34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Time Sparsification of EEG Signals in Motor-Imagery Based Brain Computer Interfaces

Hiroshi Higashi and Toshihisa Tanaka

***Abstract*— We propose a method of sparsifying EEG signals in the time domain for common spatial patterns (CSP) which are often used for feature extraction in brain computer interfaces (BCI). For accurate classification, it is important to analyze the period of time when a BCI user performs a mental task. We address this problem by optimizing the CSP cost with a time sparsification that removes unnecessary samples from the classification. We design a cost function that has CSP spatial weights and time window as optimization parameters. To find these parameters, we use alternating optimization. In an experiment on classification of motor-imagery EEG signals, the proposed method increased classification accuracy by 6% averaged over five subjects.**

I. INTRODUCTION

Brain computer interfaces (BCI) capture brain activity associated with mental tasks and/or external stimuli and convert them into a device command [1]. BCIs provide a non-muscular communication and control channel to the external world [1]–[3]. Noninvasive measurement devices such as electroencephalogram (EEG), magnetoencephalogram (MEG), and functional magnetic resonance imaging (fMRI) are widely used to observe brain activity. Among them, EEG is considered to be a practical measurement method for use in engineering applications because of its simplicity and low cost.

Motor imagery based BCI (MI-BCI) is a promising realization of a BCI [2], [3]. One of its feature components is called the mu rhythm, which disappears around the motor cortex when the body moves. It is also known that the brain area where the mu rhythm disappears depends on the imaginary task: i.e., it changes in tasks going between hand and foot movement imagery, etc. [2], [3]. Therefore, by accurately extracting these changes from the measured EEG signals in the presence of measurement noise and spontaneous components related to other brain activities, we can classify the EEG signal associated with the imagination of different motor actions such as movement of the right hand, left hand, or feet.

The common spatial patterns (CSP) method is a well-known approach to extracting brain activity for an MI-BCI [1], [4], [5]. CSP gives a spatial weight to each electrode in a multichannel EEG measurement system. The weights are determined from learning data in such a way that the variances of the signal extracted by making linear combination of the multichannel signals and the spatial weights differ between two classes (e.g. left and right hand movement imageries). Recently, variants of CSP using the subband decomposition [6], [7], temporal/spectral filter combined methods [8]–[11], and regularization [12], [13] have been proposed to improve classification accuracy.

Generally, the learning data for CSP is measured by asking a user to follow a cue and performing the movement imagery task indicated by the cue. Therefore, CSP and its extensions use a time window that extracts signals observed after the cue. Since the user performs the task after the cue, the EEG features corresponding to the task are observed later than the cue. Therefore, the signals observed during the period when the user does not perform the task should be removed from the signal used for classification. This motivated us to add the time sparsification of the EEG signal to the CSP algorithm in order to remove unnecessary samples. In this paper, we propose a method for designing a sparse time window to be applied to the EEG signals in the learning and testing processes of CSP. We call this method time-sparse CSP (TSCSP). TSCSP sparsifies the signal in the time domain by applying a sparse time window to a segmented signal with a finite number of samples. The length and center of the window are found by minimizing a cost function including the spatial weights and the time window as optimization parameters. This cost function is based on CSP. We solve the optimization problem of the cost function by using alternating optimization, and we obtain the optimal spatial weights by using the selected time window.

TSCSP is demonstrated in classification experiments. The experimental results suggest that the use of a sparse time window designed by TSCSP can improve the classification performance of the MI-BCI.

It should be emphasized that the sparse time window can be incorporated with any of the previously proposed CSP variants [6]–[13] to increase the classification performance of BCIs.

II. COMMON SPATIAL PATTERNS

CSP is an effective supervised feature extraction method for a two-class MI-BCI. Let us review the basic CSP method [4], [5].

Let $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ be a signal observed in a multichannel measurement system, where M is the number of the channels and N is the number of samples. CSP finds a spatial weight vector, $\boldsymbol{w} \in \mathbb{R}^{M}$, in such a way that the variance of a signal extracted by the linear combination, $\boldsymbol{w}^T\boldsymbol{X}$, is minimized in

H. Higashi and T. Tanaka are with the Department of Electronic and Information Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan, higashi@sip.tuat.ac.jp, tanakat@cc.tuat.ac.jp

H. Higashi and T. Tanaka are also with the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, Japan.

a class [4], [5]. The time variance of the extracted signal of $\boldsymbol{X}$ is given by

$$\sigma^2(\boldsymbol{X}, \boldsymbol{w}) = \frac{1}{N}\sum_{n=1}^{N}\left|\boldsymbol{w}^T(\boldsymbol{x}_n - \boldsymbol{\mu})\right|^2, \quad (1)$$

where $\boldsymbol{x}_n$ is the nth row of $\boldsymbol{X}$, the time average of the observed signal is given by $\boldsymbol{\mu} = (1/N)\sum_{n=1}^{N}\boldsymbol{x}_n$, and $\cdot^T$ denotes the transpose of a vector or a matrix.

