

CSP-EK: CSP filter adaptation using extended Kalman filter for BCI applications

Ali Mobaien, Reza Boostani
Biomedical Engineering Group
Faculty of Electrical & Computer Engineering
Shiraz University
Shiraz, Iran
ali.mobaien@gmail.com, boostani@shirazu.ac.ir

Abstract—Brain Computer Interface (BCI) provides a communication channel via computer between the mind and environment. Extracting suitable and discriminant features is one of the most important stages in BCI Applications. Common spatial pattern (CSP) is a well-known feature extraction method; however, due to the non-stationary nature of EEG signals CSP should be updated through the time.

This paper proposes a novel adaptation method using extended Kalman filter (EKF) for CSP feature elicitation and classification for BCI systems. In this method, EKF updates CSP filters in both supervised and unsupervised schemes. The proposed method was applied to data of BCI competition-III containing two- and multi-task imagery movements. Results demonstrate a considerable improvement in terms of classification accuracy by the proposed method in comparison with standard CSP method.

Keywords-brain computer interface; common spatial patterns; extended Kalman filter; adaptation; supervised; unsupervised

I. INTRODUCTION

A brain computer interface (BCI) is a system which allows an individual to communicate with outside world by translating his brain signal changes [1]. These changes (patterns) can be emerged by performing specific real or imagery movements or different cognitive activities. Each of these patterns can be detected and translated to a specific action by a trained classifier program in the computer [1]. The main challenging problem in BCI is to extract and classify these patterns, especially while using non-invasive signal acquisition methods such as electroencephalogram (EEG). EEG signals are filtered by soft and hard tissues of the head while passing from the source to scalp. This causes to distort temporal and local information of the brain signals. Furthermore, in EEG recording process each electrode records a combination of signals from several sources. To overcome these drawbacks, some approaches assume that brain source signals distort only spatially while passing through different layers of brain and scalp [2]. Common spatial patterns (CSP) is one the most popular methods which uses this assumption. CSP is a method to extract the uncorrelated components from multi-channel data [3][4]. The output of this method is a set of spatial filters which can discriminate two classes of data by extracting components of maximum variance (energy) for one class and minimum variance for the others [2][5].

EEG signals have a non-stationary nature due to several factors such as changing in the firing patterns of neurons through the time, changes in condition of recording environment, unsteady electrode impedance during recording, fatigue and etc. These non-stationary conditions make adaptation essential for feature extraction and classification in EEG based BCI systems [1].

Kalman filter is known as one of the greatest data fusion algorithms. This is the optimal estimator for one dimensional linear system with Gaussian error statistics. Kalman filter can be used for smoothing noisy data and providing estimations of parameters of interest, such as states of a dynamic system. Small computational requirements and perfect recursive properties make this filter one of the most favorite estimators [6]. Kalman filter estimates system states by combining noisy measurements with predictions of interested parameters. These predictions are based on the system dynamic model and the last state of the parameters. In the original Kalman filter, dynamic and observation models should necessarily be linear, while in its extended version, one of these models or both of them can be nonlinear [7]. In this study, CSP filter elements were considered as system parameters and were adapted using an extended Kalman filter (EKF).

Lowne *et al.* [8] proposed an adaptive logistic classifier for BCI experiments which used features from a nonlinear mapping. The adaptation algorithm was implemented using an EKF. Tsui *et al.* [1][9] used this idea to adapt a linear discriminant analysis (LDA) classifier with logarithmic band power features for BCI applications. In this work, the same idea was used to update CSP filters which were trained for feature extraction and classification. In the classification step as a new EEG test trial arrives, an adaptation algorithm based on EKF runs. This adaptation algorithm is executed in both supervised and unsupervised manners, where the label of incoming data is used and not used in the training, respectively. The proposed method was investigated by applying it to dataset 'a' from BCI competition-III. Classification was done for two-task and multi-task scenarios.

In the next Section the employed dataset is introduced. In Section III the theoretical details behind the proposed method and in Section IV, results of different tests are presented. The outcomes of these results are discussed in Section V and are followed with some concluding comments in Section VI.

