Feature Extraction of Individual Differences for Identification Recognition Based on Resting EEG

Rui Xu, Dong Ming*, Yanru Bai, Jing Liu, Hongzhi Qi, Qiang Xu, Peng Zhou, Lixin Zhang, and Baikun Wan

Department of Biomedical Engineering,
College of Precision Instrument and Opto-electronics Engineering,
Tianjin University, Tianjin 300072, China
richardming@tju.edu.cn

Abstract. Biometric recognition based on individual difference was commonly used in many aspects in life. Compared with the traditional features used in person identification, EEG-based biometry is an emerging research topic with high security and uniqueness, and it may open new research applications in the future. However, little work has been done within this area. In this paper, four feature extraction techniques were employed to characterize the resting EEG signals: AR model, time-domain power spectrum, frequency-domain power spectrum and phase locking value. In our experiments using 20 healthy subjects, the classification accuracy by support vector machine reached 90.52% with AR model parameters, highest of the four kinds of features. The results show the potential applications of resting EEG signal in person identification.

Keywords: individual differences, person identification, resting EEG, AR model, support vector machine.

1 Introduction

In recent years, biometric recognition has received general concerns all over the world and played a key role in identification recognition applied to access control system, building gate control, digital multimedia access, transaction authentication or secure teleworking [1]. Each kind of biometric features needs to be evaluated in some aspects, such as universality, uniqueness, permanence and so on [2]. Common-used features including fingerprint, eye retinas and irises, and facial patterns, have achieved wide applications. But there exists limitations to the traditional biometric recognition methods to some extent. Therefore, it is necessary to come up with novel methods that could compensate these limitations.

Our attention is paid on the individual differences based on resting EEG signals which is a novel attempt in this field of biometric recognition. Genetically-specific features of EEG form the basis of a person identification method. Compared with other biometric features like fingerprint, this modality has several advantages: 1) it is

* Corresponding author.

P.L.P. Rau (Ed.): CCD/HCII 2013, Part I, LNCS 8023, pp. 503–509, 2013.

confidential, 2) it is difficult to mimic, and 3) it is almost impossible to steal [1]. It has been shown in previous studies that the brain-wave pattern of every individual is unique and that electroencephalogram (EEG) can be used for biometric identification [3]. In 1999, the group of M. Poulos et al first made an attempt to experimentally investigate the connection between a person's EEG and generally-specific information using AR Modeling and achieved correct classification scores at the range of 72% to 84%, which showed the potential of EEG used for person identification [4]. Then Palaniappan et al extracted biometric features based on the spectral power of EEG signals for classification by a fuzzy Neural Network [5]. More recently, Isao Nakanishi et al adopted the concavity and convexity of spectral variance in the alpha band of EEG as the features with low computational load and obtained Equal Error Rate of 11% [6].

Autoregressive (AR) methods have been used in a number of studies to model EEG data by representing the signal at each channel as a linear combination of the signal at previous time points [7-10]. This method provides a compact, computationally efficient representation of EEG signals, and AR model parameters are invariant to scaling changes in the data that can arise from inter-subject variations, such as scalp and skull thickness [10]. Time-Domain and Frequency-Domain Power Spectrum are two methods to estimate the power distribution among different time and frequencies. Unlike these methods mentioned above, Phase Locking Values (PLV) was developed to quantify in a statistical sense the phase synchronization of such systems from experimental data and to characterize their coupling [11-13]. It was first proposed to estimate the instantaneous phase relationship between two neuroelectric or biomagnetic signals and to apply it to intracortically recorded signals in humans [14]. Support vector machine was adopted to classify these feature parameters.

The main purpose of this study is to highlight the individual differences in resting EEG by using four kinds of feature extraction techniques and apply this feature into biometric recognition on the basis of support vector machine. The reminder of this paper is organized as follows: Section 2 briefly describes the methods of feature extraction and classification. Section 3 summarizes the results. Finally, Section 4 concludes.

2 Methods

2.1 Subjects

Twenty healthy right-handed subjects (mean age: 23.20±1.28) participated in this present study. None of the subjects reported any neurological or chronic diseases. The written informed consents were given by all the subjects prior to the experiments.

