<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Feature selection and channel optimization for biometric identification based on visual evoked potentials</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Bai, Y; Zhang, Z; Ming, D</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2014</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/204089">http://hdl.handle.net/10722/204089</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Proceedings of the International Conference on Digital Signal Processing. Copyright © IEEE.</td>
</tr>
</tbody>
</table>
Feature selection and channel optimization for biometric identification based on visual evoked potentials

Yanru Bai1,2, Zhiguo Zhang2, Dong Ming1,*

1. Department of Biomedical Engineering, College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, China
2. Department of Electrical and Electronic Engineering, the University of Hong Kong, China
* Corresponding author (Dong Ming, e-mail: richardming@tj.edu.cn)

Abstract—In recent years, biometric identification has received general concerns around the world, and become a frontal and hot topic in the information age. Among the internal biometric traits, electroencephalogram (EEG) signals have emerged as a prominent characteristic due to the high security, uniqueness and impossibility to steal or mimic. In this paper, individual difference of visual evoked potentials (VEPs) with cognition task were investigated, in addition, a feature selection and channel optimization strategy was developed for the VEPs based biometric identification system, where three different methods, including genetic algorithm (GA), Fisher discriminant ratio (FDR), and recursive feature elimination (RFE) were employed. In our experiments with 20 healthy subjects, the classification accuracy of support vector machine (SVM) reached up to 97.25% with AR model parameters, compared to 96.25% before optimization, and 32 channels of most discriminative were eventually selected from 64 channels with best performance. Results in this study revealed the feasibility of VEPs based EEG to be used for biometric identification. The proposed optimization algorithm was shown to have the ability to effectively improve the identification accuracy as well as simplifying the system. Further investigate may provide a novel idea for the effect and potential of EEG for biometric identification.

Keywords—biometric; visual evoked potentials (VEPs); features selection; channels optimization

I. INTRODUCTION

In recent years, biometric identification which refer to an automatic recognition of individuals based on features derived from their physiological or behavioral characteristic, has received general concerns all over the world. Methods based on the use of electroencephalogram (EEG) as a biometric are relatively novel compared to the other established biometric tools, such as fingerprint, iris, and face. EEG has been proposed as an alternative biometric due to its advantages of uniqueness, high confidentiality, impossibility to mimic, and robustness [1]. It has been reported in previous studies that the brain activity is determined by individual’s unique pattern of neural pathways [2], which implies the potential of EEG for biometric identification.

There are mainly two types of experimental paradigms to generate specific EEG signals for biometrics in the existing literatures [3]: one is resting state. The earliest research on biometrics using resting EEG was carried out by M. Poulos in 1999, who proposed to use the parameters of the AR model as features for classification using a Learning Vector Quantization network (LVQ), where the classification performance of 72%-84% was obtained based on the experiments involving four subjects with eyes closed [4]. The other paradigm is related to evoked potentials with stimulus or mental tasks. It was indicated in previous work that the levels of thought process across subjects were different even for similar mental activity [5]. In our work, individual difference of visual evoked potentials (VEPs) with cognition task were investigated. The underlying hypothesis underpinning this approach is that the perception of VEPs related to recognition and memory, which is known to be distinct among humans and therefore a candidate for biometrics [6, 7].

Feature extraction and classification algorithms have always been the hot issues in the area of EEG based biometrics. In this work, individual classification were evaluated via SVM, and features were extracted by means of four algorithms. In addition, another topic worthy of attention refers to the selection of the most promising features for individual identification, and how to simplify channels to make the system better fit practical applications. For this purpose, three technologies including genetic algorithm (GA), Fisher discriminant ratio (FDR), and recursive feature elimination (RFE) were employed and compared with each other. Results indicated that the optimal features gave improved performance with lower design complexity, which meet our current requirements well.

II. METHODS

Flowchart of signal processing and technical processes is demonstrated in Fig.1, which includes five main parts. More details will be described below.

