

Feature Extraction and Classification of Four-Class Motor Imagery Signals Based on LCD and CSP

Qingsong Ai^{1,2}, Yi Xie¹ and Kun Chen^{1,2+}

¹ School of Information Engineering, Wuhan University of Technology, Wuhan, China

² Key Laboratory of Fiber Optic Sensing Technology and Information Processing (Wuhan University of Technology), Ministry of Education, Wuhan, China

Abstract. The common spatial pattern (CSP) can effectively extract the spatial information of motor imagery (MI) signals, but ignores time and frequency domain information of EEG signals. In order to overcome this problem, a new method is proposed in this paper, which combines the CSP method with the time-frequency analysis method Local Characteristic-scale Decomposition (LCD) to extract the time-frequency information and improve the classification accuracy. The effectiveness of the algorithm was verified by conducting experiments with the BCI competition dataset. The results show that the proposed method improves the recognition rate of MI signals, and has potential for the application of portable BCI systems in rehabilitation field.

Keywords: brain-computer interface, motor imagery, common spatial pattern, local characteristic-scale decomposition, time-frequency analysis

1. Introduction

A motor imagery brain computer interface (BCI) [1] is significant for stroke patients, which can help the recovery of damaged nerve and play an important role in the rehabilitation training assistant [2]. The application spectrum of BCI also extends to navigation, healthcare, military services, robotics, virtual gaming, communication and controls, etc. The classification accuracy of motor imagery depends largely on features extracted from EEG signal. In general, the two-class motor imagery classification accuracy rate has reached more than 90% [3], while multi-class motor imagery movement classification accuracy rate is low, which needs further research to be applied MI BCI in the field of rehabilitation exercise.

At present, the most common way to extract the features of motor imagery EEG is common space pattern (CSP) [4]. However, CSP has deficiencies. First of all, it needs multi-channel information to improve the classification results; secondly, it ignores the information of EEG signal in time and frequency domain. While time and frequency domain information for motor imagery classification is particularly important, effective feature extraction method in frequency domain is needed.

Because the EEG signal is nonlinear and non-stationary, the classical time domain or frequency domain analysis methods are difficult to analyse it effectively, while time-frequency analysis can make it. The time-frequency analysis methods include wavelet decomposition [5], empirical mode decomposition (EMD) [6], local mean decomposition (LMD) [7], the intrinsic time scale decomposition method (ITD) [8], et al.

At present, many researchers combine common space pattern with time-frequency analysis method such as wavelet decomposition [9], EMD decomposition [10], LMD, and ITD. However, there are many problems in parameter setting problem, end effects, unexplained negative frequency problem, time consuming problem, etc.

⁺ Corresponding author. Tel.: +02787651806; fax: +02787651806.
E-mail address: kunchen@whut.edu.cn.

Local characteristic-scale decomposition (LCD) [11] is proposed based on the study of intrinsic time-scale decomposition (ITD). By using this method, a complicated multi-component signal could be adaptively decomposed into the summation of a number of physically meaningful intrinsic scale components (ISCs). This method is superior to EMD algorithm in end effects, iteration number and decomposition time, and it is expected to achieve real-time online analysis of EEG signal.

In this study, LCD time-frequency decomposition is used to process part of the motor imagery EEG channels to get ISC components and extract the time-frequency features. Then, to get the spatial features, the ISC components and EEG channels are processed by CSP. After fusing the time-frequency and spatial features, features are ranked and then a classifier is used to classify the motor imagery tasks. Experiments were carried out with the 2008 BCI competition data to compare our algorithm with other algorithms. The results show the effectiveness of our algorithm.

2. Materials and Methods

2.1. Experiment Data

In this paper, using the 2008 BCI competition Data Sets 2a [15], provided by Graz University, we benchmarked the performance of our algorithm to see how good the algorithm is. The EEG data was collected from 22 channels over the sensorimotor area.

The dataset includes four imagination of movements: left hand, right hand, both feet and tongue. 9 subjects recorded EEG data sets at different two days. During the experiment, the subjects sat in front of a computer, according to computer prompts to react. An experiment lasted 8 seconds. The experimental timing scheme of the paradigm is shown in Fig. 1.

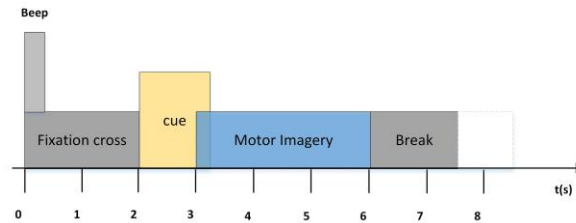


Fig. 1: Timing scheme of the paradigm.

