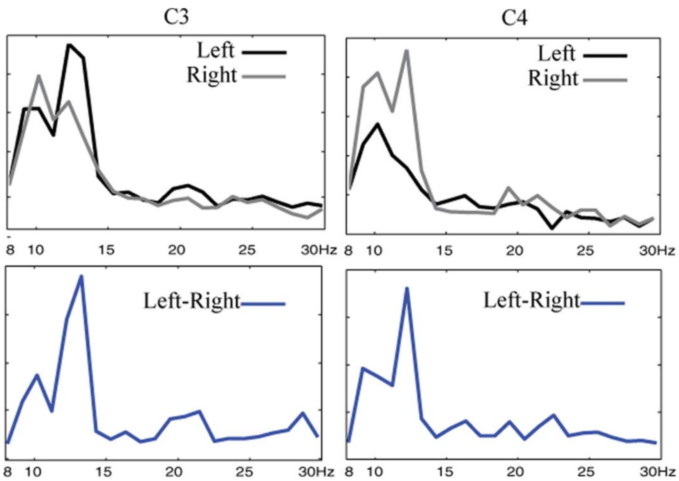


Fig. 2. PSD analysis for the second subject. There are significant differences in the average power spectrum at C3 and C4 for two performed tasks, especially around 12–13 Hz. The black line stands for imagining left hand movements, the gray line stands for imagining right hand movements, and the blue line stands for the absolute weight difference between two tasks.

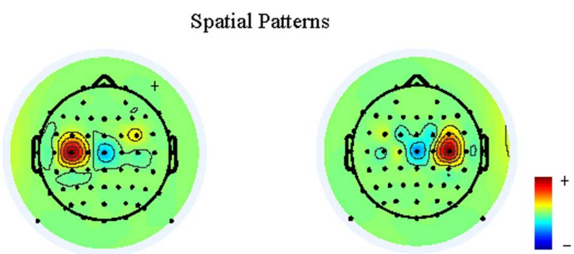


Fig. 3. The two most significant spatial patterns extracted by CSP method for the second subject.

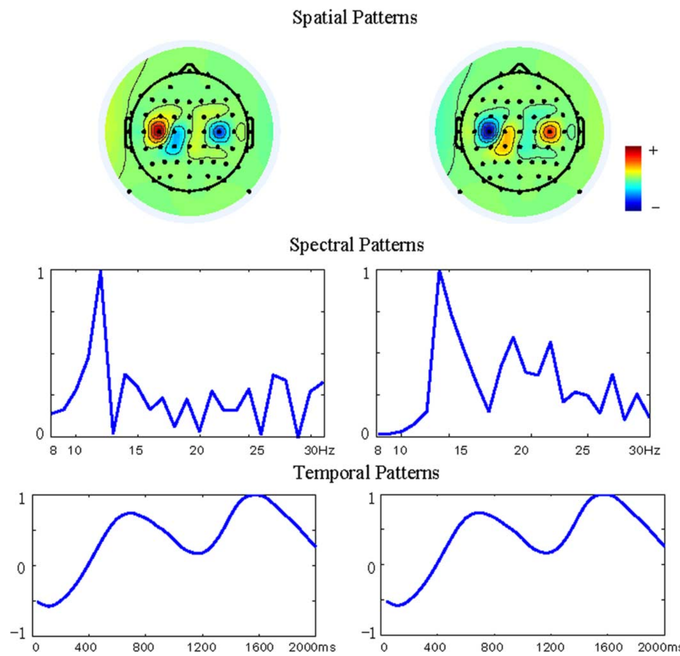


Fig. 4. The spatial, spectral, and temporal patterns of two most significant features in the proposed tensor-based scheme for the second subject.