Figure 1. Process Flowchart

This paper was supported by National Natural Science Foundation of China (No. 81222021, 61172008, 81171426, 81127003), National Key Technology R&D Program of the Ministry of Science and Technology of China (No.2012BAD4B02) and Program for New Century Excellent Talents in University of the Ministry of Education of China (No. NCET-10-0618).
A. Experimental Paradigms and EEG Recording

Twenty healthy right-handed subjects (mean age: 22.5±3) participated in our experiments of VEPs with cognition task. All participants had normal or corrected-to-normal visual acuity and no registered neurological or psychiatric disorders. Ingestion of caffeine and alcohol containing drinks were not allowed on experimental days. All details in the experiments were informed to the participant and the written informed consent was obtained prior to the experiment.

Pictures presented to participants as stimulus in visual evoked task were acquired from “Revised Snodgrass and Vanderwart object pictorial set” [8], which consists of two hundred and sixty color pictures. These pictures are common and meaningful, identified and named easily, such as a bike, a banana, etc. All pictures had been standardized on variables of central relevance to memory and cognitive processing. A total of forty trials were performed for each subject, and in each trial, a randomly selected picture was centrally displayed only once on the computer monitor located 1m away from the subject, and the stimulus duration was 0.3 second. Fig.2 illustrates examples of stimulus presentation procedure. Participants were asked to recognize and remember the picture being presented and EEG signals with length of one second after stimulus onset were recorded for further analysis [6].

Figure 2. Procedure of the stimulus presentation in the visual evoked task

During the experiments, participants were seated in a comfortable chair in a silent, temperature-controlled room. They were asked to focus their attention on the stimuli and relax their muscles. Brain signals were recorded using a 64-channel EEG cap with Ag-AgCl scalp electrodes placed according to the international 10-20 system. The recording was undertaken with a sampling rate of 1000 Hz and a bandpass filter (0.05–100 Hz). All channels were referenced to right mastoid and grounded central region, and then re-referenced to bilateral mastoid, and the impedances of all electrodes were kept lower than 10 kΩ. Electrooculographic (EOG) signals were simultaneously recorded using surface electrodes to monitor ocular movements and eye blinks.

B. Preprocessing

In order to effectively enhance the signal-to-noise ratio (SNR) and reliably extract the VEPs from background noise, signal preprocessing including re-reference, EOG interference removing with ICA [9], and down-sampling were applied.

C. Feature Extraction

Many of algorithms had been introduced to extract distinguishable information from EEG signals, for the purpose of individual identification. In this paper, four types of algorithms, including AR modeling (AR), power spectrum in time domain (TPS) and frequency domain (FPS), and phase-locking value (PLV) were adopted.

1. AR Modeling (AR)

In early studies of EEG based biometric identification, AR modeling was the most popular algorithm of feature extraction, which can be described by a linear difference equation in the time domain as follows:

\[ x(n) = \sum_{i=1}^{p} a_i x(n-i) + w(n) \]  

where a current sample of time series \( x(n) \) is a linear function of previous samples plus an independent and identically distributed white noise input \( w(n) \) [10], and \( P \) denotes the number of time points in the past to be used to model the current time point. In the present work, an order of 4 was adopted according to Akaike Information Criteria (AIC). The multivariate AR model coefficient matrices from all 64 channels were calculated as a final feature vector, that is, the feature vector has 64×4 dimensions.

2. Time-Domain Power Spectrum (TPS)

TPS analysis provides basic information of how the power distributes as a function of time, which is defined as (2):

\[ TPS(i) = \frac{1}{N} \sum_{n=1}^{N} [x(n)]^2 \]  

where \( x(n) \) is time series of EEG signals, \( N \) is the total number of time points and \( i = 1, \ldots, 64 \) denotes the EEG channels.

In order to minimize the difference among channels, a normalization method was employed:

\[ TPS_A(i) = \frac{TPS(i)}{\sum_{k=1}^{N} TPS(k)} \]  

The normalized TPS features from 64 channels were concatenated into a vector of 64×1 dimensions.

