Subject-To-Subject Transfer for CSP based BCIs: Feature Space Transformation and Decision-Level Fusion

Dominic Heger, Felix Putze, Christian Herff, and Tanja Schultz

***Abstract*— Modern Brain Computer Interfaces (BCIs) usually require a calibration session to train a machine learning system before each usage. In general, such trained systems are highly specialized to the subject's characteristic activation patterns and cannot be used for other sessions or subjects. This paper presents a feature space transformation that transforms features generated using subject-specific spatial filters into a subject-independent feature space. The transformation can be estimated from little adaptation data of the subject. Furthermore, we combine three different Common Spatial Pattern based feature extraction approaches using decision-level fusion, which enables BCI use when little calibration data is available, but also outperformed the subject-dependent reference approaches for larger amounts of training data.**

I. INTRODUCTION

Common Spatial Patterns (CSPs) [1], [2] are a widely used spatial filtering method for Brain Computer Interfaces (BCIs). They can be seen as a standard feature extraction approach to detect event related (de-)synchronization in EEG data, as they have good discriminative abilities and allow some physiological validation (e.g. using topographical plots). However, because of the strong inter-subject and inter-session variabilities of the EEG, CSPs are usually calculated subject-dependently from calibration data recorded immediately before the start of the BCI session. This is a time consuming and cumbersome procedure for BCI users. Therefore, minimizing the calibration time before BCI usage is one of the major challenges in BCI research. Subject-to-subject transfer, i.e. to learn generalizing patterns from a pool of subjects and to transfer them to a new subject can help to overcome long enrollment times. Few studies have addressed subject-to-subject transfer to enable BCI use with very little or without calibration data.

Krauledat et al. [3] identified prototypical spatial filters in multiple previously recorded sessions of a subject using a clustering approach. The prototype filters have good generalization abilities and can allow BCI use directly after a very short recalibration time for bias adaptation. However, a large number of sessions for the subject has to be available before the user can benefit from the zero training approach.

Falzi et al. [4] used an ensemble learning based approach to create a subject-independent BCI. They constructed subject-dependent classifiers for different frequency bands and sparsely combined their outputs using quadratic regression with l_1 norm penalty regularization. They could achieve results comparable to subject-dependent reference methods using a bias-correction that was applied as an offline post-processing step.

Kang et al. proposed an approach called Composite CSPs [5] to reduce calibration time for CSP based BCIs. They evaluated two different methods to use linear combinations of covariance matrices estimated from other subjects' data for CSP calculation. For a small number of calibration trials their method outperformed the traditional CSP approach.

Lotte et al. [6] combined different feature extraction methods and classifiers to find the best setup for subject-independent BCIs using pooled data from different subjects for training. They achieved the best results using linear classifiers and filter bank common spatial patterns, however the performance was still considerably lower than subject-dependent classification. In a later study [7] they regularized CSPs and LDA using covariance matrices from other subjects and proposed a sequential forward selection algorithm to select a subset of subjects that maximizes the performance for the target subject.

Reuderink et al. [8] estimated and adapted the whitening transform from pre-trial data before CSP calculation in order to reduce the influence of non-stationarities by normalizing second order covariance statistics. Their approach could achieve about the same performance for subject-independent as for subject-dependent BCI operation. However, their evaluation requires pre-trial data, which is only available in a synchronous (trial-based) BCI protocol. It is not clear if the approach is also applicable to user-paced BCI interaction (e.g. using pre-trial data from idle phases). The same data set and a similar pre-processing as in this study was used.

Tu et al. [9] presented a framework for subject-to-subject transfer on feature extraction and classification level. They extracted generalizing and subject-specific filters banks from a set of candidate filters generated using extreme energy ratio features by solving optimization problems with l_1 norm regularization. They employed a two level ensemble learning strategy. In the first level, they generated learners for both filter banks and combined both learners in the second level. Evaluations showed a successful subject-to-subject transfer.

The previously proposed approaches that aim to overcome the long calibration time before BCI use are quite diverse, however the problem is far from being solved. With this paper we contribute the following to this line of research:

- We show that a feature space transformation can effectively transform features from subject-dependently calculated spatial filters into a subject-independent feature space and that the transformation parameters can be estimated from little subject-dependent data.

