CSP-ECOC Combination On BCI Application

Prapong PRECHAPRAPRANWONG[1] Kenji NAKAYAMA[2] Akihito HIRANO[3]
Kanazawa University, Graduate School of Natural Science and Technology
Kakuma-machi, Kanazawa, Ishikawa, 920-1192, Japan
Tel: +81-76-234-4896
Fax: +81-76-234-4960
Email: prapong@leo.ec.t.kanazawa-u.ac.jp[1]
nakayama@t.ec.kanazawa-u.ac.jp[2]
hirano@t.kanazawa-u.ac.jp[3]

ABSTRACT Brain-computer interface (BCI) aims to create a new communication channel between human and computer by using brain signals, such as brainwaves or electroencephalograms (EEGs). However there are disadvantages in EEG based BCI ,e.g., its noise-contamination, non-stationary, and low spatial resolution. To alleviate these problems, some filtering techniques are applied. In such techniques, the common spatial pattern (CSP) is one of outstanding spatial filtering techniques that mostly applied in motor imagery based BCI. The objective of CSP is to find the most discrimination between two data sets by maximizing the variance of one data set and concurrently minimizing the variance of another data set. In this study, we focus on applying CSP to the cognitive mental tasks, e.g., multiplication, virtual letter composing, object rotation, etc. Thus, the feature extraction method is changed from the standard way, which uses variance of CSP transformed signals as a measure of the energy in corresponding frequency bands, to a stationary approach by using the Fourier transformation. This experimental also deals with the multi-class problems hence the multi-class extensions of CSP are investigated. We purpose to combine CSP spatial filter with error-correcting output code framework (ECOC)which, boosts up the discrimination between two mental tasks and corrects some classification errors. The simulation results from 3 subjects confirm that CSP-ECOC combination method can increase the accuracy rate from 88% to 90% for Subject S1, increase the accuracy rate from 66% to 84%, and suppress the error rate from 26% to 6% for Subject S6 and finally, this method can provide the maximum accuracy rate at 96%(increase from 80%) for Subject S2.

1 Introduction

The electroencephalogram (EEG) based brain-computer interface (BCI) is mainly used in non-invasive BCI approaches. However EEG signals have some shortcomings in noise-contamination and low spatial resolution. The voltage potentials from sources can contribute within a small radius through scalp toward each electrode that makes the observation signals obscure and noisy [1]. Moreover, EEGs are also reported that they are inherent non-stationary caused by changes in the individual subject's brain across experimental sessions [2]. To remedy these such problems, some spatial filtering techniques are employed to reduce noise and get more localized signals. The examples of the spatial filtering techniques that are applied in BCI framework, such as bipolar filtering, common average reference method, Laplace filtering, and finally, the statistical linear transformation based spatial filtering that linearly transform raw EEGs to new feature spaces [1].

In this study, we focus on the CSP based spatial filtering. We concentrate in the concept of CSP that aims to find the most discrimination between two data sets by optimizing the ratio between within-class scatter and between-class scatter of those two data sets [6].

CSP have been successfully applied to classify motor imagery based EEG [1]. However, this study investigates the CSP method on the cognitive mental tasks (e.g., mathematical multiplication, letter composing, 3D object rotation and visual number counting), which have different manners.

In previous researches, Nakayama et al. [8], [9] used the Fourier transform based features and some optimized preprocessing with the multilayer neural network (MLNN) and successfully to classify 5-class mental tasks at accuracy 78-88%. In this study, we follow this line by adding CSP spatial filtering to enhance discriminate between the mental tasks and provide more classification accuracy. Furthermore, refer to some discussions in [2], the significant information of mental tasks reveals in low frequency band and dimminshes in high frequency band. In order to extract more information in low frequency band and reduce noise in high frequency band, we also propose a modified method to the sampling reduction procedure.

This work engages with 5-class mental task problems, but originally CSP is designed for a binary-class problem. To deal with a multi-class problem, Dornhege et al.[3] have proposed many multi-class extension approaches for extend two-class CSP to multi-class application. Those methods are called, CSP within multiple-binary classification (CSP-IN), binary combination with one versus the rest strategy based CSP (CSP-OVR) and

CSP with simultaneous diagonalization method (CSP-SIM). We investigate those mentioned approaches based on MLNN and for the CSP-IN approach, we propose the ensemble of binary classifiers combine with the error-correcting output code (ECOC) framework. ECOC is a general framework to solve multi-class problems by reducing the multi-class problems to several binary class problems with error-correcting property. We combine this property with the CSP that theoretically make the classes most discriminant. We purpose to use CSP-ECOC combination to boost up the accuracy rates for EEG classification.