2.2 EEG Recordings

Subjects were comfortably seated in an electrical shield room during the experiment. They were asked to keep their body still and eyes closed for two minutes during the EEG recording.

The resting EEG was collected using a Neuroscan Synamps apparatus with an electrode cap of 64 electrodes (referenced to channel Cz). The recording was undertaken with a sampling rate of 1000 Hz and a bandpass filter (0.05–100 Hz). As Cz was set as the reference, the EEG at Cz, CPz, Pz and other central channels was very weak and it was difficult to extract features from these channels. Then offline, we changed the reference from Cz to left and right mastoid, precisely the mean value of EEG at mastoids.

2.3 Feature Extraction

Four kinds of approaches, including AR Modeling, Power Spectrum in Time Domain and Frequency Domain and Phase-Locking Values, were adopted to extract the individual difference information from resting EEG.

AR Modeling. The AR model is described by a linear difference equation in the time domain as

$$x(k)=\sum_{i=1}^{P}a(i)x(k-i)+e(k) \tag{1}$$

where a current sample of the time series $x(k)$ is a linear function of previous samples plus an independent and identically distributed (i.i.d) white noise input $e(k)$ [15], and P denotes the number of time points in the past that are used to model the current time point.

In our study, the model order was set to 4 according AIC. In detail, the first four parameters of AR Modeling of each channel of EEG were used as the features for identification, totally forming a feature vector with 64×4 dimensions per sample.

Time-Domain Power Spectrum. This spectrum feature is defined as

$$TPS(c)=\frac{1}{N}\sum_{n=1}^{N}[x_c(n)]^2 \tag{2}$$

where $x(n)$ is time series of EEG signals, N is the number of points and $c=1,\cdots,64$ denotes the EEG channel.

In order to minimize the difference between EEG from different channels, a standardization method was employed:

$$TPS_S(c)=\frac{TPS(c)}{\sum_{k=1}^{64}TPS(k)} \tag{3}$$

The TPS feature TPS_S of each subject was calculated to a vector of 64×1 dimensions.

Frequency-Domain Power Spectrum. The power spectrum in frequency domain reveals energy changes in different frequencies.

$$FPS(c) = \frac{1}{N} \sum_{n=1}^{N} [\tilde{x}(n)]^2 \tag{4}$$

where $\tilde{x}(n)$ was the EEG amplitude in a given frequency range, and N was the total number of frequency points in this frequency range. The standardization method was also used as:

$$FPS_S(c) = \frac{FPS(c)}{\sum_{k=1}^{64} FPS(k)} \tag{5}$$

In our study, EEG signals were transferred into frequency domain using Fast Fourier Transform (FFT). The whole frequency range was divided into five parts. The FPS calculation was taken within each of these five frequency ranges, and thus the FPS feature of each subject was a vector of 64×5 dimensions.

Phase-Locking Values. In order to obtain the instantaneous phase, we adopted the Hilbert transformation to transfer the EEG $x(t)$ to a complex-valued signal.

$$\tilde{x}(t) = Hilbert(x(t)) \tag{6}$$

The instantaneous phase φ was obtained analytically:

$$\varphi(t) = \mathrm{Im}(\ln(x(t))) \tag{7}$$

The phase difference between two channels of EEG signals $x_1(t)$ and $x_2(t)$ was calculated as $\Delta\varphi = \varphi_1 - \varphi_2$. So the PLV could be obtained by

$$PLV = \left| \left\langle e^{i\Delta\varphi} \right\rangle \right| \tag{8}$$

where $\langle \cdot \rangle$ meant time average.

The computation of Phase-Locking Values was made between each of 676 pairs of EEG recorded from different channels, obtaining a feature vector with 676 dimensions per sample.

2.4 Classification

The features extracted by the above mentioned techniques were classified by support vector machine (SVM) to obtain the statistical classification accuracy of all samples. Simultaneously, *k*-fold cross validation was applied during the classification to make the accuracy reliable.

