Combining Spatial & Spectral Filters in ERD-based Multiclass BCI

E. B. Sadeghian, M. H. Moradi

***Abstract*— Electroencephalography-based brain computer interface is the most appropriate way to translate human thoughts into commands. Motor imagery activities appear as changes in μ and/or β rhythms which varies extremely from one subject to another. ERD/ERS patterns is the most common feature that represent these rhythmic information which are hidden in time, frequency, and space in the sense of brain's topographic modulations. In this paper we present most recent and powerful techniques of single trial motor imagery classification of optimization the spatial and spectral filters simultaneously, and apply their multiclass extension to a 4- class motor imagery data from BCI Competition III. Our results show a significant improvement in comparison with winner results of that competition. These are: Common Spatial Patterns (CSP) and its two extensions to the Common Spatio-Spectral Patterns (CSSP), Common Sparse Spectral Spatial Patterns (CSSSP), and also the frequency tuned version of CSP, i.e. the Sub Band CSP (SBCSP). These methods extract our ERD related features, which are then fed to 6 support vector machine classifiers to classify between 4 different movement imageries.**

Keywords- Multi Class BCI, Spatio Spectral filters, single trial classification, common spatial patterns

I. Introduction

PEOPLE suffering from Amyotrophic Lateral Sclerosis (ALS) lose their muscle movement degeneratively and at a later stage may become completely paralyzed. Motor imagery based brain-computer interface (BCI) addresses this problem by making it possible to translate human intentions directly to the outside world. Although ECoG signal provides better temporal and spatial resolution than EEG, electroencephalographic based BCI is chosen because of its simplicity, inexpensiveness and applicability.

Imagination of limb movements are captured in two major changes in spontaneous EEG: ERP components, also referred as slow cortical potential (SCP) e.g. P300 component, error potential, movement related potential (MRP); also called readiness potential (RD) and oscillatory features i.e. Event Related Desynchronization and Synchronization (ERD/ERS). ERD is caused by suppression of idle EEG signal in the specific region of motor and somatosensory cortex due to loss of synchrony in μ and β bands, roughly defined in the 12-16Hz and 18-24Hz respectively, and appears both for real and imagined movements. The event in the opposite direction is called Event Related synchronization, which follows the former. In order to design an ERD based BCI system, one could use of any feature that extracts high frequency components (mainly energy) of a signal, but the challenging issue would be to zoom on most informative frequency band, spatial region of brain, and best time point of signal simultaneously, for system training. It is due to the interdependency of these three parameters, that one needs a method to optimize them simultaneously.

The first way out in multichannel recordings is the Common Spatial Pattern (CSP) [1] i.e. the method of spatially tuning and projection which will be explained in detail later in this paper. However, the spatial filter that CSP provides must only be applied to the informative frequency band/sub band. More specifically the oscillatory information of some data are spread over wide range of μ and/or β band, but frequency band of interest in most data are seen to be focused on a non specific brain rhythm e.g. 9-18Hz and is neurophysiologically specific to each subject [2]. In general, applying CSP or any other feature extraction method to noisy unfiltered EEG, yields to bad generalization and distorts classifier's performance. Therefore, here we present two extensions of CSP, i.e. methods of CSSP and CSSSP which address the same issue and find spatial and spectral patterns simultaneously.

This paper is organized as follow: Section II has the detailed explanations and formulations of three methods. Section III, introduces the dataset used and training phase of parameter tunings. Finally there is Comparison between all methods in result and conclusion sections.

II. Methods

A. Common Spatial Patterns (CSP)

Method of CSP was first introduced by Muller-Gerking *et al.* [1] and is one of the most successful approaches of motor imagery recognition in multichannel EEG.

This method is originally designed for differentiating between only two classes, but can be easily extended to multiclass problems [5], [6]. Let us take the example of

Manuscript received September 8, 2007.
E. B. Sadeghian is a M.Sc student with the Biomedical Engineering Department, Amir Kabir University of Technology (Tehran Polytechnic), Tehran, Iran (e-mail: e_bsadeghian@ aut.ac.ir).
M. H. Moradi is with the Biomedical Engineering Department, Amir Kabir University of Technology, Tehran, Iran (e-mail: mhmoradi@ aut.ac.ir).

discriminating left hand versus right hand imaginary movement. In this case the neurophysiological background of CSP would be if we focus on left hand area of motor cortex, there can be seen an attenuated motor rhythm during imagination of left hand and a strong motor rhythm (idle rhythm) during imagination of right hand movement. This discrepancy between left and right hand imaginary movement is exactly what CSP tries to magnify, i.e. to find directions (spatial filters) in such a way that maximize variance for one class and at the same time minimize variance of the other class. Having two distributions of high dimensional spatio-temporal EEG, CSP finds very few spatial filters in such a way that variances of resulting projected trials are the most discriminative features for classification.