3. Frequency-Domain Power Spectrum (FPS)

The power spectrum in frequency domain reveals how the power distributes as a function of frequency, which is defined as:

\[ FPS(i) = \frac{1}{N} \sum_{n=1}^{N} [\tilde{x}(n)]^2 \]  

where \( \tilde{x}(n) \) is the EEG amplitude in a given frequency range, and \( N \) is the total number of frequency points in the frequency range. Likewise, the power value of each channel was normalized according to (5) and then be concatenated into one feature vector.

\[ FPS_A(i) = \frac{FPS(i)}{\sum_{k=1}^{N} FPS(k)} \]  

In our study, EEG signals were firstly transferred into frequency domain using Fast Fourier Transform (FFT), then the whole frequency range was equally divided into five segments, finally, FPS values in each sub-band of each channel were computed respectively. Thus the final FPS feature was a vector of 64×5 dimensions.

4. Phase-Locking Value (PLV)

In the present work, an order of 4 was adopted according to Akaike Information Criteria (AIC). The multivariate AR model coefficient matrices from all 64 channels were calculated as a final feature vector, that is, the feature vector has 64×4 dimensions.
In this work, PLV was used to investigate the phase synchronism between two channels of EEG signals, which is defined as (6):

$$PLV = \frac{1}{N} \sum_{i=1}^{N} e^{i\Delta \varphi}$$

where $\Delta \varphi = \varphi_1 - \varphi_2$ is the phase difference. The calculation of PLV was performed pairwise among 26 channels from contralateral hemisphere, thus a feature vector with 676 dimensions was obtained.

D. Classification

Features extracted by the above mentioned methods were classified by support vector machine (SVM) to obtain the statistical classification accuracy of all samples, where different individuals represent different categories. SVM is chosen instead of other candidate classifiers due to its quality of applying to small, nonlinear and high-dimensional samples.

SVM is a computational method based on the statistical learning theory, of which the central idea is to separate data from different classes by a hyper-plane with largest margin. In practice, original input space is mapped into a high-dimensional dot product space called feature space, and in the feature space the optimal hyper-plane is determined to maximize the generalization ability of the classifier. In this work, classifications were achieved on the platform LibSVM developed by Chin-Jen Lin [11]. In addition, 10-fold cross validation was applied to make the accuracy reliable.

E. Features selection and channels optimization

• Genetic Algorithm (GA)

GA is an adaptive probabilistic search algorithm for global optimization inspired by the laws of natural selection and genetics. GA follows natural evolutionary model, and starts with an initial population of individuals, which consist of a fixed length continuous or discrete strings analogous to the chromosomes in a DNA. Each individual represents a possible solution to a given optimization problem and over successive generations evolves toward a set of more optimal or fit individuals.

![Figure 3. Evolution process of GA](image)

Fig. 3 illustrates the evolution process of GA which is governed by three basic operations: selection, crossover and mutation. Each individual in the population was assigned a fitness value based on a fitness function according to how good a solution it was to the problem. In the present work, fitness was evaluated on the basis of the classification accuracy on SVM. The ‘selection’ process mimicked the survival of the fittest in nature, and chose those individuals from the current population that would be allowed to reproduce, with fitter individuals producing on average more offspring than less fit ones. It should be noted that in the process of features selection in this work, we took channel as a unit, which meant features from the same channel would be taken as a whole to reserve or remove together. The best individuals to be selected were those who obtained the highest classification accuracy in the generation. ‘Crossover’ was an analogy to mating between individuals and exchanged subparts of two chromosomes to form new individuals for the next generation. A portion of the new individuals was created through ‘mutation’, which randomly changed the allele values of some locations in the chromosome and helped maintain diversity within the population.

When a new generation of individual was produced through the three operators, repeat the above process until we reached the pre-set generation, or when the fitness was no longer rising, then the algorithm ended.