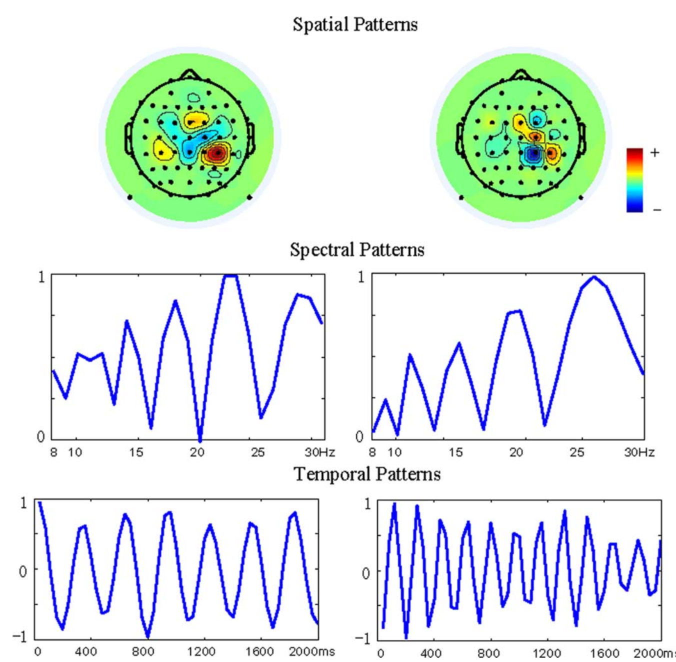


Fig. 5. The spatial, spectral, and temporal patterns of two removed redundant features in the proposed tensor-based scheme for the second subject.

and temporal patterns of two most significant features extracted by the proposed scheme for the second subject. Comparing the spectral patterns with the absolute difference between two tasks in PSD analysis, consistent characteristics were identified between them, and the most discriminative frequency bands were both concentrated around 12–13 Hz. In each spatial pattern, the highest absolute weights with opposite directions at C3 and C4 mean the C3 and C4 take leading roles for two tasks respectively. For comparison, the spatial, spectral and temporal patterns of two removed redundant features are also shown in Fig. 5, it is evident that the physiologically meaningful patterns are extracted by the proposed scheme.

In NMWF analysis, tensors were identical to the previous ones. The two most significant spatial, spectral and temporal patterns for the second subject are shown in Fig. 6. Although the patterns presented significant relations to two tasks respectively, they were less distinctive on classification than the patterns obtained by the proposed scheme and CSP. For example, the spatial patterns revealed the contralateral area for each task, but C3 and C4 could not be identified as the evidently highest weight channels in the centro-parietal region. This is because NMWF is optimal for data reconstruction but not for classification, while the proposed tensor scheme and CSP directly take discriminative features into account.

Based on the selected parameters in the training data, a validation procedure was conducted on the testing data for performance evaluation. For all algorithms, the number of features was reduced to 2–10 according to the training performance (more features cannot improve the training accuracy). Classification results are listed in the Table I. For the fourth and fifth subjects, all algorithms performed poorly. For the first, second, and third subjects, PSD obtained the accuracies from 62.1% to 75.0%. Since PSD calculated power spectral density features from one-way EEG data (time) at given channels and frequency bands which actually reflected the average spectral power distribution in a segment of time, and

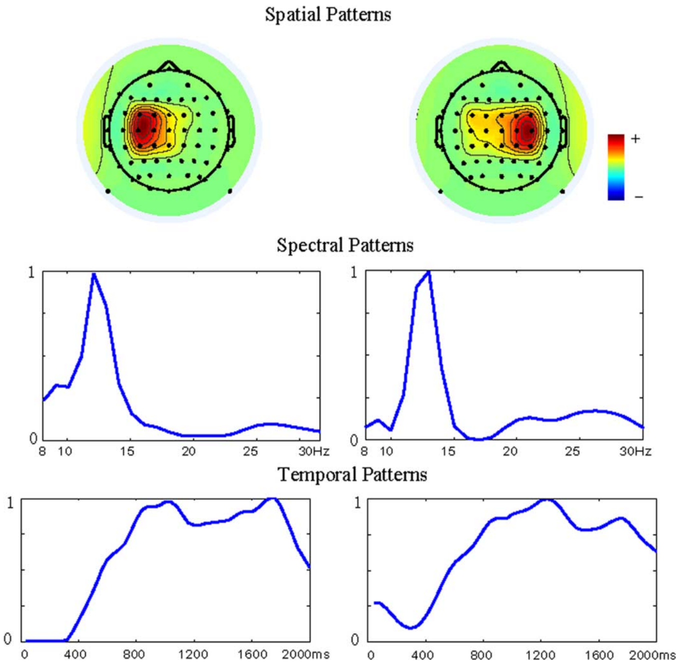


Fig. 6. The two most significant spatial, spectral, and temporal patterns extracted by NMWF for the second subject.