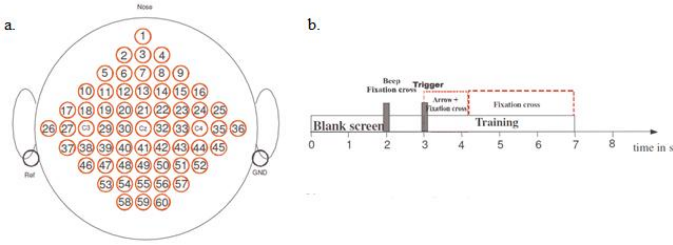


Figure 1 a. Position of EEG electrodes, b. Timing of the paradigm.

II. DATASET

In this study, dataset 'a' from BCI competition-III was used [10]. This is a cued motor imagery, multi-task dataset, provided by the Laboratory of Brain-Computer Interfaces (BCI-Lab), Graz University of Technology. The dataset consists of 4 classes (left hand, right hand, foot, and tongue) from 3 subjects (ranging from quite good to fair performance). 60 EEG channels (according to the scheme in Fig.1a) and 60 trials per class (task) were recorded. The EEG was sampled with 250 Hz and was filtered between 1 and 50Hz and also a Notch filter was applied to eliminate the electric city. The data of all runs was concatenated.

The paradigm of recording was to perform imagery movements according to a cue. The order of cues corresponding to each movement was random. After each trial begun, the first 2s were quite, at $t=2s$ an acoustic stimulus indicated the beginning of the trial, and a cross “+” is displayed; then from $t=3s$ an arrow to the left, right, up or down was displayed for 1 s; 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 $t=7s$ (Fig.1b).

The analysis was done on 5 Laplacian derivation channels named as LC_3 , LC_1 , LC_z , LC_4 and LC_2 where

$$LC_3 = C_3 - \frac{1}{4}(FC_3 + CP_3 + C_5 + C_1) \quad (1)$$

$$LC_1 = C_1 - \frac{1}{4}(FC_1 + CP_1 + C_3 + C_z) \quad (2)$$

$$LC_z = C_z - \frac{1}{4}(FC_z + CP_z + C_1 + C_2) \quad (3)$$

$$LC_2 = C_2 - \frac{1}{4}(FC_2 + CP_2 + C_z + C_4) \quad (4)$$

$$LC_4 = C_4 - \frac{1}{4}(FC_4 + CP_4 + C_2 + C_6) \quad (5)$$

and $C_5 = ch_{27}$, $C_6 = ch_{35}$, $C_1 = \frac{1}{2}(ch_{29} + ch_{30})$, $CP_z = ch_{41}$,

$C_2 = \frac{1}{2}(ch_{32} + ch_{33})$, $FC_3 = ch_{18}$, $FC_z = ch_{21}$, $FC_4 = ch_{24}$,

$FC_1 = \frac{1}{2}(ch_{19} + ch_{20})$, $FC_2 = \frac{1}{2}(ch_{22} + ch_{23})$, $CP_3 = ch_{38}$,

$CP_4 = ch_{44}$, $CP_1 = \frac{1}{2}(ch_{39} + ch_{40})$, $CP_2 = \frac{1}{2}(ch_{42} + ch_{43})$.

' ch_{xy} ' stands for BCI recording channels (Fig.1b). The Laplacian derivation was chosen, because it showed better results. Also the analysis was done on data from 3.5 to 7 seconds, which is the respond period to the cues (for subject 2 and class 1 analysis was done on data from 3.5 to 6.6 seconds). Data was subsequently band pass filtered between 8 to 30 Hz by a Butterworth filter of order 4.