All authors are with the Cognitive Systems Lab, Institute for Anthropomatics, Karlsruhe Institute of Technology, Adenauerring 4, 76131 Karlsruhe, Germany. dominic.heger@kit.edu

- We propose to combine three different approaches for feature extraction using decision-level fusion to create a classifier that can achieve good performance with very little calibration data and also outperform the reference methods in our experiment when more calibration data is available.

II. MATERIAL AND METHODS

A. DATA CORPUS

We used the EEG Motor Movement/Imagery Dataset freely available from PhysioNet [10]. It consists of 64 channel EEG recordings of 109 different subjects assessed using the BCI2000 instrumentation system [11] at 160 Hz sampling rate. For each subject, we chose to use the runs 6, 10, and 14 of the recording session, where subjects had to perform two different classes of motor imagery: moving both fists versus moving both feet. Therefore, we used 45 trials per subject. Four subjects have been left out as their recordings consist of less trials. The terms subject-independent and session-independent can be used synonymously in the following, as there is one recording session per subject.

The left part of Figure 1 shows the partitioning of the data for the evaluations in this paper. We split the data into three different sets: The first 50 sessions form an independent set of subjects (S1) that do not occur in the data sets used for testing. Each of the remaining sessions were chronologically split into a first part for calibration (S2) and a second part for testing (S3). In section III we compare results for different ratios of calibration and test set size.

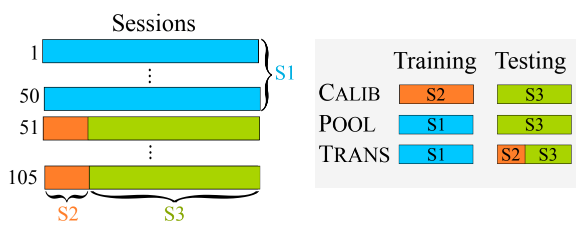

Fig. 1. Left: Partitioning of the data set into S1, S2, and S3. Right: Data usage of the three approaches CALIB, POOL, and TRANS in training and testing phases.

B. PREPROCESSING

We extracted trials of 3.5 seconds length starting at 0.5 seconds after each stimulus. We rereferenced the trial data to common average reference and removed signal offsets and linear trends from each trial. A 5^{th} order Butterworth filter between 8-30 Hz was applied to bandpass filter the data.

For each of the approaches in this paper we applied spatial filters based on CSPs regularized with diagonal loading, where the regularization parameter was estimated using 5-fold cross-validation (see [12] for a comparison of different regularization techniques for CSPs). We used the two most discriminative spatial filters for each of the two classes (first and second columns of the CSP transformation matrix, i.e. 4 spatial filters).

C. FEATURE EXTRACTION

For feature extraction, we employed three different approaches based on logarithmic variance features [1] calculated from the pre-processed and spatially filtered EEG data. We call the approaches CALIB, POOL, and TRANS throughout this paper:

CALIB: Logarithmic variance features based on the output of spatial filters, subject-dependently calculated from the calibration data S2. This can be seen as a standard design for subject and session dependent BCIs (e.g. [1], [2]).

POOL: Logarithmic variance features based on the output of spatial filters calculated from the pool of all sessions of the independent data set S1 combined into one single session. This is a simple method to create a subject-independent BCI.

TRANS: Logarithmic variance features based on spatial filters calculated from each session of the data set S1 transformed into a subject-independent feature space. This approach allows to train an subject-independent classifier that can be used in combination with a subject-dependently learned feature space transformation.

The right part of Figure 1 summarizes, which parts of the data set are used by CALIB, POOL, and TRANS during training and testing. For the evaluation (using data set S3) of the approach TRANS, the feature space transform was estimated on the data set S2 for each subject.