TABLE I
CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON THE FIRST DATASET (8–30 Hz)

8-30Hz	1st Sub.	2nd Sub.	3rd Sub.	4th Sub.	5th Sub.	Avg.
PSD[19]	72.1%	75.0%	62.1%	55.7%	40.0%	61.0%
CSP[16]	80.0%	88.6%	82.1%	61.4%	65.7%	75.6%
NMWF[28]	81.4%	84.3%	71.4%	56.4%	50.7%	68.8%
TbS	85.7%	95.0%	73.6%	60.7%	66.4%	76.3%

TbS is the proposed tensor-based scheme for signal trial EEG signal classification.

did not directly take the discriminative analysis on the spatial and spectral modes, it was not very effective for classification; CSP achieved remarkable performance by constructing discriminative spatial patterns from two-way EEG data; with multiway analysis, NMWF and the proposed scheme both obtained high classification accuracies. The accuracies of the proposed scheme were higher than NMWF, because GTDA, used in the scheme for feature extraction, extracted multilinear discriminative features for classification, whereas NMWF was rather a multiway decomposition technique than classification technique. For the five subjects, the proposed scheme achieved accuracies compared favorably to CSP: the average of accuracies was 76.3%, and CSP was 75.6%. Experimental results suggested that the proposed scheme performs comparably to CSP on properly preprocessed EEG data.

To investigate the robustness of all algorithms, the feature extraction without considering prior knowledge was also demonstrated. Raw EEG data were processed only by filtering through the general EEG wave spectral range (4–45 Hz) and the tensor objects were constructed in the general spatial-spectral-temporal range (62 channel; 4–45 Hz; 1–2000 ms, step by 20 ms). As listed in Table II, classification accuracies dropt for all algorithms. Especially, the average classification accuracy

TABLE II
CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON THE FIRST DATASET (4–45 Hz)

4-45Hz	1st Sub.	2nd Sub.	3rd Sub.	4th Sub.	5th Sub.	Avg.
PSD[19]	68.6%	77.1%	56.4%	50.7%	43.6%	59.3%
CSP[16]	46.4%	45.0%	54.3%	54.6%	52.9%	50.6%
NMWF[28]	78.5%	79.3%	65.7%	54.3%	47.1%	65.0%
TbS	84.3%	88.6%	72.1%	52.1%	52.9%	70.0%

TbS is the proposed tensor-based scheme for signal trial EEG signal classification.

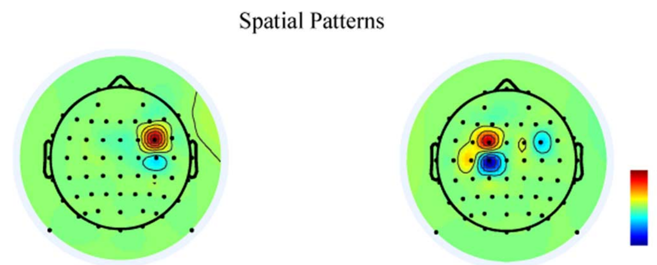


Fig. 7. The two most significant spatial patterns extracted by CSP for the second subject. The patterns have no significance in discrimination.

of CSP dramatically decreased to 50.6%. Since CSP detects change of rhythmic activities based on the temporal variances of signals, it is particularly noise prone and ineffective for the feature extraction in the general EEG wave range. Fig. 7 shows the two most important spatial patterns for the second subject extracted by CSP. The patterns had no significance in discrimination entirely. However, by using the multiway discriminative analysis, the proposed scheme can still maintain high accuracy. For the first and second subjects, the accuracies of the proposed scheme were still higher than 80%, and discriminative spatial and spectral patterns can still be obtained as shown in Fig. 8. In summary, the performance of CSP highly depends on filtering data through a specific preidentified frequency band, while the proposed scheme is more robust than CSP in identifying the discriminative patterns and features for noisy data.

B. Results on Dataset 2 and Dataset 3

Dataset 2 contains data from two subjects performing the figure perception and the mental arithmetic task, and Dataset 3 contains data from two subjects performing the memory tasks for identifying, remembering English words. Compared with motor imagery tasks (Dataset 1), the discriminative spectral properties for cognitive and memory tasks in Dataset 2 and Dataset 3 are not specifically identified.