Assume that we have sets of learning data, $\mathcal{C}_1$ and $\mathcal{C}_2$. $\mathcal{C}_d$ contains the signals belonging to class d, $d \in \{1, 2\}$ is a class label, and $\mathcal{C}_1 \cap \mathcal{C}_2 = \emptyset$. CSP finds $\boldsymbol{w}$ in (1) that minimizes the intra-class variance in $\mathcal{C}_c$ under the normalization of samples, where c is a class label. More specifically, for c fixed, CSP finds $\boldsymbol{w}_c$ by solving the following optimization problem [4], [5];

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & E_{\boldsymbol{X}\in\mathcal{C}_c}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] \\ \text{subject to} \quad & \sum_{d=1,2} E_{\boldsymbol{X}\in\mathcal{C}_d}[\sigma^2(\boldsymbol{X}, \boldsymbol{w})] = 1, \end{aligned} \quad (2)$$

where $E_{\boldsymbol{X}\in\mathcal{C}_d}[\cdot]$ denotes the expectation over $\mathcal{C}_d$. The problem (2) can be rewritten as

$$\min_{\boldsymbol{w}} \ \boldsymbol{w}^T\boldsymbol{\Sigma}_c\boldsymbol{w}, \quad \text{subject to } \boldsymbol{w}^T(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w} = 1, \quad (3)$$

where $\boldsymbol{\Sigma}_d$, $d = 1, 2$, are defined as

$$\boldsymbol{\Sigma}_d = E_{\boldsymbol{X}\in\mathcal{C}_d}\left[\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\right]. \quad (4)$$

The solution of (3) is a generalized eigenvector corresponding to the smallest generalized eigenvalue of the generalized eigenvalue problem:

$$\boldsymbol{\Sigma}_c\boldsymbol{w} = \lambda(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\boldsymbol{w}. \quad (5)$$

Although the solution of (3) corresponds to the smallest eigenvalue in (5), using several eigenvectors as the spatial weights raise the classification accuracy [14]. $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_M$ are the M generalized eigenvectors obtained by solving (5), where $\hat{\boldsymbol{w}}_i$ corresponds to the ith smallest eigenvalue. We assume that the $2r$ eigenvectors are used as the spatial weights for classification of unlabeled data, $\boldsymbol{X}$. We obtain the feature vector, $\boldsymbol{y} \in \mathbb{R}^{2r}$, from $\boldsymbol{X}$ defined as

$$\begin{aligned} \boldsymbol{y} =& [\sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_1), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_r), \\ & \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_{M-r+1}), \ldots, \sigma^2(\boldsymbol{X}, \hat{\boldsymbol{w}}_M)]^T. \end{aligned} \quad (6)$$

III. TIME-SPARSE COMMON SPATIAL PATTERNS (TSCSP)

We propose a method for finding a sparse time window to remove unnecessary samples. Instead of (1), TSCSP uses

$$\sigma^2_{\mathrm{TW}}(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{b}) = \frac{1}{\|\boldsymbol{b}\|}\sum_{n=1}^{N} b_n|\boldsymbol{w}^T(\boldsymbol{x}_n - \boldsymbol{\mu})|^2, \quad (7)$$

as a feature vector, where, $\boldsymbol{b}$ is a binary vector working as a sparse time window defined as $\boldsymbol{b} = [b_1, \ldots, b_N]^T$, $b_i \in \{0, 1\}$, $i = 1, \ldots, N$, and $\|\cdot\|$ is the Euclidean norm of a vector. $\|\boldsymbol{b}\|$ represents the length of the time window because b_i takes 0 or 1. The main problem of TSCSP is to decide the spatial weights, $\boldsymbol{w}$, and the time window, $\boldsymbol{b}$.