In doing so, sum of variances of all projected trials is set to 1, therefore maximization of sum of variances of all projected trials of class 1 yields to minimization of sum of variances of all projected trials of class 2.

Specifically, we begin with two groups of N×T recorded trials, i.e. X_c^i the raw data of trial i belonging to class c $\{1,2\}$, with N the number of channels and T the number of samples in time. Optimization problem is formulated as:

$$\max_P \; P^T \sum_1 P, \; s.t. \; P^T\left(\sum_1 + \sum_2\right)P = 1 \qquad (1)$$

Where P is our projection matrix to be found, $\sum_1$ and $\sum_2$ are the covariance matrices of class 1 and 2, respectively:

$$\sum_1 = \left\langle X_c^{i^T} * X_c^i \middle/ trace\left(X_c^{i^T} * X_c^i\right) \right\rangle , \; c \in \{1,2\} \qquad (2)$$

Where $\langle \; \rangle$ stands for averaging over all trials of class c.

A solution of (1) is as follows (For further explanations see [1].): First we decompose $\sum_1 + \sum_2$ into its eigenvectors to find the whitening transformation of interest:

$$\sum_1 + \sum_2 = B\lambda B^T \qquad (3)$$

Where B is an N×N matrix of eigenvectors and τ is the corresponding matrix of eigenvalues. Then with whitening transformation i.e. $W = \lambda^{-1/2} B^T$, we have:

$$W\left(\sum_1 + \sum_2\right)W^T = I \qquad (4)$$

Now, if we decompose each individual transformed covariance matrices by their eigenvectors, it reads:

$$W \sum_1 W^T = RDR^T , \; W \sum_2 W^T = R(I - D)R^T \qquad (5)$$

Finally, comparing (1) with (5) gives a solution of P that is:

$$P = R^T W \qquad (6)$$

Using our projection matrix P, we can retain only 2m most important spatial filters (e.g. m=1 means only first and last rows of matrix P) and successfully filter our trials as:

$$Z_c^i = P X_c^i \qquad (7)$$

The feature vector to be extracted is composed of the variances of 2m filtered trials normalized by the total variance of the projections retained, and log-transformed:

$$f_p^i = \log\left(\frac{\mathrm{var}_p^i}{\sum_{p=1}^{2m} \mathrm{var}_p^i}\right) \quad p = 1,\ldots,m, N-m+1,\ldots,N \qquad (8)$$

Since the motor imagery idle rhythms are highly variable between subjects, CSP, despite its strength in finding spatial patterns, might fail when opposed to unfiltered EEG signal. Therefore first thing that comes to mind is to find the best informative frequency sub-band from the wide range of 8-30Hz, which gives the best classification accuracy.

Although this solution is simple, accurate and straightforward, it adds a very time consuming, exhaustive manual tuning to training phase of any classifier. We have called this Sub-Band CSP (SBCSP), and executed it for results comparisons.

B. *Common Spatio-Spectral Pattern (CSSP)*

It is known in literature, that generally augmenting the dimension of any extracted feature or raw signal with its own delayed version is good for classification purposes and enhances the robustness of method. Here in CSSP the common CSP is applied to the concatenation of X_c^i and $X_c^{i^\tau}$ (i.e. the delayed trial by τ samples in time) in channel dimension. This time the same exact optimization in CSP is applied on $\hat{X}_c^i$ with 2N×T dimension.

This method is proposed by Steven Lemm *et al.* in [2], where they show that *addition of one delay tab to our trials as supplementary channels, yields to a specifically tuned FIR filter at each individual channel.* Therewith incorporates spectral filters into spatial ones, i.e. Spatio Spectral Patterns. Of course this method depends on parameter τ to be selected by some validation approaches on training data.

C. Common Sparse Spectral Spatial Pattern (CSSSP)

This method is recently proposed by Guido Dornhege *et al.* [3], and in contrast to CSSP allows for a simultaneous optimization of spatial and spectral filters together. In particular, here a trial is filtered with an unknown FIR filter prior to the calculation of spatial patterns, then these filter coefficients are optimized in such a way that the total optimization cost function in (1) become maximized. Since solution of our FIR filter is found through optimization techniques such as gradient descend or line search methods, therefore it depends greatly on initial points.