The raw EEG data were preprocessed by the band filter in general EEG wave spectral range (4–45 Hz), and then segmented into epochs (1–2000 ms for Dataset 2, 1–1000 ms for Dataset 3). For each trial, the tensor was constructed in the general EEG spectral range (62 channel; 4–45 Hz, the general EEG wave frequency range; 1–2000 ms for Dataset 2, step by 50 ms, 1–1000 ms for Dataset 3, and step by 20 ms).

In order to reveal the most discriminative patterns and test the stability of the proposed scheme, each session data were taken as a training dataset in turn, and other sessions were used as a testing dataset, i.e., a 4 times cross-validation

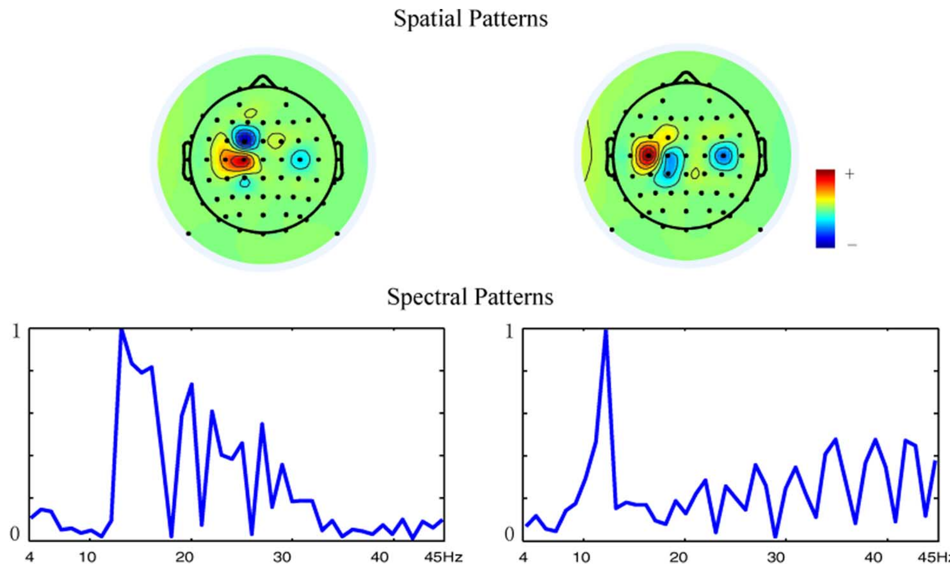


Fig. 8. The two most significant spatial and spectral patterns extracted by the proposed tensor-based scheme for the second subject. The patterns still have significance in discrimination.

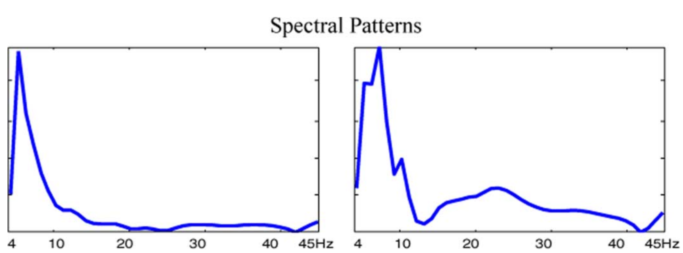


Fig. 9. The most significant spectral patterns extracted by the proposed tensor-based scheme from the session 1 data of Dataset 2. The left is for the sixth subject and the right is for the seventh subject.

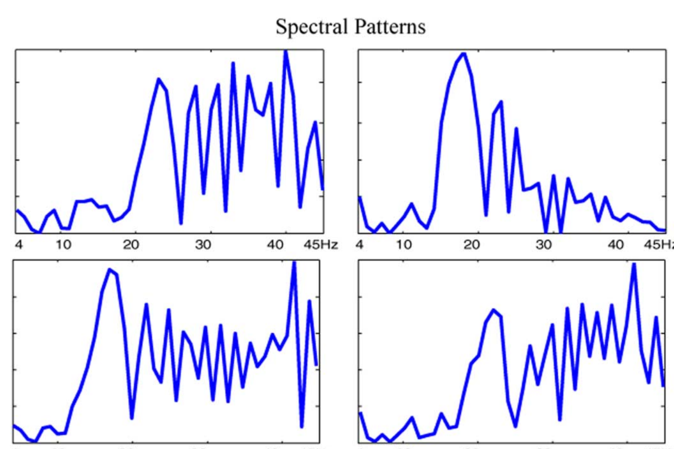


Fig. 10. The two most significant spectral patterns extracted by the proposed tensor-based scheme from the session 1 data of Dataset 3. The left is for the eighth subject and the right is for the ninth subject.