III. METHODS

A. Common Spatial Patterns [5]

In the following, EEG signal trials corresponding to each class are represented by a matrix $X_{k_i, TxN}^{(i)}$, where 'i' indicates the trial class, k_i is the trial number in that class, T is the number of samples and N is the number of recording channels. The problem consists of finding a spatial filter vector w that maximizes the below expression:

$$\max \frac{w^T \overline{C_i} w}{w^T \overline{C_j} w}, \quad w \in R^N \quad (6)$$

where N is the number of available channels, $\overline{C_i}$ and $\overline{C_j}$ represents the spatial covariance matrices for class 'i' and 'j'. These covariance matrices are normalized and averaged across all trials in the corresponding class. w is the answer of the problem which should be found. Suppose that $X_{k_i}^{(i)}$,

$X_{k_j}^{(j)} \in R^{TxN}$ are band pass filtered EEG trials recorded for two classes. The covariance matrices of each trial and each class are given by

$$C_{k_i} = E[(X_{k_i}^{(i)} - E(X_{k_i}^{(i)}))(X_{k_i}^{(i)} - E(X_{k_i}^{(i)}))^T] \quad (7)$$

$$C_{k_j} = E[(X_{k_j}^{(j)} - E(X_{k_j}^{(j)}))(X_{k_j}^{(j)} - E(X_{k_j}^{(j)}))^T] \quad (8)$$

where $E[.]$ represents expectation operator. The normalized and averaged covariance matrices are

$$\overline{C_i} = \frac{1}{n_i} \sum_{k_i=1}^{n_i} \frac{C_{k_i}}{\text{trace}(C_{k_i})} \quad (9)$$

$$\overline{C_j} = \frac{1}{n_j} \sum_{k_j=1}^{n_j} \frac{C_{k_j}}{\text{trace}(C_{k_j})} \quad (10)$$

Here n_i and n_j are the number of train trials in classes 'i' and 'j'. One solution to (6) can be obtained by simultaneous diagonalization of $\overline{C_i}$ and $\overline{C_j}$. Let $C_c \triangleq \overline{C_i} + \overline{C_j}$. Suppose that the eigenvalue decomposition of C_c is

$$C_c = U_c \Lambda_c U_c^T \quad (11)$$

where Λ_c is a diagonal matrix and the columns of U_c are the eigenvectors corresponding to the eigenvalues in Λ_c . Let $P \triangleq \Lambda_c^{-0.5} U_c^T$, C_c can be whitened by

$$PC_c P^T = P \overline{C_i} P^T + P \overline{C_j} P^T = I = S_i + S_j \quad (12)$$

Here $S_i \triangleq \overline{PC_i} P^T$ and $S_j \triangleq \overline{PC_j} P^T$. The eigenvalue decomposition of S_i and S_j can be obtained from

$$S_i = U_{S_i} \Lambda_{S_i} U_{S_i}^T \quad (13)$$

$$S_j = U_{S_j} \Lambda_{S_j} U_{S_j}^T \quad (14)$$

If (12) be pre-multiplied and post-multiplied by $U_{S_i}^T$ and U_{S_i} respectively, we have

$$I = U_{S_i}^T S_i U_{S_i} + U_{S_i}^T S_j U_{S_i} = \Lambda_{S_i} + U_{S_i}^T S_j U_{S_i} \quad (15)$$

so

$$U_{S_i}^T S_j U_{S_i} = I - \Lambda_{S_i} = \Lambda_{S_j} \quad (16)$$

The sum of diagonal elements in Λ_{S_i} and Λ_{S_j} is equal to 1 (equation 16), therefore, when the diagonal elements in Λ_{S_i} decrease, those in Λ_{S_j} increase, and vice versa. Finally a projection matrix W can be obtained from

$$W = U_{S_i}^T P \quad (17)$$

Here the rows of projection matrix W are the spatial filters that we are looking for. Each of these filters results a projected signal in a new space which has maximum variance for one class and minimum variance for the other. A set of projected signals Y in new space, can be obtained from

$$Y = XW^T \quad (18)$$

where X is an EEG trial. The first component of Y , corresponding to the outcome of the first filter, is expected to have maximum variance for one class and minimum variance for the other, and vice versa for the last component, with the components in between having decreasing or increasing variances, respectively. These variance values can be utilized as features for classification. A computationally efficient way for extracting these features is taking the extreme diagonal elements of $WC_x W^T$ where the feature vector f_x can be obtained from the first and last m diagonal elements d_x , such that $f_x = \ln(d_x)$. The logarithm of the diagonal elements is computed to approximate the distribution of the features to a normal distribution [3][4].