Just like CSSP, this method adds its own parameters to be determined through some validation procedures on training data. These parameters are filter length T, which is not as determinative as the second parameter, the sparsity constant C. The latter parameter compensates for the increased complexity of selecting a large T, with only a few nonzero coefficients of our FIR solution.

D. Multiclass Extensions

As it is previously mentioned, CSP and all its extensions are originally designed for discrimination between two distributions only. Therefore, one way to apply them on multiclass paradigm is to solve optimization problems and find patterns between each possible pair of classes separately. Accordingly, for 4-class data one would have 6 CSPs, each maximizing the difference between 2-class variances. Features extracted this way should then be fed into 6 binary classifiers to decide between each of possible pairs. Therefore a new projected trial with all 6 CSPs will be assigned to the class for which most classifiers are voting.

The second way is to compute CSPs for each class versus others, i.e. "one versus rest" approach, which in this case yields to 4 projection matrices. Followed by this, one could use either a multiclass classifier or 6 binary ones.

More preferable approach is selected to be the latter by [4] - [6].

III. Application

A. Dataset

We have performed our analysis on dataset IIIa from BCI Competition III [7],[8] provided by the Laboratory of Brain-Computer Interfaces (BCI-Lab), Graz University of Technology (Prof. Gert Pfurtscheller, Alois Schlögl). This data includes cued motor imageries of 4 classes (left hand, right hand, foot, tongue) provided with 3 subjects, 60 channels and 60 trials per class. It has been filtered between 1and 50 Hz and sampled with 250 Hz. The experiment consists of several runs (at least 6) with 40 trials each. After trial begins, the first 2s were quite, at t=2s an acoustic stimulus indicated the beginning of the trial, and a cross "+" is displayed; then from second 3 to 4.25 an arrow to the left, right, up or down was displayed; at the same time the subject was asked to imagine a left hand, right hand, tongue or foot movement, respectively, until the cross disappeared at t=7s. Each of the 4 cues was displayed 10 times within each run in a randomized order. Fig. 1 depicts this paradigm.

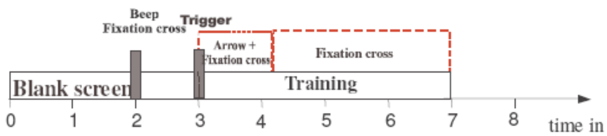

Fig. 1. Timing of the paradigm.

B. Preprocessing

Prior to each processing method, all 60 channels were band pass filtered from 8Hz to 30Hz, applying a causal IIR filter. This broad frequency band focuses on μ and β oscillations that contain the most key information used in BCI research It has shown to surpass other narrower frequency bands and improves the classification accuracy [4],[5].

C. Training

Here, in all feature extraction methods the training is done on all available runs of data except the last run which is considered as test data for evaluation purpose.

The validation procedure used for parameter tunings was 10×10 fold cross validation, in which each time we shuffled the data, split it further into a 9by1 ratio of train to test sets.

One rule to be observed specially in multiclass data is to select test and train sets in such a way that each set contains equal samples of each class; therewith the classes are balanced throughout and validation results would be more generalizable and realistic.

In feature extraction phase we have followed the approach of "one versus rest" CSPs and retained two most discriminative spatial patterns of each projection matrix, wherein $m=2$. Doing so CSPs were extracted on 60× (1.5s* sampling rate) trial matrices and our feature vector dimension for each incoming trial was 1×16. We have applied 6 binary linear SVM classifiers to extracted feature vectors, wherein a trial is assigned to class c only if all 3 engaged classifiers of class c would vote for that.

All 6 classifiers were trained by features of one specified time point, which was the time point with minimum training error and was determined through 10×10 fold cross validation on training set. More precisely we have swept the 3-8s time interval of training trials by 0.1s steps and run a cross validation on each.

D. Kappa, the comparable performance measure

Using averaged classification accuracy as performance measure makes it difficult to have a fair comparison between multiclass problems. Wherein a classification accuracy of 75% in a two-class problem is equivalent to an accuracy of 62.5% in a four-class problem. While the kappa value in both cases equals 0.5 indicating the midpoint between random and perfect classification. In general kappa varies between -1 and 1 with zero value representing the random performance. Wherever the number of items to be classified in each class is equal as here is the case the kappa's relationship with classification accuracy, despite the complexity of its derivation reduces to a line.