Each subject's data set consisted of two parts, the data set T and E, including the 576 experiments totally. Further details of the dataset can be found in [15].

2.2. Time-Frequency Analysis and Feature Extraction

Based on the definition of the ISC component, a real value signal $x(t)(t > 0)$ can be decomposed into numbers of ISCs by using the LCD method following the way in [11]. The ISC component's conditions are to eliminate the situation of riding wave, guarantee the waveform single and ensure the smoothness and symmetry of the ISC component waveform. They assure that the ISC component possesses a single mode between two adjacent extrema and corresponds with the sine curve locally. Hence the IF of the ISC has physical significance.

So a signal $x(t)$ is decomposed into n ISCs and a residue $u_n(t)$ as

$$x(t) = \sum_{p=1}^n ISC_p(t) + u_n(t) \quad (1)$$

To guarantee that the ISC components meet the definitions, a criterion for the sifting process should be determined. In this paper, the standard deviation (SD) is adopted.

$$SD = \sum_{t=0}^T \left[\frac{|h_{ik}(t) - h_{i(k-1)}(t)|^2}{h_{i(k-1)}^2(t)} \right] \quad (2)$$

Note that there is still much time consuming, the maximum contribution 3 channels C3, C4 and Cz, are selected to be processed with LCD. And each channel data get 3 ISC compositions.

The obtained 9 ISC components are integrated with the original 22 channels of EEG signals to the data of 31 channels. Then, the Hjorth parameters of the data is extracted.

Hjorth[19] introduced a set of three parameters to describe the EEG signal: Activity, the signal power; Mobility, the mean frequency; and Complexity, the change in frequency.

Thus, after processing the EEG signals with LCD, the time-frequency features in a trial are extracted as $F_1 = [f_{11}, f_{12}, \dots, f_{1K}] \in R^{1 \times K}$, where K is 93 in this paper.

2.3. Spatial Feature Extraction

In this paper, Common Spatial Pattern (CSP) [14] is used for feature extraction. We denote the EEG data of trial i for class A by X_{Ai} , which is a matrix of size N by M . Here N represents the number of channels and M represents the Number of sample points in time domain of a trial. In the theory, there are two classes A and B , and a projection matrix W is gained after CSP.

Using the projection matrix W , the data from each trial X can be projected as

$$Z = WX \quad (3)$$

After CSP projection, total $2m$ rows are selected to represent each trial. Let z_p be defined as the variance of row p of Z . Then usually the p^{th} component of the feature vector for the trial is computed as the logarithm of the normalize variance as in (4):

$$f_p = \log \left[\frac{\text{var}(z_p)}{\sum_{p=1}^{2m} \text{var}(z_p)} \right] \quad (4)$$

The feature vector $f = (f_1, \dots, f_{2m})$ is then used for designing classifier for motor imagery tasks. In this paper, we set the parameter N as 31, which is the number of all channels and 9 ISC components.

Because this paper designs the four-class MI tasks classification, CSP needs to be expanded to meet the requirements. There are two kinds of commonly used expansion methods: one-to-one and one-to-other. In this paper, the one-to-other way was chosen to expand the CSP. Therefore, four projection matrices f^1, f^2, f^3, f^4 will be generated for each trial. Each projection matrix is concatenated to form a whole spatial feature vector F_2 .

$$F_2 = [f^1, f^2, f^3, f^4] \in R^{1 \times (4 \times 2m)} \quad (5)$$

2.4. Feature Fusion

There are two kinds of feature fusion method: serial feature fusion and parallel feature fusion. Among them, the serial feature fusion is to connect a variety of features after they were normalized, which is simple. This paper uses serial feature fusion strategy to obtain EEG features.

Assuming the EEG feature vector after processing and serial feature fusion is $F \in R^{1 \times (K+8m)}$, then

$$\left\{ \begin{array}{l} F = [F_1, F_2] \\ F_1 = \left[\frac{f_{11}}{\|f_{11}\|}, \frac{f_{12}}{\|f_{12}\|}, \dots, \frac{f_{1K}}{\|f_{1K}\|} \right] \in R^{1 \times K} \\ F_2 = \left[\frac{f^1}{\|f^1\|}, \frac{f^2}{\|f^2\|}, \frac{f^3}{\|f^3\|}, \frac{f^4}{\|f^4\|} \right] \in R^{1 \times 8m} \end{array} \right. \quad (6)$$

In the formula, F_1 is the time-frequency domain feature vector of the 9 ISC components extracted based on LCD and the original 22 channels of EEG signals. F_2 is the spatial feature vector extracted by CSP method from N signals. In this paper, N is 31.