B. Classification

1) Two-task classification

In this study, based on the previous research, only one spatial filter and therefore one feature was used to classify EEG test trials. To accomplish this goal, after training CSP filters, the train data of class 'i' and 'j' are transformed to Y space, such that

$$Y_{k_i}^{(i)} = X_{k_i}^{(i)} W^T \quad (19)$$

$$Y_{k_j}^{(j)} = X_{k_j}^{(j)} W^T \quad (20)$$

Y is a $T \times N$ matrix where each of its columns, is the spatial decomposition of X on the corresponding row of W . Each of the spatial filters are tested to get the most discriminant one.

Suppose that w_{\max} is a spatial filter which gives the most discriminant features for two classes of data, the corresponding extracted components can be obtained from

$$y_{k_i}^{(i)} = X_{k_i}^{(i)} w_{\max}^T \quad (21)$$

$$y_{k_j}^{(j)} = X_{k_j}^{(j)} w_{\max}^T \quad (22)$$

The corresponding features are

$$f_{k_i}^{(i)} = \ln(\text{var}(y_{k_i}^{(i)})) \quad (23)$$

$$f_{k_j}^{(j)} = \ln(\text{var}(y_{k_j}^{(j)})) \quad (24)$$

The means of these features for each class and all trials in that class can be obtained from

$$\bar{f}_i = \frac{1}{n_i} \sum_{k_i=1}^{n_i} f_{k_i}^{(i)} \quad (25)$$

$$\bar{f}_j = \frac{1}{n_j} \sum_{k_j=1}^{n_j} f_{k_j}^{(j)} \quad (26)$$

These means are used to classify new EEG test data. When an EEG trial arrives, feature f_x for that trial and its Euclidean distances from two class feature means are calculated. The new data is labeled to the class with nearer distance.

$$\text{if } |f_x - \bar{f}_i| < |f_x - \bar{f}_j| \Rightarrow \text{class : 'i'} \quad (27)$$

$$\text{if } |f_x - \bar{f}_i| > |f_x - \bar{f}_j| \Rightarrow \text{class : 'j'} \quad (28)$$

As suggested in [1], a logistic function was used for classifying, such that

$$y = \frac{1}{1 + \exp(-(f_x - th))}, \text{ th} = \frac{1}{2}(\bar{f}_i + \bar{f}_j). \quad (29)$$

Suppose that $\bar{f}_i > \bar{f}_j$,

$$\text{if } y > 0.5 \Rightarrow \text{class : 'i'} \quad (30)$$

$$\text{if } y < 0.5 \Rightarrow \text{class : 'j'} \quad (31)$$

2) Multi task classification

There are different solutions for multi-task classification problems. Here, a sequential classification method is used to classify multi-task imagery movements. This is done by using a two-stage classifying strategy. To do this, a set of ten weight vectors (or the most discriminant CSP spatial filters) are trained using CSP method. Four of these vectors are trained as 'one class versus all' method, and named as $w_{1vs.all}$, $w_{2vs.all}$, $w_{3vs.all}$ and $w_{4vs.all}$. For example, to train $w_{1vs.all}$, ten trials of class 1 are used as class 'i' trials and thirty trials of class 2, 3 and 4 (ten of each) are used as class 'j' trials. The other six weight vectors are trained for two-class cases. These vectors are named as $w_{1vs.2}$, $w_{1vs.3}$, $w_{1vs.4}$, $w_{2vs.3}$, $w_{2vs.4}$ and $w_{3vs.4}$. To get these vectors, ten trials of each class are used as train data. For classification, in the first stage one vs. all classifiers and in the second stage two-task classifiers are used.