was conducted on Datasets 2 and 3, respectively. For each training dataset, four features with the largest Fisher scores were considered, and the corresponding patterns were selected as the significant discriminative patterns. For Dataset 2, each discriminative spectral pattern showed highest weights in low-frequency band (4–14 Hz). Fig. 9 gives an illustration of the most significant spectral patterns in absolute weights for each subject. For Dataset 3, the discriminative spectral patterns indicated close relation to high-frequency band (12–45 Hz). Fig. 10 gives an illustration of the two most significant spectral patterns in absolute weights for each subject. The classification accuracies for the two Datasets are listed in Tables III and IV, respectively. The proposed scheme acquired high accuracies for both datasets, while the performance of CSP was poor in the general EEG wave spectral range. For both of the datasets, by filtering the EEG data through the discriminative frequency band identified by the proposed scheme, the performances of CSP were substantially improved. For comparison, CSP is also applied in the EEG data filtered through insignificant frequency bands. The corresponding results showed that the discriminative spectral patterns extracted by the proposed scheme gave effective instructions for CSP.

TABLE III
CLASSIFICATION ACCURACIES OF TWO ALGORITHMS ON THE SECOND DATASET

	Filter band(Hz)	6^{th} Sub.	7^{th} Sub.
TbS	4-45	92.5%	86.2%
CSP[16]	4-45	74.9%	61.4%
	4-14	97.1%	75.8%
	15-45	72.1%	56.5%

TbS is the proposed tensor-based scheme for signal trial EEG signal classification.

At the end of this Section, it can be concluded that CSP and the proposed tensor based scheme can both achieve good performance in classification of properly preprocessed EEG data. However, the proposed scheme is more robust than CSP to extract discriminative features for noisy data, and it is most useful when the spectral characteristics of tasks are unavailable.

TABLE IV
CLASSIFICATION ACCURACIES OF TWO ALGORITHMS ON THE THIRD DATASET

	Filter band(Hz)	8th Sub.	9th Sub.
TbS	4-45	75.3%	70.5%
CSP[16]	4-45	56.9%	51.9%
	4-11	65.6%	52.4%
	12-45	73.5%	80.9%

TbS is the proposed tensor-based scheme for signal trial EEG signal classification.

V. CONCLUSION

In this paper, a tensor-based scheme is proposed for the single trial EEG classification, which incorporates merits from the spatial-spectral-temporal domain based EEG signal representation, GTDA based multilinear discriminative subspace reservation, Fisher score based redundant indiscriminative pattern elimination, and SVM based classification.

Unlike conventional algorithms, such as PSD and CSP, the proposed scheme includes the specific information endorsed by multiple-modalities, and improves identifying EEG activity for classification in the spatial-spectral-temporal domain. Furthermore, benefiting from GTDA based multilinear discriminative subspace reservation, and Fisher score based redundant indiscriminative pattern elimination, it is efficient for extracting multiway discriminative subspace from high dimensionality and high order EEG representation. Consequently, the proposed scheme is more powerful for EEG classification than conventional algorithms as well as NMWF which is practically a multiway decomposition technique.

Evaluations on three datasets confirmed the efficiency of the proposed scheme. In the motor imagery task, the proposed scheme performed comparably to CSP. It is worthy emphasizing that the proposed scheme is more robust than CSP by taking multi-way discriminative patterns into account. The robustness of proposed scheme is very useful because it can be applied to extract the discriminative features and patterns when the spectral characteristics in some paradigms are not available. In cognitive and memory tasks, although the discriminative spectral properties are not specifically identified, the scheme extracts the discriminative patterns and features in the general EEG wave range and achieves high classification accuracy. Besides, the discriminative spectral properties identified by the proposed scheme could be taken as effective instructions for CSP.

Our studies show that the proposed tensor-based scheme is an effective and robust data exploratory tool in classifying EEG signals, especially in the case of lacking prior neurophysiologic knowledge.