In order to find these parameters, we design the following optimization problem

$$\min_{\boldsymbol{w},\boldsymbol{b}} \ f(\boldsymbol{b})J(\boldsymbol{w}, \boldsymbol{b}), \quad \text{subject to } \boldsymbol{b} \in \{\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_L\}, \quad (8)$$

where

$$J(\boldsymbol{w}, \boldsymbol{b}) = \frac{E_{\boldsymbol{X}\in\mathcal{C}_c}[\sigma^2_{\mathrm{TW}}(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{b})]}{\sum_{d=1,2} E_{\boldsymbol{X}\in\mathcal{C}_d}[\sigma^2_{\mathrm{TW}}(\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{b})]}, \quad (9)$$

$f(\boldsymbol{b})$ is a function of $\boldsymbol{b}$, and c is a class label chosen from 1 and 2. Regarding the sparse time window, $\boldsymbol{b}$, we choose $\boldsymbol{b}$ out of candidates representing $\{\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_L\}$, where L is the number of the candidates for $\boldsymbol{b}$ and $[\hat{\boldsymbol{b}}_i]_j \in \{0, 1\}$. $f(\boldsymbol{b})$ is defined in such a way that it is always positive and monotonically decreases with respect to $\|\boldsymbol{b}\|$.

We prevent overfitting of the spatial weights by combining the $f(\boldsymbol{b})$ in the cost function of (8). In general, if the signal length is short, it leads to overfitting and poor classification ability of the optimized parameters [15]. When the $\|\boldsymbol{b}\|$ is small (meaning that the time widow is short), $f(\boldsymbol{b})$ weights $J(\boldsymbol{w}, \boldsymbol{b})$ with a large coefficient. Then the short time window is not easily chosen by $f(\boldsymbol{b})$ in (8). Therefore, by adding $f(\boldsymbol{b})$ to the problem, we can design the parameters taking account of the signal length and prevent overfitting. We can use any function satisfying the terms as $f(\boldsymbol{b})$.

Because the number of the candidates for $\boldsymbol{b}$ is limited, we can search the optimal parameters of (8) by solving L generalized eigenvalue problems. However, it needs a large computation cost if L is large. Therefore, to optimize $\boldsymbol{w}$ and $\boldsymbol{b}$, we propose the following alternating optimization procedure based on alternating least squares. The procedure separates the optimization problem into two subproblems: one for finding $\boldsymbol{w}$ and the other for finding $\boldsymbol{b}$. The subproblems are alternately solved and the parameters are updated.

The first subproblem is to optimize $\boldsymbol{w}$ while fixing $\boldsymbol{b}$. Define

$$\boldsymbol{R}_d = E_{\boldsymbol{X}\in\mathcal{C}_d}\left[\frac{1}{\|\boldsymbol{b}\|}\sum_{n=1}^{N} b_n(\boldsymbol{x}_n - \boldsymbol{\mu})(\boldsymbol{x}_n - \boldsymbol{\mu})^T\right]. \quad (10)$$

for $d = 1, 2$. Then (8) can be written as

$$\min_{\boldsymbol{w}} \quad J(\boldsymbol{w}|\boldsymbol{b}) = \frac{\boldsymbol{w}^T\boldsymbol{R}_c(\boldsymbol{b})\boldsymbol{w}}{\boldsymbol{w}^T\left(\boldsymbol{R}_1(\boldsymbol{b}) + \boldsymbol{R}_2(\boldsymbol{b})\right)\boldsymbol{w}}. \quad (11)$$

The solution of (11) is given by the generalized eigenvector corresponding to the largest generalized eigenvalue of the generalized eigenvalue problem:

$$\boldsymbol{R}_c(\boldsymbol{b})\boldsymbol{w} = \lambda(\boldsymbol{R}_1(\boldsymbol{b}) + \boldsymbol{R}_2(\boldsymbol{b}))\boldsymbol{w}. \quad (12)$$

The second subproblem is to choose the sparse time window from the candidates. Define the N-dimensional vector,

$$\boldsymbol{z} = \left[|\boldsymbol{w}^T(\boldsymbol{x}_1 - \boldsymbol{\mu})|^2, \ldots, |\boldsymbol{w}^T(\boldsymbol{x}_N - \boldsymbol{\mu})|^2\right]^T. \quad (13)$$

TABLE I

CLASSIFICATION ACCURACY [%] GIVEN BY 5×5 CV. THE FIGURES IN THE ROUND BRACKETS BESIDE THE ACCURACIES REPRESENT THE STANDARD DEVIATION OF ACCURACY IN CV AND THE NUMBER OF DIMENSIONS OF THE FEATURE VECTOR.