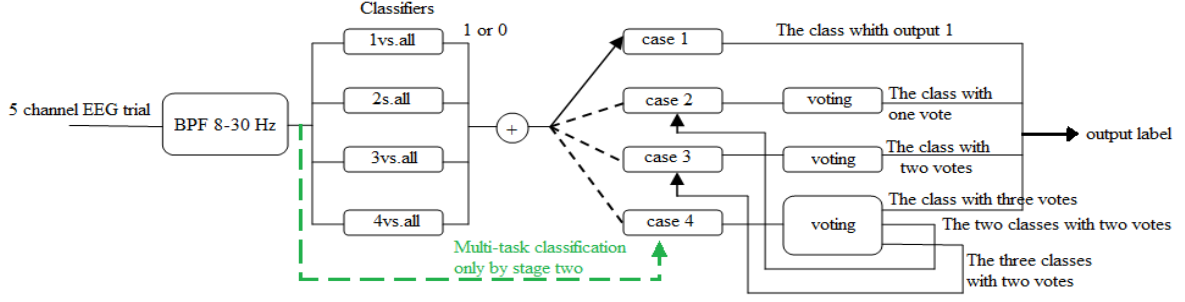


Figure 2. A block diagram of multi-task classifying strategy.

As a new EEG trial arrives, in the first stage four 'one versus all' classifiers, label that trial. Classifiers output 0 if input data is detected as class 'all' and output 1 if it is detected as one of the 1 to 4 classes. Four cases can happen,

- Case 1: Only one classifier outputs '1'.
- Case 2: Two classifiers output '1'.
- Case 3: Three classifiers output '1'.
- Case 4: Four classifiers output '1' or '0'.

In the next stage input data is labeled using two-task classifiers by four rules corresponding to each of the cases.

- Rule 1: The final label of input data is the same as classifier label with output '1'. For example, if only classifier 2 vs. all outputs '1', the final label is 'class 2'.
- Rule 2: The input data is subsequently labeled using the two-task classifier which is corresponded to the two classifiers with output '1'. For example, if classifier 1vs.all and classifier 3 vs. all output '1', the final label is the output of classifier 1 vs. 3 for that data.
- Rule 3: The input data is labeled using the two-task classifiers which are corresponded to the three 'one vs. all' classifiers with output '1'. The label with two votes from these classifiers is known as final label. If all three labels had one vote, the input data cannot be labeled.
- Rule 4: All two-task classifiers are used for classifying input data. If one of the labels has three votes, it is known as final label. If only two labels had two votes, case 2 is used to get the final label. If three labels had two votes, case 3 is used to get the final label.

Another way to classify multi-task data could be using only stage two of the proposed strategy. In this case it is supposed that case 4 happened and classification follows. Fig 2 presents a block diagram of the multi-task classifying strategy.

C. CSP filter adaptation using EKF

Here, spatial filter w_{csp} is adapted by each new EEG trial using below equations,

$$w_{init} = \frac{w_{csp}}{\sqrt{e^{th}}} \quad (32)$$

Prediction equations:

$$w_t^p = w_{t-1}^u \quad (33)$$

$$P_t^p = P_{t-1}^u + Q_{t-1} \quad (34)$$

where

$$Q_{t-1} = \max\{0, (u_{t-1}^u - u_{t-1}^p)\} I \quad (35)$$

$$u = y(1 - y) \quad (36)$$

$$y = \frac{w C_x w^T}{w C_x w^T + 1} \quad (37)$$

Script 'P' stands for 'prediction', 'u' stands for 'update', 't' indicates current step and 't-1' indicates former step. P is the covariance matrix of w , I is an identity matrix, $(u_{t-1}^u - u_{t-1}^p)$ is information gain from the last updating, and u represents the uncertainty of the classifier output y .

Update equations:

Suppose that $\bar{f}_i > \bar{f}_j$,

$$\text{if } y > 0.5 \Rightarrow w_t^u = w_t^p + k_t(z_t - y_t^p) \quad (38)$$

$$\text{if } y < 0.5 \Rightarrow w_t^u = w_t^p - k_t(z_t - y_t^p) \quad (39)$$

$$p_t^u = P_t^p - k_t G w_t^p C_x P_t^p \quad (40)$$

$$k_t = \frac{P_t^p}{G s_t^2 + C_{cons}} C_x w_t^p \quad (41)$$

where

$$G = \frac{2}{(1 + w_t^p C_x w_t^p)^2} \quad (42)$$

$$s^2 = w C_x P C_x w^T \quad (43)$$

$$z_t = \begin{cases} 1, y > 0.5 \\ 0, y < 0.5 \end{cases} \quad (44)$$

C_{cons} adjusts adaptation speed. Using a suitable value for C_{cons} is important. Here this value was set manually for each case. Consider that Tsui *et al.* [1] used supervised learning paradigm, where they knew class label of incoming data. As proposed in [8], In this study, an assumed target function is used for z_t . Unlike [1] and [8], that run adaptation algorithm when one class is detected, in this work adaptation algorithm is run whether class 'i' or class 'j' is detected. This is done by manipulating update equation in a way that detection of class 'j' has the same effect on w as detection of class 'i'.