2 Common spatial pattern

CSP algorithm [6] aims to find spatial filters that project the original signals to the most difference in the temporal variance of signals between two sets of signals i.e., maximize the variance of a data set, simultaneously, minimize the variance of data set of another data set. In short, the two data sets will be transformed into the most differentiated direction in the term of variance of these data sets.

2.1 CSP algorithm

1. Given 2-class mental tasks of EEG signals X_1 for class**1** and X_2 for class**2** with dimension of $[N \times K]$, where N is the number of signal's channels and K is the number of samples in time domain.

2. Compute the normalized auto-covariance matrices S_i for each class
$$S_i = \frac{X_i X_i^T}{trace(X_i X_i^T)}, i \in \{1, 2\}$$

3. First, the whitening transformation should be preformed by compute the sum of both auto-covariance matrices
$$S_{sum} = S_1 + S_2$$

4. Then, decompose the eigenvector and the eigenvalue of the matrix S_{sum} as:
$$S_{sum} = U \Lambda U^T$$
where U and Λ are the eigenvector matrix and eigenvalue matrix of S_{sum}
whitening transformation matrix can be received from:
$$W = \Lambda^{-\frac{1}{2}} U^T$$

5. Apply the whitening transformation to both auto-covariance matrices then we got the whitening transformed covariance as:
$$\hat{S}_1 = W S_1 W^T \text{and } \hat{S}_2 = W S_2 W^T$$
Next, CSP spatial filters are computed from these whitening transformed covariance matrices

6. Refer to the concept of CSP, these whitening transformed covariance matrices (i.e.,$\hat{S}_1$ and $\hat{S}_2$) should share a common eigenvector matrix and corresponding eigenvalues of the sum of $\hat{S}_1$ and $\hat{S}_2$ should be one. Therefore, $\hat{S}_1$ and $\hat{S}_2$ can be decomposed as:
$$V^T \hat{S}_1 V = D$$
$$V^T (\hat{S}_1 + \hat{S}_2) V = I \text{ and } V^T \hat{S}_2 V = I - D$$
where V is a common eigenvector matrix of $\hat{S}_1$ and $\hat{S}_2$, D is a eigenvalue matrix of $\hat{S}_1$,and I is identity matrix

7. Then, we obtain the spatial filters of CSP transformation matrix as:
$$\hat{V} = V^T W$$

8. Practically, we choose only few most important eigenvectors from by sorting the eigenvalues in D in descending order and choose $2m$ eigenvectors ($2m < N$) that corresponding to the m largest and the m smallest eigenvalues. Then we obtain:
$$\hat{V}_m = [\hat{v}_1, ..., \hat{v}_m, \hat{v}_{N-m+1}, ..., \hat{v}_N]$$
where $\hat{v}_i$ is a eigenvector the corresponding to a eigenvalue in D

9. Finally, the projected signals are defined as:
$$Z_i = \hat{V}_m X_i, \; i \in \{1, 2\}$$

2.2 Extension to multi-class CSP

There are three approaches of multi-class extensions, discussed in [3].

1) CSP-IN: separate the problem to several sub-binary problems and perform CSP spatial filters for those pairs of problems

2) CSP-OVR: use several binary CSPs on one multi-classifier

3) CSP-SIM: find CSP by using joint approximate diagonalization (JAD) method and apply on one multi-classifier

CSP-IN and CSP-OVR approaches are similar in the concept of extension by adding more binary CSP processes to form a multiple classification. However, CSP-IN uses several binary classifiers to deal with each binary problem, but CSP-OVR solves these binary problems with one multi-class classifier.