Method	Subject					Ave.
	aa	*al*	*av*	*aw*	*ay*	
CSP	78.57 (±5.97, 2)	94.79 (±2.72, 8)	67.36 (±6.09, 7)	95.43 (1.86, 6)	88.93 (±4.46, 3)	85.21 (±4.22)
TSCSP1 ($q = 0.01$)	81.93 (±5.03, 4)	96.93 (±2.45, 8)	72.57 (±3.64, 2)	95.93 (2.39, 3)	92.86 (±3.96, 4)	88.04 (±3.49)
TSCSP1 ($q = 0.1$)	81.71 (±4.70, 5)	96.79 (±2.47, 7)	72.36 (±2.47, 2)	95.93 (2.28, 10)	92.71 (±3.34, 7)	87.90 (±3.24)
TSCSP1 ($q = 1$)	80.93 (±4.08, 6)	96.79 (±2.47, 7)	72.64 (±4.88, 3)	95.71 (4.88, 3)	93.07 (±3.87, 6)	87.83 (±3.47)
TSCSP1 ($q = 10$)	80.50 (±5.31, 8)	96.71 (±2.51, 7)	71.36 (±3.73, 2)	95.43 (2.58, 3)	93.21 (±4.31, 6)	87.44 (±3.69)
TSCSP1 ($q = 100$)	78.93 (±6.46, 3)	96.86 (±2.54, 7)	71.57 (±4.15, 2)	95.36 (2.42, 5)	93.36 (±3.94, 6)	87.21 (±3.90)
TSCSP2 ($q = 0.01$)	84.29 (±3.65, 5)	**98.86** (±1.62, 10)	78.71 (±4.64, 3)	96.43 (2.13, 5)	96.07 (±3.22, 4)	90.87 (±3.05)
TSCSP2 ($q = 0.1$)	85.43 (±3.60, 4)	98.71 (±1.59, 10)	**79.21** (±4.43, 5)	**96.64** (2.69, 2)	96.93 (±2.39, 8)	**91.39** (±2.94)
TSCSP2 ($q = 1$)	**85.93** (±4.23, 2)	98.71 (±1.59, 10)	78.36 (±4.13, 2)	96.57 (2.24, 6)	96.71 (±1.98, 1)	91.26 (±2.83)
TSCSP2 ($q = 10$)	84.64 (±4.58, 1)	98.71 (±1.67, 9)	77.07 (±5.64, 2)	96.29 (2.41, 2)	**97.00** (±1.84, 6)	90.74 (±3.23)
TSCSP2 ($q = 100$)	81.07 (±6.88, 1)	98.50 (±1.76, 9)	73.86 (±6.70, 5)	96.29 (2.18, 2)	96.93 (±2.03, 6)	89.33 (±3.91)

Its expectation over $\mathcal{C}_d$ is defined as $\boldsymbol{s}_d = E_{\boldsymbol{z}\in\mathcal{C}_d}[\boldsymbol{z}]$ for $d = 1, 2$. Accordingly, (8) can be written as

$$\min_{\boldsymbol{b}} \quad f(\boldsymbol{b})J(\boldsymbol{b}|\boldsymbol{w}) = f(\boldsymbol{b})\frac{\boldsymbol{b}^T\boldsymbol{s}_c}{\boldsymbol{b}^T(\boldsymbol{s}_1+\boldsymbol{s}_2)}, \qquad \text{subject to} \quad \boldsymbol{b}\in\{\hat{\boldsymbol{b}}_1,\ldots,\hat{\boldsymbol{b}}_L\}. \tag{14}$$

Since $f(\boldsymbol{b})J(\boldsymbol{b}|\boldsymbol{w})$ is calculated by performing $2N$ times multiplications after obtaining $\boldsymbol{s}_d$ and $f(\hat{\boldsymbol{b}}_i)$ can be calculated in advance, calculating $f(\boldsymbol{b})J(\boldsymbol{b}|\boldsymbol{w})$ does not cost much. Therefore, we calculate $f(\boldsymbol{b})J(\boldsymbol{b}|\boldsymbol{w})$ by using all of the candidates and select the time window that gives the minimum value of $f(\boldsymbol{b})J(\boldsymbol{b}|\boldsymbol{w})$ as the optimal window.

We alternately optimize $\boldsymbol{w}$ and $\boldsymbol{b}$ by solving optimization problems (11) and (14). We initialize $\boldsymbol{b}$ as a vector of which all elements are one, and we update $\boldsymbol{w}$ first in the alternating optimization.