In the approach TRANS, 200 spatial filters are calculated during training (50 sessions x 4 spatial filters). Since many of the available filters are not discriminative for larger groups of subjects, we selected only a small subset of spatial filters that are most discriminative for all of the subjects in the data set S1. As a selection criterion, we define the discriminative ratio $dr(w_i)$ of filter w_i, as the absolute value of the ratio of average logarithmic variance features between the two classes for all n sessions, as follows:

$$dr(w_i) = \frac{1}{n}\sum_{p=1\ldots n} |vq(w_i, X^p) - 1|, \tag{1}$$

$$vq(w_i, X^p) = \frac{\frac{1}{n_{p+}}\sum_{j_+} log(var(w_i \cdot x^p_{j_+}))}{\frac{1}{n_{p-}}\sum_{j_-} log(var(w_i \cdot x^p_{j_-}))}, \tag{2}$$

where w_i is the i^{th} spatial filter generated from the independent data set S1. $x^p_{j_c} \in X^p$ is the ${j_c}^{\text{th}}$ trial in session p that has the class label $c \in \{+, -\}$ and n_{pc} is the number of trials of class c in session p. If the average of the log variance features is identical for both classes, the variance quotient becomes $vq(w_i) = 1$. This indicates that features generated by the spatial filter w_i cannot separate the two classes for this particular session. In this case the variance quotient does not contribute to the corresponding discriminative ratio. The larger the difference of average log variance features between the two classes for all sessions, the larger $dr(w_i)$. For the evaluations in this paper we chose to use the $k = 10$ spatial filters with the largest $dr(\cdot)$, which appeared to cover a reasonable amount of filters with high discriminative power according to the distribution of $dr(\cdot)$ values.

To transform subject-specific features into a subject-independent feature space, we applied a simple linear function to the logarithmic variance features y generated from spatial filter w and subject p:

$$f_w^p(y) = a_w^p + b_w^p \cdot y. \tag{3}$$

We determined the transformation parameters a_w^p and b_w^p to scale and shift the input features of the first class (+) near a predefined target value $t_+ = 1$ and the features of the second class (-) near a different target value $t_- = -1$. The parameters can be estimated by minimizing the corresponding quadratic error function as an overdetermined equation system using QR decomposition:

$$\arg\min_{a_w^p, b_w^p} (f_w^p(y) - t_c)^2, \quad \forall\, y \in Y_{c,w}^p,\; c \in \{+,-\}, \tag{4}$$

where $Y_{c,w}^p$ is the set of logarithmic variance features generated using spatial filter w and subject p that belong to class c.

For classifier training, the transformation parameters a_w^p and b_w^p were estimated for each spatial filter and each subject in the independent data set S1. For the evaluation (data set S3) they were estimated on the calibration data S2 of the target subject. To increase the robustness of the estimates for a_w^p and b_w^p, we employed bootstrap aggregating (bagging) [13]. For this purpose, we calculated the median result of 100 iterations of drawing random samples with replacement from the training data and calculating the optimization problem (4).

D. *CLASSIFICATION AND DECISION-LEVEL FUSION*

For each of the three approaches CALIB, POOL, and TRANS, features of the test data set S3 were extracted and classified using Linear Discriminant Analysis (LDA). In addition, we combined the three approaches by decision-level fusion. In this approach we aggregated the outputs of the three independent classifiers of CALIB, POOL, and TRANS for each trial using weighted majority voting. The squared distance from the separating hyperplane was used as voting weight, which is an indicator for the classifiers' confidence. In the following, we call this approach FUSED.

III. RESULTS

Figure 2 shows the recognition accuracies averaged across the 55 test subjects (S3) of the approaches CALIB, POOL, TRANS, and FUSED for different ratios of splitting the sessions into calibration set (S2) and test set (S3). Please note that only 45 trials per session were available for the evaluations. Therefore, a large number of calibration trials leads to a small number test trials, which can reduce the reliability of the prediction (e.g. the results for 40 calibration trials were produced by only 5 test trials per session). Due to interindividual differences (including low and high performers), the performances of the 55 different sessions in the test set vary strongly (std. up to 11-24 %, depending on the number of calibration trials). However, the session-wise performance differences among the four approaches are quite stable, which allows to show significant pairwise performance improvements of the proposed approaches.