CSP-SIM derives from the concept of 2-class CSP algorithm that CSP algorithm will find a simultaneous diagonalization of both covariance matrices whose eigenvalues sum to one. Thus it is possible to extend to many classes if we can approximate a simultaneous diagonalization for many classes problem. However, there is no general strategy to choose the appropriate CPS patterns for multi-class CSP. (e.g., 2-class problem uses the highest or the lowest eigenvalue). Dornhege et al. [3] have also proposed the heuristic way to solve this problem by using some score strategy. Given $\mathbf{D}$ is an approximate simultaneous diagonal matrix, computed by joint approximate diagonalization (JAD). $\mathbf{D}$ is in a form of concatenated eigenvalue matrices i.e., $\mathbf{D} = [D_1, D_2, ..., D_n]$, n is the number of classes then choose the appropriate eigenvectors for each class i corresponding to the highest score of eigenvalue in each sub-matrix D_i by following criteria: $score(\lambda_{ij}) = max(\lambda_{ij}, 1/(1+(N-1)^2\lambda_{ij}/(1/\lambda_{ij})))$. Note that if one eigenvector is selected more than once, replace it by the eigenvector with the next highest score.

2.3 Binary CSP with error-correcting output code

CSP-IN is the first algorithm of CSP's multi-class extensions.The idea of CSP-IN is separating the multi-class problem to sub-binary problem and simply doing the CSP within binary classification with one versus the rest or pairwise strategies [3]. We adopt this idea and implement on the error-correcting output code framework henceforth, called CSP-ECOC. We aim that combination between error-correcting properties of ECOC and discriminative feature extraction of CSP could provide some enhancements in classification.

3 Error-correcting output codes

Error-correcting output codes (ECOC) are approved as a general framework to combine binary problems to address the multi-class problem [11].

In ECOC framework consists of two steps: a coding step, where a codeword is assigned to each class, assume if N-class problem is classified by binary classifiers, we need n binary classifiers to form n different binary-discriminated partitions. To supervise those classifiers, a set of binary target with length n is assigned for each class, called a codeword. Arranging the codewords as rows of matrix, finally, we can define a codeword matrix M ,where $M \in \{0,1\}^{N\times n}$.

Another step is a decoding step, where a testing output vector searches for the most similar codeword in the codeword table. The performance of ECOC mostly relies on the codeword table that applied to the system. The regulations of designing codewords have been discussed in many researches. We can categorize the method of generating codewords into three types:
1) generating from the algebraic coding theory methods
2) generating by randomization
3) generating unique codewords for a particular data set

In this study, we use the generalized algebraic coding theory, i.e., exhaustive ECOC (E-ECOC). For decoding, we use the L1-norm distance. We prefer to use L1-norm distance instead of the hamming distance because it is more flexible to adjust the rejection threshold regarding for unintended EEG.

3.1 Coding strategy: Exhaustive ECOC

Dietterich and Bakiri [11] have proposed a code and a procedure for generating a well balance hamming distance between rows and include all possible non-trivial and non redundant $2^{(N-1)}-1$ length codes for N-class problem, called exhausitive ECOC (E-ECOC). This code is recommended to use for $3 \le N \le 7$. The procedure for generating E-ECOC is as follows: Assign the first row is all ones. 2^{nd}row consists of $2^{(N-2)}$ zeros, followed by $2^{(N-2)}-1$ ones. 3^{rd}row consists of $2^{(N-3)}$ zeros, followed by $2^{(N-3)}$ ones, followed by $2^{(N-3)}$ zeros, followed by $2^{(N-3)}-1$ ones. i^{th}row consists of alternating $2^{(N-i)}$ zeros and ones. Finally, the last row contains $0,1,0,1,0,1,\ldots,0$.

For $8 \le N \le 11$, Dietterich and Bakiri have suggested selecting some good subset of columns from the exhaustive code by optimization algorithm. For $N > 11$, the random code generation with hill-climbing procedure is recommended.

The exhaustive ECOCs for 5 classes, obtained from the generation procedure which has been detailed in [11] are shown in Table 1 below:

Table 1: Codeword table of E-ECOC for 5-class problem

	Classifier														
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}
ω_1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ω_2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
ω_3	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
ω_4	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
ω_5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Performance of correcting codes

The efficiency of ECOC is determined by the capability of error-correcting. It is depended on distance between codewords that can be calculate by $\lfloor(d-1)/2)\rfloor$ bits, which d is the distance between row's codewords in the codeword matrix. For example, in Table 1, distance d between each row of E-ECOC is 8 so, the capability of error correcting $\lfloor(d-1)/2)\rfloor$ is 3 bits. Because the distance d corresponds to the code's performance, so it is possible to increase the performance of coding by expanded the length of codeword. In spite of that does not mean, we should generate the number of bit in codewords as long as we can, in order to enlarge the distance d. Inversely, the longer codeword means the more classifiers and more computation. The issues about optimization of codeword length are discussed in recent ECOC researches, but they are not in our study's scope.