For a balanced four-class problem, kappa is captured from averaged accuracy by:

$$Kappa = \frac{\overline{C}r - 0.25}{0.75} \qquad (9)$$

IV. RESULTS

Tables I to III have our results of applying all described methods on final run with training data taken from all previous runs of 3 subjects' datasets. Here columns 2and 3 depict the captured accuracy of method CSP applied with 8 and 16 dimensional feature vectors wherein we reserved 1 and 2 most important spatial patterns respectively. Results of SBCSP are captured by applying CSP on the best frequency sub-band which is chosen through an exhaustive manual search on all possible sub-bands in 8-30Hz range. These best selected sub-bands of all subjects are also depicted in table IV.

All results provided here correspond to maximum captured accuracy of sample by sample classification of test data, which was also the case in BCI competition III-dataset IIIa. Time courses of classification accuracies generated applying method of CSSSP on 3 subjects' test data are shown in Fig. 2 to Fig.4.

TABLE I
CLASSIFICATION RESULTS OF SUBJECT K3

	CSP 8	CSP 16	SBCSP	CSSP	CSSSP
Classification Accuracy	97.5%	97.5%	100%	100%	100%
Kappa	96.6%	96.6%	100%	100%	100%

TABLE II
CLASSIFICATION RESULTS OF SUBJECT K6

	CSP 8	CSP 16	SBCSP	CSSP	CSSSP
Classification Accuracy	70%	72.5%	77.5%	82.5%	75%
Kappa	60%	63.3%	70%	76.6%	66.6%

TABLE III
CLASSIFICATION RESULTS OF SUBJECT L1

	CSP 8	CSP 16	SBCSP	CSSP	CSSSP
Classification Accuracy	85%	87.5%	95%	90%	90%
Kappa	80%	83.3%	93.3%	86.6%	86.6%

TABLE IV
BEST SELECTED FREQUENCY SUB-BAND OF EACH SUBJECT

	K3	K6	L1
Best Frequency Range	9-16Hz	18-24Hz	9-15Hz

In confusion matrices shown in tables V to VII, actual imageries appear above while classified ones are captured on left side. These tables compare the results of applying CSP and CSSP in classification of each type of imageries separately, where upper and lower digits indicate the number of classified imageries using CSP and CSSP, respectively.

Since test data contains 10 samples of each imagery, to be classified, the digits in confusion matrices represent the number of classified imageries each with respect to 10.

TABLE V
CONFUSION MATRIX OF SUBJECT K3
THE ABBREVIATIONS LH, RH, F, AND T STAND FOR LEFT HAND, RIGHT HAND, FOOT, AND TONGUE MOVEMENT IMAGERY RESPECTIVELY. UPPER DIGITS RESULT FROM METHOD OF CSP, AND LOWER ONES RESULT FROM METHOD OF CSSP

	LH	RH	F	T
LH	10 10	0 0	0 0	0 0
RH	0 0	10 10	0 0	0 0
F	0 0	0 0	9 10	0 0
T	0 0	0 0	1 0	10 10

TABLE VI
CONFUSION MATRIX OF SUBJECT K6

	LH	RH	F	T
LH	6 9	0 0	0 0	1 1
RH	2 1	8 9	3 2	1 2
F	2 0	1 0	7 8	0 0
T	0 0	1 1	0 0	8 7

TABLE VII
CONFUSION MATRIX OF SUBJECT L1

	LH	RH	F	T
LH	9 10	0 0	0 0	0 0
RH	0 0	10 10	0 0	0 0
F	1 0	0 0	10 10	4 4
T	0 0	0 0	0 0	6 6

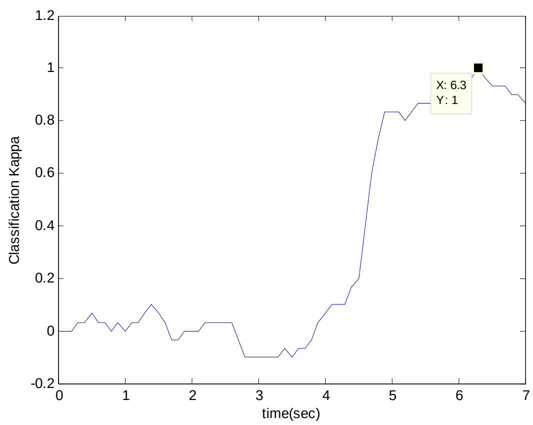

Fig. 2. Time course for classification kappa generated by applying CSSSP on subject k3's test set, plotted from 0s to 7s with respect to trigger point.