TABLE I. RESULTS OF TWO-TASK CLASSIFICATION

Data without artifacts		class	1vs.2	1vs.3	1vs.4	2 vs.3	2 vs.4	3 vs.4	1 vs. all	2 vs. all	3 vs. all	4 vs. all
Subject 1	No adaptation		82.94	93.79	96.87	97.69	100	75.96	84.49	89.53	80.62	74.41
	Supervised adaptation		83.72	93.79	98.43	98.46	100	86.04	85.65	89.53	82.55	86.43
	Unsupervised adaptation		83.72	93.79	98.43	98.46	100	86.04	86.04	89.53	82.55	87.20
Subject 2	No adaptation		59.09	77.94	86.56	67.64	74.62	63.76	77.03	68.88	60.76	64.44
	Supervised adaptation		66.66	80.88	86.56	73.52	77.61	75.36	79.25	77.03	82.96	78.51
	Unsupervised adaptation		63.63	80.88	86.56	72.05	79.10	73.91	79.25	77.03	82.96	77.03
Subject 3	No adaptation		74.60	71.87	83.33	77.04	93.65	70.31	74.80	81.88	57.48	78.74
	Supervised adaptation		77.77	75.00	87.87	90.16	95.23	70.31	74.80	84.25	76.37	85.00
	Unsupervised adaptation		79.36	73.43	87.87	90.16	96.82	70.31	74.80	85.82	76.37	85.82
Data with artifacts		class	1vs.2	1vs.3	1vs.4	2 vs.3	2 vs.4	3 vs.4	1 vs. all	2 vs. all	3 vs. all	4 vs. all
Subject 1	No adaptation		80.62	93.75	96.87	98.12	98.75	81.25	83.75	87.18	80.93	77.81
	Supervised adaptation		81.87	93.75	98.75	98.75	98.75	86.25	85.00	87.18	81.56	85.31
	Unsupervised adaptation		81.87	93.75	98.75	98.75	98.75	86.25	85.00	87.18	81.25	85.93
Subject 2	No adaptation		58.00	77.00	73.00	70.00	75.00	67.00	77.50	68.00	62.50	69.00
	Supervised adaptation		60.00	83.00	76.00	72.00	75.00	73.00	78.50	75.00	79.50	75.00
	Unsupervised adaptation		59.00	82.00	76.00	72.00	75.00	73.00	78.50	75.50	79.50	75.50
Subject 3	No adaptation		71.00	73.00	82.00	84.00	89.00	66.00	66.50	89.00	72.00	76.00
	Supervised adaptation		77.00	74.00	84.00	92.00	94.00	69.00	67.50	89.50	75.00	81.50
	Unsupervised adaptation		77.00	77.00	84.00	92.00	94.00	69.00	70.00	89.50	75.00	82.00

The CSP filter can also be updated using a supervised learning paradigm. In this case $z_t = 1$ if a true positive is detected for class 'i' or a false positive is detected for class 'j' and $z_t = 0$ if a false positive is selected for class 'i' or a true positive is detected for class 'j'. This means

$$z_t = \begin{cases} 1, y > 0.5 \\ 0, y < 0.5 \end{cases}$$

In this case the update equations for w are

$$\text{If class 'i' happened} \Rightarrow w_t^u = w_t^p + k_t(z_t - y_t^p) \quad (45)$$

$$\text{If class 'j' happened} \Rightarrow w_t^u = w_t^p - k_t(z_t - y_t^p) \quad (46)$$

For multi-task classifier adaptation, as a new EEG trial arrives, all one-versus-all classifiers and three of two-task classifiers are adapted. These three classifiers are corresponding to the label of input data for supervised method

and corresponding to the final label for unsupervised one.