3.2 Decoding strategy: L1-norm distance

In decoding step, the output vectors from the base classifiers are compared to the patterns in codeword table for searching the most similar pattern. To find the similarity, the most easily techniques like L1-norm based distance is applied. Given a testing output from the b classifiers is combined to a vector: $Y = [y_1, y_2, \ldots, y_b]^T$,where y_j is the output of j^{th} base classifier. The L1-norm distance is defined by

$$d_j = \sum_{j=1}^{b} |C_{ij} - y_i| \tag{1}$$

where C_{ij} is a target value at i^{th} row, j^{th} column in codeword matrix.

4 Methodology

4.1 Data acquisition

In this study, we use the brainwave data sets that are available from the web site of Colorado state university [10]. To perform the experimental, the data sets of subject No.1, 2 and 6 are chosen, henceforth, called S1, S2 and S6, respectively. The S1's data set was used in conventional method [8], [9]. The conventional method also worked well on S2's data set but did not work well on S6's ,hence they are chosen for comparison. Each data set was consisted of 7 channels of signals from 7 electrodes ,which were placed at 6 positions: C3, C4, P3, P4, O1, and O2 of the International 10-20 system and 1 EOG. The signals were recorded at sampling rate of 250Hz for 10 seconds (total 2500 samples per each channel). These 5 mental tasks are performed in data sets:

Baseline: do nothing, but relaxed
Multiplication: do non-trivial multiplication
Letter composition: mentally compose the letter
Rotation: rotate a complex 3D object in mind
Counting: write the numbers one by one in mind

4.2 Preprocessing and feature extraction

The EEG data sets in our study are 5-class mental tasks. Therefore, the multi-class extension techniques for CSP are required. We perform 3 methods of multi-class extensions, i.e., CSP-OVR, CSP-SIM and CSP-ECOC.

CSP based spatial filtering

CSP is a supervised spatial analysis thus, it needs some initialization. First, the spatial filters are determined from the class-separated, trial-concatenated EEGs of training set. We use 6 channels of EEGs, i.e., C3, C4, P3, P4, O1, and O2 henceforth, called Ch1, Ch2, Ch3, Ch4, Ch5, Ch6 and EOG as Ch7 for short. Note that Ch7 (EOG) is not included for computing the CSP spatial filters, but it is used for detecting eye's artifacts in later. After CSP spatial filters are received, commonly, only few effective spatial filters are enough for discrimination. For CSP-OVR method and CSP-ECOC method, the number of spatial filters is varied from 1 to 4 and compared for finding the best accuracy rate. For CSP-SIM method, it has a difference scheme because it is based on the approximated joint diagonalization. In this method, we have available CSP's spatial filters equal to the number of original channels. In our case, we use 6-channel EEGs for 5-class mental tasks, thus, we have only 1 available spatial filter per class. Finally, the CSP transformation matrix is formed by those selected spatial filters.

Fourier transformation

Most of the researches in BCI, applied CSP to event-related EEG or motor imaginary based EEG that power of bandpass filtered EEGs are estimated by variances of interval signals and the activities are detected by the changing of variances. CSP analysis can be directly applied to extract those variance based features. However, this study deals with the mental tasks that rely on stationary of EEG signals in duration. For that reason, we attempt to use CSP spatial filtering without considering the time-domain information. Thus, Fourier transform based features that neglect time-domain information are considered. The Discrete Fourier Transformation (DFT) of the EEG signal $x(n)$ is given by

$$X(k) = \sum_{n=0}^{N-1} x(n) exp(-j\frac{2\pi}{N}kn), k = 0, 1, 2, \ldots, N-1 \quad (2)$$

where N is the number of EEG samples

Then the spectral characteristics at frequency k of EEG are obtained by coefficients $X(k)$.

Sample reduction method

After perform the Fourier transform on brainwaves, in order to reduce contaminated noises in brainwaves and also reduce computation in classification, the sizes of input are reduced by averaging the absolute values of Fourier coefficients in interval [8], [9]. Note that those values contain the spectral information of the corresponding frequencies.

In our work, Fourier transformed features of all frequency bands (0.1-100 Hz) are employed and reduced sampling by averaging the values in interval. However, the significant information of mental tasks explicitly reveals in the low frequency region and diminishes in high frequency region. Thus, sampling reduction with uniform resolution [8] may inattentively miss some information in low frequency band and accent some noises in high frequency band to input patterns.