ACKNOWLEDGMENT

The authors would like to thank the technical writer, M. Kyle, for assistance with the presentation of this paper. The authors would also like to thank Associate Editor Prof. A. Bezerianos and anonymous reviewers for their constructive comments on this paper.

REFERENCES

[1] B. Blankertz, G. Curio, and K. R. Müller, “Classifying single trial EEG: Towards brain computer interfacing,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, vol. 14, pp. 157–164.
[2] C. Guger, H. Ramoser, and G. Pfurtscheller, “Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI),” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 447–456, Dec. 2000.
[3] D. Cai, X. He, and J. Han, “SRDA: An efficient algorithm for large scale discriminant analysis,” *IEEE Trans. Knowledge Data Eng.*, vol. 20, no. 1, pp. 1–12, Jan. 2008.
[4] D. Tao, X. Li, X. Wu, and S. J. Maybank, “General tensor discriminant analysis and gabor features for gait recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[5] X. He and P. Niyogi, “Locality preserving projections,” *Adv. Neural Inf. Process. Syst.*, vol. 16, pp. 1–8, 2003.
[6] E. Osuna, R. Freund, and F. Girosi, “Training support vector machines: An application to face detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1997, pp. 130–136.
[7] E. Donchin, K. M. Spencer, and R. Wijesinghe, “The mental prosthesis: Assessing the speed of a P300-based brain-computer interface,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 8, no. 2, pp. 174–179, Jun. 2000.
[8] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for EEG-based brain-computer interfaces,” *J. Neural Eng.*, vol. 4, pp. R1–R13, 2007.
[9] G. Blanchard and B. Blankertz, “BCI competition 2003—Data set IIa: Spatial patterns of self-controlled brain rhythm modulations,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1062–1066, Jun. 2004.
[10] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, “Combining features for BCI,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, vol. 15, pp. 1115–1122.
[11] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, “Increase information transfer rates in BCI by CSP extension to multi-class,” in *Proc. Adv. Neural Information Process. Syst.*, 2004, vol. 16, pp. 773–740.
[12] G. Pfurtscheller and C. Neuper, “Motor imagery activates primary sensorimotor area in humans,” *Neurosci. Lett.*, vol. 239, pp. 65–68, 1997.
[13] H. Lee, Y. D. Kim, A. Cichocki, and S. Choi, “Nonnegative tensor factorization for continuous EEG classification,” *Int. J. Neural Syst.*, vol. 17, no. 4, pp. 1–13, 2007.
[14] H. Sheikh, D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, “Electroencephalographic (EEG)-based communication: EEG control versus system performance in humans,” *Neurosci. Lett.*, vol. 345, no. 2, pp. 89–92, 2003.
[15] H. Sun and L. Zhang, “Subject-adaptive real-time BCI system,” in *Int. Conf. Neural Inf. Process.*, Kitakyushu, Japan, Nov. 2007, pp. 1037–1046.
[16] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, “Optimal spatial filtering of single trial EEG during imagined hand movement,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Apr. 2000.
[17] D. Tao, X. Li, X. Wu, and S. J. Maybank, “Geometric mean for subspace selection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2006.
[18] D. Cai, X. He, J. Han, and H.-J. Zhang, “Orthogonal Laplacianfaces for face recognition,” *IEEE Tran. Image Process.*, vol. 15, no. 11, pp. 3608–3614, Nov. 2006.
[19] P. D. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.
[20] J. Li, L. Zhang, and Q. Zhao, “Pattern classification of visual evoked potentials based on parallel factor analysis,” in *Int. Conf. Cognitive Neurodynamics*, Shanghai, China, Nov. 2007, pp. 571–575.
[21] J. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, “Non-invasive brain-actuated control of a mobile robot by human EEG,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1026–1033, Jun. 2004.
[22] J. R. Wolpaw, D. J. McFarland, and T. M. Vaughan, “Brain-computer interface research at the Wadsworth Center,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 222–226, Jun. 2000.
[23] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” *Clin. Neurophysiol.*, vol. 113, pp. 767–791, 2002.
[24] J. Ye, R. Janardan, and Q. Li, “Two-dimensional linear discriminant analysis,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2005, vol. 17, pp. 1569–1576.
[25] L. R. Tucker, “Some mathematical notes on three-mode factor analysis,” *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[26] L. Song, E. Gordon, and E. Gysels, "Phase synchrony rate for the recognition of motor imagery in BCIs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, vol. 18, pp. 1265–1272.