IV. RESULTS

The proposed method was applied on dataset 'a' from BCI competition-III. Here both two-task and multi-task classification scenarios were considered. For two-task problem the method was applied to all two-class pairs and to all 'one versus all' cases from each subject. As it was mentioned, the first ten trials of each class were used to train CSP filters, the rest were used as test data. Classifications were performed in two methods, classifications without adaptation and with adaptation. The adaptation also carried out by supervised and unsupervised manners. The results are reported for these conditions and their performance is evaluated by accuracy defined as $[TP/(TP + FP)]$, where TP is the number of true positives and FP is the number of false positives.

TABLE II. RESULTS OF MULTI-TASK CLASSIFICATION

Data without artifacts		run 1					run 2	run 3	run 4
		total	class 1	class 2	class 3	class 4	total	total	total
Subject 1	No adaptation	78.29	96.87	75.38	50.76	90.62	78.29	78.29	78.29
	Supervised adaptation	82.55	92.18	76.92	75.38	85.93	81.87	83.33	82.94
	Unsupervised adaptation	83.72	95.31	76.92	76.92	85.93	84.37	83.72	84.10
Subject 2	No adaptation	48.14	42.42	33.33	80.00	35.29	48.14	48.14	48.14
	Supervised adaptation	55.55	48.48	42.42	71.42	58.82	54.81	52.59	51.85
	Unsupervised adaptation	54.81	48.48	39.39	71.42	58.52	53.33	52.59	52.59
Subject 3	No adaptation	56.69	78.78	53.33	41.93	51.51	56.69	56.69	56.69
	Supervised adaptation	61.41	78.78	66.66	41.93	57.57	59.84	60.62	61.41
	Unsupervised adaptation	61.41	75.75	66.66	45.16	57.57	60.62	60.62	61.41
Data with artifacts		run 1					run 2	run 3	run 4
		total	class 1	class 2	class 3	class 4	total	total	total
Subject 1	No adaptation	79.06	97.50	76.25	52.50	90.00	79.06	79.06	79.06
	Supervised adaptation	83.12	90.00	77.50	82.50	82.50	82.81	82.18	81.87
	Unsupervised adaptation	83.75	92.50	77.50	81.25	83.75	84.06	83.75	84.37
Subject2	No adaptation	46.00	36.00	38.00	76.00	34.00	46.00	46.00	46.00
	Supervised adaptation	50.00	42.00	44.00	58.00	56.00	49.00	48.50	48.00
	Unsupervised adaptation	48.50	42.00	40.00	56.00	56.00	48.00	47.50	47.50
Subject 3	No adaptation	58.00	76.00	68.00	40.00	48.00	58.00	58.00	58.00
	Supervised adaptation	62.00	76.00	80.00	42.00	50.00	61.00	62.00	61.50
	Unsupervised adaptation	61.00	72.00	80.00	44.00	48.00	61.00	61.00	61.00

The mentioned dataset has marked trials with artifact. To see the effects of artifacts on accuracies, classifications were performed on data with artifacts and on data without artifacts. In the case with artifacts, train and test data for two-task problem and only test data for multi-task problem contained artifacts.

Table 1 represents results for two-task problem and for different subjects. Consider that the selection of input task data was random, it caused near but different results for adaptive classification in different runs. Here the best result for each case is reported. The results for all subjects and most cases show improvements as adaptation algorithm was taking into account. In some cases (such as class 1vs.3 for subject 1) adaptation did not caused better accuracies but worked as well as algorithm with no adaptation. In most cases unsupervised

adaptation results as good as supervised one and in some cases (for example class 2vs.4 and data with artifact for subject 2) it out performed supervised algorithm. Subject 1 had the best performance. The best improvements in accuracies with about 13 to 19 percents, happened for subject 1 class 3vs.4 and 4vs.all, subject 2 class 3vs.4, subject 3 class 2vs.3 and subject 3 class 2vs.all (data with no artifact).