To alleviate those problems, the non-uniform resolution sampling reduction is proposed. In this method, the intervals of averaging are not equal, but they are increased by the factor of 2^k , where $k = 2, 3, \ldots, K+1$, K is the number of reduced sample.Note that the samples on only the right side of Fourier transformed signal are used because they are symmetrical to the samples on the left side.

Non-linear normalization

The information of mental tasks may widely distribute, not only in the peak's frequency band. Important information for classification may be included in small non-prominent frequency band. Moreover, naturally in the neural networks, large inputs play an important role. To avoid the neural network's biased learning, this non-linear normalization is applied to the input data [8].

$$f(x) = \frac{\log(x - x_{min} + 1)}{\log(x_{max} - x_{min} + 1)} \quad (3)$$

Channel-concatenated patterns

To create input pattern for MLNNs, Fourier transform based features from each channel are concatenated

Table 2: Examples of number of input dimension which use in experimental

Method	#SF (channel)	#Class (pair)	#Sample (samples)	#Input (samples)
CSP-OVR	2	5	5	55
CSP-SIM	1	5	5	30
CSP-ECOC	4	1	5	25

by order to create the patterns and applied to classification process. For examples, CSP-OVR method, which is based on one over the rest strategies, required 5 pairs of problems for 5-class problem. Assume that we required 2 spatial filters for each pair ,thus 5 sets of 2 channel-signals are acquired. if each signal reduces sampling to 5 samples then we will get (5 pairs × 2 channels) × 5 samples = 50 samples. To detect some eye's artifact, one more channel of EOG signal is included, so totally 55 samples per one input pattern are acquired.

The examples of the input number of patterns are shown in table 2

4.3 Classification

We focus on backpropagation based MLNN for classification. MLNNs are performed both in multi-class classification approaches (for CSP-OVR and CSP-SIM) and binary classification approaches (for CSP-ECOC).

The parameters of neural network are setting as shown in Table 3 and Table 4.

Table 3: MLNN's setting for CSP-OVR and CPS-SIM

Parameter	
Input node	Refer to Table 2
Output node	5
Hidden node	20
Iteration	100000
Learning rate	0.01
Activation function	Tanh - Logistic sigmoid
Initial weight Range	-0.1 – 0.1
Threshold of rejection	0.6
Generalization method: Random noise	-0.1 – 0.1

4.4 Validation and evaluation

This experimental are subject-specific classification. Data of each subject are separately applied to classifiers. The experiments are evaluated by using 5-fold cross validation. EEG data of S1 and S6 (who preformed completely 2 sessions, 10 trials of EEG recordings) are separated to 8 trials for training and 2 trials for testing. Similarly, EEG data of S2 (who performed 1 session of EEG recordings) are separated to 4 trials for training and 1 for testing. Each trial contains 5 mental tasks of EEGs. So we totally have 50 data for S1, S6 and 25 data for S2.

Table 4: MLNN's setting for CSP-ECOC

Parameter	
Input node	Refer to Table 2
Output node	1
Hidden node	10
Iteration	80000
Learning rate	0.01
Activation function	Tanh - Logistic sigmoid
Initial weight range	-0.1 – 0.1
Threshold of rejection	0.32
Generalization method: Random noise	-0.1 – 0.1

To evaluate the classification performance, a correct classification rate (P_c), an error classification rate (P_e) and rate of correct and error classification (R_c) are used.

$$P_c = \frac{N_c}{N_t} \times 100\% \tag{4}$$

$$P_e = \frac{N_e}{N_t} \times 100\% \tag{5}$$

$$R_c = \frac{N_c}{N_c + N_e} \tag{6}$$

$$N_t = N_c + N_e + N_r \tag{7}$$

where N_c, N_e and N_r is the number of correct classifications, number of error classifications and number of rejections, respectively.

Finally, the results of simulation are averaged by 5 times experimental.

5 Experimental Results and Discussion

Experimental results

The results of classification are demonstrated in Table 5. S1's EEG data set, which has worked well with the conventional method (ORG) [18]-[19], also worked well with the overall CSP method by accuracy rates are 84%-90% ,although it seems that the preformance are slightly degraded.

The CSP filtering works well on S2's EEG data set, the correct classification rates are obviously increased to 96% by CSP-ECOC method, increased to 92% and 88% by CSP-SIM method and CSP-OVR method, respectively. Moreover, the error classification rates can be decreased to 0% by CSP-OVR and CSP-SIM method.