The feature vector extracted with the sparse time window is defined as follows. By solving (12) with $\boldsymbol{b}$, we obtain M spatial patterns as $\hat{\boldsymbol{w}}_i$ for $i = 1,\ldots,M$ where $\hat{\boldsymbol{w}}_i$ is the eigenvector corresponding to the ith smallest eigenvalue of (12). As suggested in [8], we use the $2r$ eigenvectors to form a feature vector. Accordingly, the feature vector, $\boldsymbol{y}\in\mathbb{R}^{2r}$, is defined as

$$\begin{aligned}\boldsymbol{y} =&[\sigma^2_{\mathrm{TW}}(\boldsymbol{X},\hat{\boldsymbol{w}}_1,\boldsymbol{b}),\ldots,\sigma^2_{\mathrm{TW}}(\boldsymbol{X},\hat{\boldsymbol{w}}_r,\boldsymbol{b}),\\ &\sigma^2_{\mathrm{TW}}(\boldsymbol{X},\hat{\boldsymbol{w}}_{M-r+1},\boldsymbol{b}),\ldots,\sigma^2_{\mathrm{TW}}(\boldsymbol{X},\hat{\boldsymbol{w}}_M,\boldsymbol{b})]^T.\end{aligned} \tag{15}$$

The TSCSP procedure is summarized in Algorithm 1 as a pseudo-code.

Algorithm 1 Time-sparse CSP

Input: $\mathcal{C}_1$, $\mathcal{C}_2$: the sets of learning data of $\boldsymbol{X}\in\mathbb{R}^{M\times N}$.
Parameter: f: the function by the length of the time window, $\{\hat{\boldsymbol{b}}_1,\ldots,\hat{\boldsymbol{b}}_L\}$: the candidates for $\boldsymbol{b}$.
Output: $\hat{\boldsymbol{w}}_i$; spatial weights ($i = 1,\ldots,M$), $\boldsymbol{b}$: the time windows.

Initialize $\boldsymbol{b}$.
Set the index of iteration as $k = 0$.
repeat
  $k \leftarrow k+1$
  Update $\boldsymbol{w}$ by solving (11).
  Update $\boldsymbol{b}$ by solving (14).
  Calculate cost, C_k from the cost, $f(\boldsymbol{b})J(\boldsymbol{w},\boldsymbol{b})$.
until $C_k - C_{k-1}$ is sufficiently small.
Obtain M spatial weights, $\hat{\boldsymbol{w}}_1,\ldots,\hat{\boldsymbol{w}}_M$, by (12).

IV. EXPERIMENT

We compare the performances of TSCSP and CSP in classifying EEG signals during motor imagery tasks.

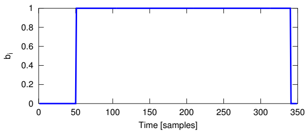
A. Data Description

We used dataset IVa from BCI competition III [16] (see `http://www.bbci.de/competition/iii/` for the details about the dataset). This dataset consists of EEG signals during right hand and right foot motor-imageries. The EEG signals were recorded from five subjects labeled *aa*, *al*, *av*, *aw*, and *ay*. 118 EEG channels were measured at positions of the extended international 10/20 system. The measured signal was bandpass filtered with a passband of 0.05–200 Hz and digitized at 1000 Hz. During each experiment, visual cues told the subject which imagery task (left hand, right hand, or right foot) should be performed. The cue was indicated for 3.5 seconds and the subject performed the motor imagery for this period. The resting interval between two trials was randomized to be from 1.75–2.25 seconds. Only EEG trials for the right hand and right foot were provided.

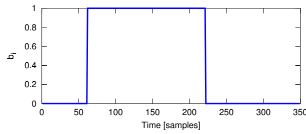
We furthermore applied a Butterworth lowpass filter whose cutoff frequency was 50 Hz and the filter order was 4 to this data. After that, the data was downsampled to 100 Hz. The dataset for each subject consisted of signals of 140 trials per class. The signal in each trial was extracted from the period of 0.5 to 3.5 seconds after a visual cue. Therefore, the length of the signal in each trial amounted to 3.0 seconds.