• Fisher Discriminant Ratio (FDR)

In this study, FDR was used to compute the discriminatory power of each feature, on which basis, the discriminating features (channels) would be selected from all features in the 64 channels. For the $k^{th}$ feature, calculation of FDR is defined as (7):

$$FDR_k = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\mu_{i}^{k} - \mu_{j}^{k}}{\sigma_{i}^{k} + \sigma_{j}^{k}}\right)^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\mu_{i}^{k} - \mu_{j}^{k}}{\sigma_{i}^{k} + \sigma_{j}^{k}}\right)^2}$$

where $N$ is the number of categories; $\mu_i^k$ and $\sigma_i^k$ are respectively the mean and standard deviation of the $k^{th}$ feature for the $i^{th}$ class, and $i$ is varied from 1 to $N$ classes. The greater the FDR values, the stronger discrimination between the categories of the features, and vice versa.

Likewise, features from the same channel would be reserved or removed as a whole. FDR value of each channel was calculated and sorted in descending order. Then the combination of features from the first one channel, the first two channels…till the first $N$ channels were successively used as input for classification. Channels combination which attained the highest accuracy would eventually be taken as the optimal one.

• SVM based Recursive Feature Elimination (RFE-SVM)

The RFE-SVM approach, proposed by Guyon [12], is very popular for feature selection and subsequent classification, especially in the area of bioinformatics. The goodness of the features was determined by the absolute value of the corresponding weights used in the SVM, which were evaluated by the influence on the margin of a trained SVM. RFE method was achieved through a series of iterative process, at each iteration the SVM was re-trained, followed by removing the ‘bad’ features which minimized the margin variance. The features (or rather channels) remaining after a number of iterations were considered to be the most discriminating and useful ones for classification.

III. Results

A. Typical VEP waveforms and individual differences

VEP waveforms from four different subjects were showed in Fig.4, where we can see that the positive and negative peaks occur at relatively dynamic latencies though the VEPs appear similar in wave shape.
In order to present the effectiveness and feasibility of the proposed method, we collected visual evoked potentials from 20 healthy participants. To be more reliable, we performed a 10-fold cross validation for which 18 subjects were randomly selected for training and the remaining 2 subjects were used for test. Fig.5 displayed the comparison of classification results for the four feature extraction algorithms. Results disclosed that the best feature extraction method for individual classification in the current problem was AR Modeling with the highest accuracy of 96.25%, followed by TPS with 90.63%, and FPS with 88.75%, finally PLV with 88.63%.

C. Performance of feature selection and channels optimization

The objective of this part is to provide further perspective of EEG based biometrics on the use of VEPs by minimizing the number of required channels from all 64 channels. According to the preliminary results, only features extracted by using AR modeling were chosen for further optimizations due to its best performance. The upper panel and the lower panel of Fig.6 displayed the performance of optimizations in term of classification accuracy and number of channels respectively. For GA, the improvement in classification was from 96.25% to 97%, with channels reducing from 64 to 37; for RFE-SVM, classification accuracy was increased to 97.25%, and channels were reduced to 32; while for FDR, no improvement was achieved for both aspects.

In summary, optimization methods are helpful for removing redundant information and simplifying the system to some degree. Among the three proposed algorithms, GA and RFE-SVM are much more preferred for the current problem, since they can not only improve the classification accuracy but also minimize the number of required channels, while FDR is not an appropriate choice here in the present work.

IV. CONCLUSION

The main contributions of this paper are two-fold. First, a novel biometric system based on VEPs with multiple features were investigated. Experimental results showed that the highest classification accuracy of 96.25% across 20 subjects was obtained on SVM using AR model parameters as features, which revealed the feasibility of VEPs based EEG to be used for biometric identification.

The objective of second part was focus on exploring effective optimization strategy for feature selection and channel simplification. By doing this, we hoped to select the most promising features with less redundancy information, meanwhile reduce the number of required channels, so that it would requires lower computational time, design complexity and cost without degradation in the classification performance. Three methods including genetic algorithm, Fisher discriminant ratio, and recursive feature elimination were employed on AR model features. It was found from the results that except FDR, the other methods both can improve the performance of classification while reducing the number of channels. The highest classification rate reached to 97.25% with 32 optimal channels selected by RFE-SVM. The positive results obtained in this work advance our knowledge on EEG individual difference and provide a contribution to improve its practical design in the applications of biometrics.
REFERENCES