2.5. Feature Ranking and Classification

MCFS [17] is to compute the score for each feature and select d features according to the scores. In this paper, we just use the MCFS to score and rank the features for the following classification stage.

LDA is frequently used as classifier for analysing EEG signal. However, the computation time and memory cost of LDA are very expensive. This paper adopts the spectral regression discriminant analysis (SRDA) classifier [18]. Specifically, SRDA only needs to solve a set of regularized least squares problems, and there is no eigenvector computation involved, which is a huge save of both time and memory.

3. Results and Discussion

3.1. Our Experiment Procedure

In order to enhance the signal-to-noise ratio (SNR), the data was band-pass filtered between 8 Hz and 30 Hz. According to the literature [16], the 2.5s-3.5s data was selected for the classification processing.

Then the C3, C4 and Cz channels were selected to be processed with the LCD. Each channel can be decomposed into 3 ISC components. After integrating the original 22 channels and the 9 ISC components, the following step is to extract time-frequency features F_1 and spatial features F_2 from the integrated data with CSP.

Finally, the two feature vectors F_1 and F_2 are fused as F . F is ranked by the MCFS method and then put into the SRDA classifier to train and classify.

3.2. Classification Performance

There were 9 subjects taking part in the experiment, and each subject had two data sets. Session T is the training set, while the session E is the evaluation set. Each of them has 288 trials.

To illustrate the effectiveness of the proposed method, the performance of our method is also compared with other competitor algorithms: discrete wavelet transform (DWT) features [12], combining power spectral density, AR parameters and Hjorth parameters [13], the CSP of 22 channels, and the CSP of 31channels as shown in Table 1. In this table, the parameter $2m$ is set as 8.

Using SRDA classifier, and keeping the network parameters unchanged, this paper compared the classification performance of the five algorithms as shown in Table 1. In this table, each data show the classification accuracy of the four-class MI tasks of the subjects, including left hand, right hand, both feet and tongue.

Table 1: Comparison of Classification Accuracy (%) of Competitor Algorithms of 9 Subjects where $2m=8$

subjects	1	2	3	4	5	6	7	8	9	Mean
DWT	53.8	31.6	58.3	38.9	31.6	42.0	40.3	57.3	57.6	45.71
PSD + AR + Hjorth	54.2	35.4	48.6	41.7	30.6	39.9	48.6	49.3	62.5	45.64
CSP of 22 channels	60.8	34.7	77.1	53.1	35.1	37.5	62.8	73.3	77.1	56.83
CSP of 31 channels	63.9	39.9	75.0	57.6	34.4	42.0	62.5	73.6	76.0	58.33
Our method	62.2	40.3	75.3	57.3	34.4	41.3	64.6	75.0	79.5	58.87

As we can see, only for subject 6, our method has no ideal result, comparing with the DWT and PSD + AR + Hjorth features. For most subjects, our methods obtained the best results under the same experimental conditions.

Comparing with the CSP of 22 channels, the CSP of 31 channels has improved the classification accuracy averagely. It shows the importance of 9 ISC components decomposed by LCD in classification. Further, since the performance of our method is better than the other two CSP methods which have no time-frequency features, it's obvious that the time-frequency features extracted are significant for the classification.

For subjects 1, 2, 7, 8 and 9, our method has the best effect among the five algorithms. And our method has improved the average accuracy to a certain degree, compared with the CSP of 22channels. The validity of the proposed algorithm can be verified through this.

To further illustrate the effectiveness of our method, the average classification accuracy of our method is also compared with other two CSP algorithms when the value of $2m$ is changing, as shown in Table 2. Each data in this table presents the average classification accuracy of 9 subjects under different value of $2m$. As we can see, for most values of $2m$, our methods obtained the best results under the same experimental conditions, comparing with other two algorithms. And our method averagely improved the performance of the CSP method, which shows the effectiveness of our proposed method.

Table 2: Average Classification Accuracy (%) of 9 Subjects under Different Values of $2m$

values of $2m$	2	4	6	8	10	12	14	16	18	20	22	Mean
CSP of 22 channels	55.5	58.7	57.5	56.8	56.7	57.6	57.5	57.1	57.1	56.8	56.9	57.1
CSP of 31 channels	55.9	58.6	57.7	58.3	58.5	59.0	59.0	58.4	57.9	58.0	57.6	58.1
Our method	56.9	59.2	58.4	58.9	59.2	58.7	58.5	58.1	58.6	58.4	58.5	58.5

Thus it can be seen, the method proposed in this study can effectively extract the time-frequency and spatial information in motor imagery EEG signal to improve the classification accuracy of four-class motor imagery tasks. The results discussed above suggest that the proposed method is effective.