When only a small number of trials was available for calibration, the subject-dependent approach CALIB suffered

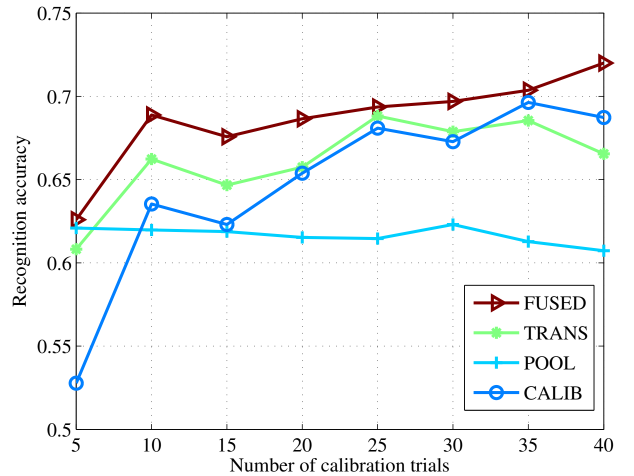

Fig. 2. Recognition accuracies averaged accross the test subjects for different ratios of splitting the session into calibration and test data.

from insufficient data, even though regularized CSPs have been used. Its performance increased strongly as more data for training was used. The subject-independent approach POOL does not use the calibration data S2, therefore its performance is quite stable and only affected by the test sets of different size. It outperformed the other approaches for a very small number of 5 trials, however its overall performance was below the other approaches for more than 5 trials (significantly below the other approaches for more than 20 trials, paired t-tests $p < 0.05$). The proposed approach TRANS realized a successful subject transfer, as it outperformed the standard subject-dependent approach CALIB for up to 30 calibration trials. The approach FUSED achieved the overall best performance of up to 72 % accuracy in the experiment. The fusion increased the performance above its fundamental approaches for a very small number of calibration trials, but also outperformed the other approaches for the maximum number of 40 calibration trials. For up to 20 calibration trials it performed significantly better than CALIB (paired t-tests $p < 0.02$). Therefore, the proposed approach FUSED offers a high flexibility and good performance.

Further analyses showed that FUSED benefits from all of its fundamental methods, e.g. POOL contributed information that led to performance improvements, even when a large number of calibration trials were available. The proposed approach TRANS is most flexible fundamental approach and strongly improved recognition performance in all evaluations of FUSED.

TABLE I

COMPARISON OF DIFFERENT MODES TO ESTIMATE THE TRANSFORMATION PARAMETERS FOR THE APPROACH **TRANS**. RESULTS SHOW RECOGNITION ACCURACY IN PERCENT.

Parameter Estimation	**Mean (std.)**	**Median**
No ($a = 0$, $b = 1$)	58.2 (12.0)	57.1
On 10 calibration trials	66.2 (14.7)	65.7
On 35 test trials	72.7 (11.8)	74.3

Table I shows the results of TRANS for the test session S3 using different modes to determine the feature space

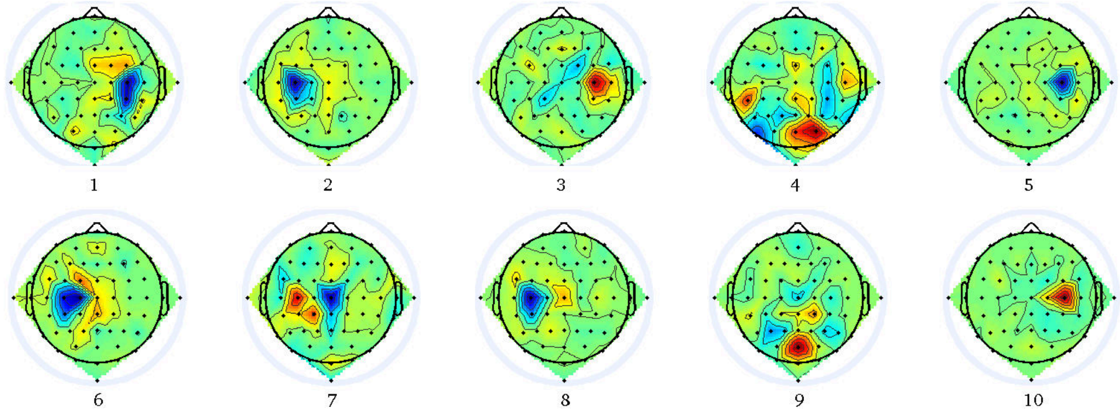

Fig. 3. Topographical plots of the 10 spatial filters selected in approach TRANS.