For S6's data set, the results also have indicated that the correct classification rates are obviously increased from 66% to 84% and the error classification rates are also significantly decreased from 26% to 6% by CSP-ECOC method. It could be seen that CSP-spatial filtering provides some trends of improvement on classification performance.

The reason of the improvement may cause from some benefits of CSP spatial filtering that make some increase of discriminated information on input patterns that expressed in Figure 2 for CSP-OVR and Figure 3 for CSP-SIM method.

Table 5: Experimental Results for S1

S1	P_c	P_e	R_c	#Classifier
ORG	88	4	0.96	1
CSP-OVR	84	6	0.93	1
CSP-SIM	84	8.6	0.90	1
CSP-ECOC	90	6	0.94	15

Table 6: Experimental Results for S2

S2	P_c	P_e	R_c	#Classifier
ORG	80	4	0.95	1
CSP-OVR	88	0	1.0	1
CSP-SIM	92	0	1.0	1
CSP-ECOC	96	4	0.96	15

Table 7: Experimental Results for S6

S6	P_c	P_e	R_c	#Classifier
ORG	66	26	0.72	1
CSP-OVR	82	8	0.91	1
CSP-SIM	76	14	0.84	1
CSP-ECOC	84	6	0.93	15

6 Conclusion

We have experimented on CSP spatial filtering and the multi-class extensions of CSP apply to EEG data sets which are available from the Colorado State University website. These mental tasks have different properties from the motor imagery mental tasks. Therefore, we change the way of standard CSP's feature extraction to more stationary method by using Fourier transform based features. We also developed the non-uniform resolution technique for sampling reduction in pre-processing process to suppress more noises in high frequency region. Finally, with the principle of CSP, which aims to discriminate the classes, combine with ECOC framework, we can boost up the correct classification rates (P_c) from 88% to 90% for Subject S1, 80% to 96% for Subject S2 and 66% to 84% for Subject S6. Furthermore, the error classification rates (P_e) can be significantly suppressed from 26% to 6% for Subject S6.

References

[1] G. Dornhege, J. d. r. Millán, T. Hinterberger, D. J. McFanland and K. R. Müller,"Toward Brain-Computer Interfacing", The MIT Press, 2007

[2] Z. A. Keirn and J.I. Aunon,"A new mode of communication between man and his surroundings" *IEEE Transactions on Biomedical Engineering*, Vol. 37, No. 12, pp.1209-1214, Dec., 1990

[3] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms", *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 6, pp. 993-1002, June, 2004

[4] M. Grosse-Wentrup and M. Buss, "Multi-class Common Spatial Patterns and Information Theoretic Feature Extraction" *IEEE Transaction on Biomedical Engineering*, Vol. 55, No. 8, pp. 1991-2000, Aug., 2008

[5] T. Yan, T. Jingtian, and G. Andong, "Multi-Class EEG Classification for Brain Computer Interfaces based on CSP" *International Conference on BioMedical Engineering and Informatics*, Vol. 2, pp. 469-472, May, 2008

[6] K. Fukunaga, "Introduction to statistical Pattern Recognition, 2nd ed.", Academic Press, 1990

[7] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification, 2nd ed.", Wiley-Interscience, 2000

[8] K. Nakayama, K. Inagaki, "A brain computer interface based on neural network with efficient pre-processing" *Proc.IEEE, ISPACS*, pp. 673-676, Dec., 2006.

[9] K. Nakayama, Y. Kaneda, and A. Hirano, "A brain computer interface based on FFT and multilayer neural network-Feature extraction and generalization" *Proc.IEEE, ISPACS*, pp. 826-829, Nov.28-Dec.1, 2007.

[10] http://www.cs.colostate.edu/eeg/

[11] T.G. Dietterich and G. Bakiri, "Solving Multi-class Learning Problems via Error-correcting Output Codes" *Journal of Artificial Intelligence Research*, Vol. 2, pp. 263-286, 1995

[12] A. Zhang, Z. Wu, C. Li and K. Fang, "On Hadamard-Type Output Coding in Multiclass Learning" *Lecture Notes in Computer Science* Vol. 2690 pp. 397-404, 2003

[13] F. J. MacWillians, and N. J. A. Sloane, "The Theory of Error-Correcting codes", Elsevier Science Publishers, 1977