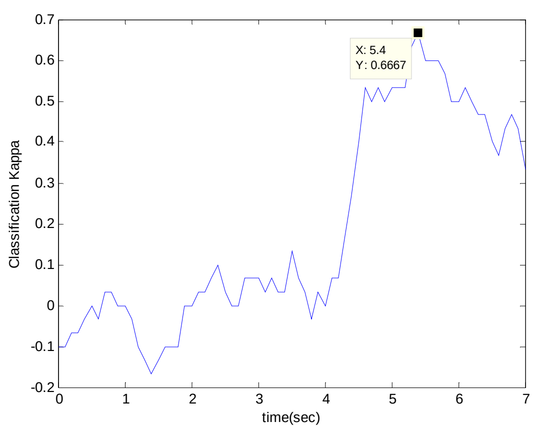

Fig. 3. Time course for classification kappa generated by applying CSSSP on subject k6's test set, plotted from 0s to 7s with respect to trigger point.

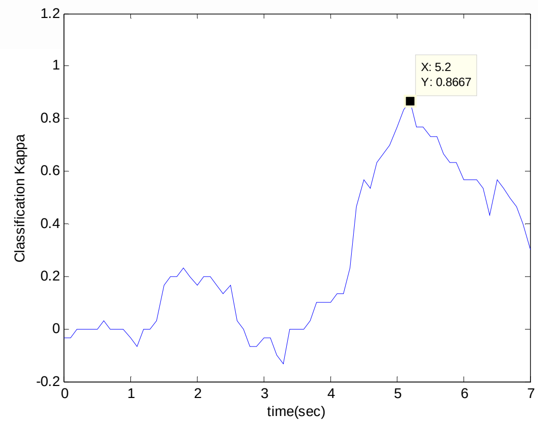

Fig. 4. Time course for classification kappa generated by applying CSSSP on subject l1's test set, plotted from 0s to 7s with respect to trigger point.

TABLE VIII
BEST RESULTS OF INDIVIDUAL SUBJECTS AND AVERAGED RESULT ACROSS ALL 3

	K3	K6	L1	Averaged Kappa
Our Results	100%	76.6%	93.3%	89.96%
Winner Results	82.2%	75.56%	80%	79.26%

V. CONCLUSION

In this paper three extensions of combining the significant method of CSP with spectral optimization was introduced and applied on a 4-class motor imagery data from BCI Competition III.

These results show a significant improvement of classification accuracy over original CSP and lead to even better accuracy as with troublous manual tuning of SBCSP.

Accuracies depicted in first row of table VIII are best results acquired among all methods applied here.

In order to simulate the real situation of a BCI system following the chronological order of recording the train and test data, we considered the last available run of the data as test set. Although the arrangement was different in the competition, i.e. test data has been randomly selected among all available trials, we believe that our results are less optimistic and more realistic because of two reasons: First, the special arrangement of the data can affect a method's performance and thereby prevents the execution of a standard comparison. Furthermore in random selection of test set, since there is more similarity between train and test trials, ultimate results would be more optimistic than real situation. With this in mind, our accuracies captured by kappa value in table VIII show a noticeable improvement in comparison with winner results of that competition.

REFERENCES

[1] H. Ramoser, J. Muller-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, Dec. 2000.

[2] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, “Spatio-spectral filters for improved classification of single trial EEG,” *IEEE Trans.*

[3] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and Klaus-Robert Müller, "Combined Optimization of Spatial and Temporal Filters for Improving Brain-Computer Interfacing", *IEEE Trans. Biomedical Eng*, vol. 53, no. 11,pp 2274-2281, Nov. 2006

[4] G. Townsend ,B. Graimann, and G. Pfurtscheller “Continuous EEG Classification During Motor Imagery—Simulation of an Asynchronous BCI” *IEEE Trans. Neural, Rehab*.vol. 12, pp. 258-265, Apr. 2006.

[5] E. B. Sadeghian, M.H. Moradi “Continuous Detection of Motor Imagery in a 4-Class Asynchronous BCI,” *29th IEEE EMBS Annual International Conference, Lyon, France*, *Aug. 23-26, 2007,* to be published.

[6] G. Dornhege, B. Blankertz, G. Curio, and K. R. Muller, “Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[7] B. Blankertz, K.-R. M¨uller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schl¨ogl, G. Pfurtscheller, J. del R. Mill´an, M. Schr¨oder, and N. Birbaumer, “The BCI competition III: Validating alternative approaches to actual BCI problems,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, 2006.

[8] BCI Competition III, http://ida.first.fraunhofer.de/ projects/bci/competition iii, 2005.