B. Result

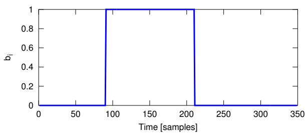
Table I shows the classification accuracy given by CSP and TSCSP with different parameters. The classification accuracies in Table I are from 5×5 cross validation (CV). Before classification, a bandpass filter between 7–30 Hz was applied to the signals as preprocessing. In TSCSP, we



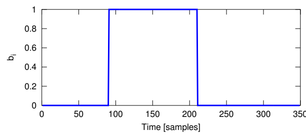
(a) $q = 0.1$



(b) $q = 1$



(c) $q = 10$



(d) $q = 100$

Fig. 1. Selected sparse time window for each q in subject *aa*.

prepared the candidates for $\boldsymbol{b}$ in such a way that

$$\hat{\boldsymbol{b}}_l = [\underbrace{0,\ldots,0}_{D_l}\underbrace{1,\ldots,1}_{W_l}0,\ldots,0]^T, \quad l = 1,\ldots,L, \tag{16}$$

where D_l and W_l represent the delay and length of the time window, respectively, and they were chosen out of $\{0, 10, \ldots, 350\}$ satisfying $D_l + W_l \leq 350$ and $W_l > 1$. Thus, the number of candidates, L, was 627. Next, we used the function,

$$f(\boldsymbol{b}) = \frac{1}{\ln(q\|\boldsymbol{b}\| + 1)}, \tag{17}$$

as $f(\boldsymbol{b})$ in (8), where q is a parameter that affects the length of the time window and q should be a positive value. In Table I, TSCSP1 is a result of not using the optimal time window for test signals. TSCSP2 is a result of using the optimal time window for test signals. The number of spatial weights representing r was tuned for each method and subject is also shown in Table I. The extracted feature vector was classified by using a support vector machine (SVM) [17] with a radial basis function (RBF) kernel. The SVM was implemented with SVM-Light [18]. The best classification performance was achieved by TSCSP2 ($q = 0.1$).

Examples of the sparse time windows selected by TSCSP are shown in Fig. 1. We can see that the selected time window becomes shorter as q becomes larger.

V. CONCLUSIONS

We described a method that selects a sparse time window for observed signals to remove unnecessary time periods from observations. On the basis of CSP, we designed the cost function with the sparse time-windowed signals. To find the spatial weights and the sparse time window satisfying the criterion, we used alternating optimization. In the experiment of classification for motor-imagery EEG signals, we showed the effectiveness for classification of the TSCSP. In future work, because TSCSP needs design and tuning for $f(\boldsymbol{b})$, we will develop a method for easily finding a suitable $f(\boldsymbol{b})$. The idea of time sparsification can be applied to the method based on CSP [6]–[11], and TSCSP can help to improve BCI classification performance and brain signal analysis.

ACKNOWLEDGMENT

This work is supported in part by KAKENHI, Grant-in-Aid for Scientific Research (B), 21360179 and Grant-in-Aid for JSPS Fellows, 24686.

REFERENCES

[1] G. Dornhege, J. d. R. Millan, T. Hinterberger, D. McFarland, and K.-R. Muller, Eds., *Toward brain-computer interfacing*. The MIT Press, 2007.

[2] S. Sanei and J. Chambers, *EEG signal processing*. Wiley-Interscience, 2007.

[3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[4] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.

[5] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

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

[7] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proceedings of The 2008 International Joint Conference on Neural Networks (IJCNN)*, 2008, pp. 2390–2397.

[8] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 9, pp. 1541–1548, 2005.

[9] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Müller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[10] R. Tomioka, G. Dornhege, G. Nolte, B. Blankertz, K. Aihara, and K. R. Müller, "Spectrally weighted common spatial pattern algorithm for single trial eeg classification," *Dept. Math. Eng., Univ. Tokyo, Tokyo, Japan, Tech. Rep*, vol. 40, 2006.

[11] H. Higashi and T. Tanaka, "Classification by weighting for spatio-frequency components of EEG signal during motor imagery," in *Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 585–588.

[12] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2936–2946, 2010.

[13] F. Lotte and C. Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[14] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[15] C. M. Bishop, *Pattern recognition and machine learning*, ser. Information science and statistics. Springer, 2006.

[16] B. Blankertz, K.-R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[17] V. N. Vapnik, *Statistical learning theory*, ser. Wiley series on adaptive and learning systems for signal processing, communications and control, S. Haykin, Ed. Wiley-Interscience, 1998.

[18] T. Joachims, "Making large-scale SVM learning practical," *Advances in Kernel Methods Support Vector Learning*, pp. 169–184, 1999.