Table 2 represents results for multi-task problem. Here the results of different runs for algorithm with adaptation are reported; also the results of run 1 are reported for each task separately. The classification for subject 2 and 3 was only done by using stage two of multi-task classification strategy. As it can be seen, adaptive classifier improved the total detection accuracies. Subject 1 had the best performance; for this subject, task 3 provided poorest results and decreased the

total accuracy. Adaptation method had adjusted different task accuracies and significantly increased task 3 detection rate. Although it caused lower detection rate for tasks 1, 2 and 4, it totally improved the results. In most cases unsupervised adaptation outperformed supervised one. For subject 2, task 3 had the best results. Adaptation method had adjusted different task accuracies which caused lower detection rate for task 3 and higher detection rate for the other tasks. The results were totally improved. Unsupervised method had worked as good as supervised one. For subject 3, task 1 had the best results. Adaptation method had adjusted different task accuracies. It caused higher detection rate for all tasks except 1 (data without artifacts). The results were totally improved. Unsupervised method had almost worked as well as the supervised one.

V. DISCUSSION

As it is clear from results, the presented method is able to improve the performance of BCI systems. It is because of the fact that EEG signals have a non stationary nature. This proves the necessity of online adaptation in BCI systems. Unsupervised method seemed to perform as good as supervised one and it is very important, because in a real application the labels of incoming data are not provided, so unsupervised learning is prior to supervised one. The algorithm presented here, was applied synchronously on a multi-task EEG dataset, but the theme of this method is capable to be performed asynchronously, where fast online adaptation is needed. Using EKF algorithm guarantees this fast, online and optimized estimation of interested parameters.

It worth to note that, because of studying reasons only one spatial filter was used as feature extractor and classifier. It is predictable that using more features will improve the performance of the BCI system under study.

VI. CONCLUSION

Because of the non-stationary nature of EEG signals in BCI applications, online adaptation for feature extraction and classification is essential. This paper proposed a novel CSP based feature extraction and classification method for two-task and multi-task scenarios, which can be adaptive through an extended Kalman filter. The usefulness of the proposed method was demonstrated by applying it to dataset 'a' from BCI competition-III. The results showed a remarkable improvement in terms of classification accuracy in comparison with standard CSP method with no adaptation. The proposed method was implemented synchronously. Because of small computational requirements and perfect recursive properties of EKF, the proposed method can be easily and effectively applied to asynchronous BCI systems where, fast, online adaptation is needed. Further studies should consider asynchronous BCI as well as synchronous one.

ACKNOWLEDGMENT

Special thanks to Prof. Gert Pfurtscheller and Dr. Alois Schlögl for provision of the data.

REFERENCES

- [1] C. S. L. Tsui, J. Q. Gan, and S. J. Roberts, "A self-paced brain-computer interface for controlling a robot simulator: an online event labelling paradigm and an extended Kalman filter based algorithm for online training," *Medical & biological engineering & computing*, vol. 47.3, 2009, pp. 257-265.
- [2] F. Davood, B. Nasihatkon, and R. Boostani, "A general framework to estimate spatial and spatio-spectral filters for EEG signal classification," *Neurocomputing* 119, 2013, pp. 165-174.
- [3] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, 1999, pp. 787-798.
- [4] 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, Dec. 2000.
- [5] O. Falzon, K. P. Camilleri, and J. Muscat, "The analytic common spatial patterns method for EEG-based BCI data," *Journal of neural engineering*, 9.4th ed., vol. 045009, 2012.
- [6] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation," *IEEE Signal processing magazine*, 29.5th ed., 2012, pp. 128-132.
- [7] G. A. Terejanu, "Extended kalman filter tutorial," [Online]. Disponible: <http://users.ices.utexas.edu/~terejanu/files/tutorialEKF.pdf>, 2008.
- [8] D. R. Lowne, C. J. Haw, and S. J. Roberts, "An adaptive, sparsefeedback EEG classifier for self-paced BCI," 3rd international workshop on brain-computer interfaces, Graz, 2006, pp. 58-59.
- [9] C. S. L. Tsui, J. Q. Gan, "Comparison of three methods for adapting LDA classifiers with BCI applications," 4th international workshop on brain-computer interfaces, Graz, 2008, pp. 116-121.
- [10] B. Blankertz, K. R. Müller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. del R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition-III: Validating alternative approaches to actual BCI problems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, 2nd ed., 2006.