4. Conclusions

This paper presents a feature extraction method for four-class motor imagery tasks based on LCD and CSP. This method used LCD and CSP to analyse EEG signals and obtain corresponding time-frequency and spatial features. The fourth BCI competition Data Sets 2a data was used to verify the proposed method. The results show that our method further improves the recognition accuracy of CSP and has important significance in the practical application of motor imagery in the field of rehabilitation.

5. Acknowledgements

This work was supported by the National Science Foundation (Grant No. 51475342), the Fundamental Research Funds for the Central Universities (Grant No. 2016-III-019) and the Fundamental Research Funds for the Central Universities (Grant No. 2017-IVA-051).

We would like to thank all reviewers for valuable comments which help improve the quality of paper.

6. References

- [1] Hamed M, Salleh S H, Noor A M. Electroencephalographic Motor Imagery Brain Connectivity Analysis for BCI: A Review.[J]. *Neural Computation*, 2016, 28(6):999-1041.
- [2] Floriana Pichiorri M D, Giovanni Morone M D, Manuela Petti M S, et al. Brain-computer interface boosts motor imagery practice during stroke recovery[J]. *Annals of Neurology*, 2015, 77(5):851–865.
- [3] Li M, Lin L, Jia S. Multi-class imagery EEG recognition based on adaptive subject-based feature extraction and SVM-BP classifier[C]// *International Conference on Mechatronics and Automation. IEEE*, 2011:1184-1189.
- [4] Önder AYDEMİR. Common spatial pattern-based feature extraction from the best time segment of BCI data[J]. *Turkish Journal of Electrical Engineering & Computer Sciences*, 2016, 24(5):3976-3986.
- [5] Hsu W Y. Motor imagery EEG discrimination using the correlation of wavelet features.[J]. *Clinical Eeg & Neuroscience Official Journal of the Eeg & Clinical Neuroscience Society*, 2014, 46(2):94-99.
- [6] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences*, 1998, 454(1971):903-995.
- [7] Smith J S. The local mean decomposition and its application to EEG perception data.[J]. *Journal of the Royal Society Interface*, 2005, 2(5):443-54.
- [8] Mark G. Frei. Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals[J]. *Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences*, 2007, 463(463):321-342.
- [9] Bai X, Wang X, Zheng S, et al. The Offline Feature Extraction of Four-class Motor Imagery EEG Based on ICA and Wavelet-CSP[C]// *Control Conference. IEEE*, 2014:7189-7194.
- [10] Gaur P, Pachori R B, Wang H, et al. Enhanced motor imagery classification in EEG-BCI using multivariate EMD based filtering and CSP features[C]// *International Brain-Computer Interface*. 2016.
- [11] Zheng J, Cheng J, Yang Y. A rolling bearing fault diagnosis approach based on LCD and fuzzy entropy[J]. *Mechanism & Machine Theory*, 2013, 70(6):441-453.
- [12] Pal M, Bhattacharyya S, Roy S, et al. A bacterial foraging optimization and learning automata based feature selection for motor imagery EEG classification[C]// *International Conference on Signal Processing and Communications*. 2015:1-5.
- [13] Rodríguezbermúdez G, Garcíaencina P J, Rocardora J. Efficient automatic selection and combination of EEG features in least squares classifiers for motor imagery brain-computer interfaces.[J]. *International Journal of Neural Systems*, 2013, 23(4):1350015.
- [14] Wu S L, Wu C W, Pal N R, et al. Common spatial pattern and linear discriminant analysis for motor imagery classification[C]// *IEEE Symposium on Computational Intelligence. IEEE*, 2013:146-151.
- [15] Brunner, C., Leeb, R., Müller-Putz, G. R., Schlögl, A., and Pfurtscheller, G. (2008). BCI Competition 2008 – Graz data set A. Available at: http://www.bbc.de/competition/iv/desc_2a.pdf.
- [16] Fang Y, Chen M, Zheng X, et al. Extending CSP to detect motor imagery in a four-class BCI[J]. 2012, 9(1):143-151.
- [17] Cai D, Zhang C, He X. Unsupervised feature selection for Multi-Cluster data[C]// *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, Dc, Usa, July*. 2010:333-342.
- [18] Cai D, He X, Han J. SRDA: An Efficient Algorithm for Large-Scale Discriminant Analysis[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2008, 20(1):1-12.
- [19] Hjorth B. EEG analysis based on time domain properties.[J]. *Electroencephalography & Clinical Neurophysiology*, 1970, 29(3):306-310.