[27] M. Grosse-Wentrup, K. Gramann, and M. Buss, "Adaptive spatial filters with predefined region of interest for EEG based brain-computer-interfaces," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, vol. 19, pp. 537–544.

[28] M. Mørup, L. K. Hansen, J. Parnas, and S. M. Arnfred, Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization. Tech. Univ. Denmark, Copenhagen, Denmark, Tech. Rep., 2006.

[29] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed," *Nature*, vol. 398, no. 6725, pp. 297–298, 1999.

[30] R. A. Harshman, "Foundation of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," *UCLA Working Papers Phon.*, vol. 16, pp. 1–84, 1970.

[31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.

[32] S. Lemm, B. Blankertz, G. Curio, and K. R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[33] D. Tao, M. Song, X. Li, J. Shen, J. Sun, X. Wu, C. Faloutsos, and S. J. Maybank, "Bayesian tensor approach for 3-D face modelling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 10, pp. 1397–1410, Oct. 2008.

[34] X. He, D. Cai, and P. Niyogi, "Tensor subspace analysis," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, vol. 18, pp. 499–506.

[35] X. Liao, D. Yao, D. Wu, and C. Li, "Combining spatial filters for the classification of single-trial EEG in a finger movement task," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 821–831, May 2007.

[36] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Supervised tensor learning," *Knowledge Inf. Syst.*, vol. 13, no. 1, pp. 1–42, 2007.

[37] J. Sun, D. Tao, S. Papadimitriou, P. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," *ACM Trans. Knowledge Discovery Data*, vol. 2, no. 3, pp. 1–37, 2008.

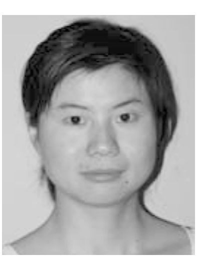
[38] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[39] X. Li, S. Lin, S. Yan, and D. Xu, "Discriminant locally linear embedding with high-order tensor data," *IEEE Trans. Syst., Man, Cybern.*, vol. 38, no. 2, pp. 342–352, Feb. 2008.

[40] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 725–738.

[41] W. Bian and D. Tao, "Harmonic mean for subspace selection," in *Proc. Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[42] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General averaged divegence analysis," in *Proc. IEEE Int. Conf. Data Mining*, 2007, pp. 302–311.

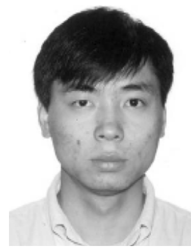


Jie Li received the B.S. and M.S. degrees in computer science and engineering from Anhui University of Technology, Maansan, China, in 2000 and 2003, respectively. She is currently working toward the Ph.D. degree in the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China.

Her research interests are in brain signal processing and brain–computer interface.

Liqing Zhang received the Ph.D. degree from Zhongshan University, Guangzhou, China, in 1988.

He is now a Professor with Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. His current research interests cover computational theory for cortical networks, brain signal processing and brain–computer interface, perception and cognition computing model, statistical learning and inference. He has published more than 120 papers in international journals and conferences.



Dacheng Tao received the Ph.D. degree from the University of London (Lon).

Currently, he holds a Nanyang titled post with the School of Computer Engineering in the Nanyang Technological University and a visiting post in Lon. He is a Visiting Professor in Xi Dian University and a Guest Professor in Wu Han University. His research is mainly on applying statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, neuroscience, and visual surveillance. He is an Associate Editor of *Neurocomputing* and the Official Journal of the International Association for Statistical Computing–*Computational Statistics & Data Analysis*.

Dr. Tao is an Associate Editor of IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING.

Han Sun received the B.S. degree in computer science and engineering, in 2007, from Shanghai Jiao Tong University, where he is now working toward the M.S. degree.

His research interests are in electrical brain activity analysis and brain–computer interface.

Qibin Zhao received the B.S. degree in computer science from Henan University of Science and Technology, Luoyang, China, in 2000, and the M.S. degree in electrical information engineering from Guangxi University, Nanning, China, in 2004. He is currently working toward the Ph.D. degree at Shanghai Jiao Tong University, Shanghai, China.

His current research interests include brain signal processing and brain–computer interface.