transformation parameters. The first row shows recognition accuracies without applying the transform (i.e. $a = 0$, $b = 1$). Row two shows the results, where the feature space transformation parameters were estimated on 10 calibration trials. Applying the transform strongly improved the performance, even in the case where as few as 5 trials per class are available for calibration. This shows the effectiveness of the proposed feature space transform. More data to estimate the transform did further improve the performance (cf. Fig. 2 approach TRANS). The third row shows the results, where the transform is estimated on the 35 test trials. This can be seen as an upper bound of finding optimal transformation parameters, however it requires knowledge of the class labels that are not available in testing.

We chose to use a linear transform because of its simplicity and its ability for easy and robust estimation. Please note that more simple linear transforms, such as additive correction of the session bias (cf. [14]) or feature scaling did not perform as effectively as the proposed feature space transform.

Figure 3 shows the topographical plots of the 10 spatial filters selected in approach TRANS. The most influential regions are at the sensorimotor areas and the occipital cortex. The spatial filter weights are fairly localized to discriminative areas and are not strongly affected by artifacts.

IV. CONCLUSION

The proposed feature space transform approach TRANS could successfully transform features based on subject-dependently trained spatial filters into an subject-independent feature space, as it outperformed the subject-dependently trained system CALIB for small numbers of calibration trials to estimate the transform. This leads to a substantially reduced calibration time for BCI users. Furthermore, the proposed fusion FUSED of subject-dependently trained (CALIB), subject-independent pooled systems (POOL), and feature space transformed (TRANS) outperformed the other BCI designs and achieved the best performance for all calibration set sizes in our experiment. Therefore, our approach was able to learn and transfer information from the independent set of sessions for the prediction of unseen subjects robustly for different amounts of calibration data.

In future work, we will evaluate our approach in an online application. Furthermore, we plan to investigate if the feature space transformation can be updated using unsupervised adaptation methods.

REFERENCES

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

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

[3] M. Krauledat, M. Tangermann, B. Blankertz, and K. Müller, "Towards zero training for brain-computer interfacing," *PLoS One*, vol. 3, no. 8, p. e2967, 2008.

[4] S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K. Müller, and C. Grozea, "Subject-independent mental state classification in single trials," *Neural networks*, vol. 22, no. 9, pp. 1305–1312, 2009.

[5] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *Signal Processing Letters, IEEE*, vol. 16, no. 8, pp. 683–686, 2009.

[6] F. Lotte, C. Guan, and K. Ang, "Comparison of designs towards a subject-independent brain-computer interface based on motor imagery," in *IEEE Engineering in Medicine and Biology Society, 2009*. IEEE, 2009, pp. 4543–4546.

[7] F. Lotte and C. Guan, "Learning from other subjects helps reducing brain-computer interface calibration time," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 614–617.

[8] B. Reuderink, J. Farquhar, M. Poel, and A. Nijholt, "A subject-independent brain-computer interface based on smoothed, second-order baselining," in *IEEE Engineering in Medicine and Biology Society, 2011*. IEEE, 2011, pp. 4600–4604.

[9] W. Tu and S. Sun, "A subject transfer framework for eeg classification," *Neurocomputing*, 2011.

[10] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[11] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "Bci2000: a general-purpose brain-computer interface (bci) system," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1034–1043, 2004.

[12] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 2, pp. 355–362, 2011.

[13] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

[14] P. Shenoy, M. Krauledat, B. Blankertz, R. Rao, and K. Müller, "Towards adaptive classification for bci," *Journal of neural engineering*, vol. 3, no. 1, p. R13, 2006.