Support vector machine developed by V. Vapnik, was a modern computational learning method based on statistical learning theory. The central idea was to separate data X into two classes by finding a weight vector $\omega \in R^d$ and an offset $b \in R$ of a hyper plane

$$\begin{aligned} &H : R^d \rightarrow \{-1,1\} \\ &x \mapsto sign(\omega \cdot x + b) \end{aligned} \tag{9}$$

with the largest possible margin [16]. x was the element of X with the dimension of d. The purpose of SVM was to separate one class from the other. In order to recognize twenty classes or subjects, we used 190 classifiers to separate every two of these twenty classes. Finally, 10-fold cross validation was adopted to calculate the classification accuracy.

3 Results

The dimensions of above-mentioned feature vectors were displayed in Table 1. These vectors with different dimensions composed the inputs of classifiers.

Table 1. The EEG features to be classified

Classes	Samples/Class	Feature Dimensions			
		AR Model	TPS	FPS	PLV
20	40	256	64	320	676

The results in this thesis showed that the best method of feature extraction for individual recognition was AR Modeling with the highest accuracy of 90.52%, followed by 82.63% with Time-domain Power Spectrum, 80.00% with Phase Locking Values and 67.37% with Frequency-domain Power Spectrum. The efficiency of AR Modeling was in accordance with what had been proposed in previous studies [17][18].

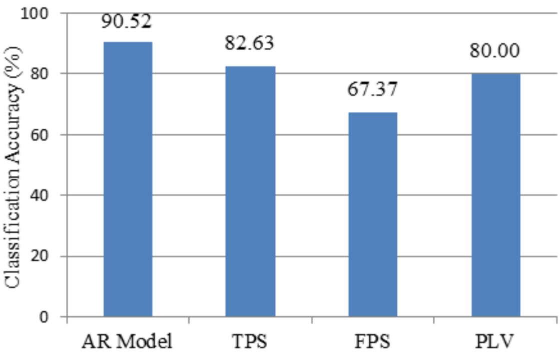

Fig. 1. The result of classification based on SVM for four kinds of features (AR model, TPS, FPS and PLV)

4 Conclusion

In this paper, a novel idea was provided for the individual differences analysis of resting EEG and for its practical design in the field of human-computer interaction especially for individual rehabilitation robots. In details, we investigated AR model, time-domain and frequency-domain power spectrum and phase locking value to extract the characteristic features of resting EEG among different subjects.

Our research proposed a feasible way for feature extraction in identification recognition, taking advantage of individual differences in EEG signals. Except for the frequency-domain power spectrum, the classification accuracies of these features were all above 80.00%, which verified the fact that EEG signals could be used as a kind of biometric feature. AR model parameters with the highest classification accuracy (90.52%) could represent individual differences in resting EEG well.

However, there may be more appropriate mental tasks for person identification than the resting EEG. In practice, different categories of EEG signals and channel optimization should be added in the following works, and EEG should be used with other biometric features to show complementary advantages in the future.

Acknowledgments. The authors would like to acknowledge all members of the Neural Engineering & Rehabilitation Lab of Tianjin University (TUNERL) for their valuable contributions to this research. This research was supported by National Natural Science Foundation of China (No. 30970875, 81171423, 61172008, 90920015), National Key Technology R&D Program of the Ministry of Science and Technology of China (No. 2012BAI34B02) and Program for New Century Excellent Talents in University of the Ministry of Education of China.

References

1. Marcel, S., del R. Millán, J.: Person Authentication Using Brainwaves (EEG) and Maximum A Posteriori Model Adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 743–748 (2007)
2. Clarke, R.: Human Identification in Information Systems: Management Challenges and Public Policy Issues. Information Technology & People 7, 6–37 (1994)
3. Marcel, S., del R. Millán, J.: Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 743–748 (2007)
4. Poulos, M., Rangoussi, M., Chrissikopoulos, V., Evangelou, A.: Person identification based on parametric processing of the EEG. In: The 6th IEEE International Conference on Electronics, Circuits and Systems, vol. 1, pp. 283–286 (1999)
5. Palaniappan, R., Ravi, K.V.R.: A new method to identify individuals using signals from the brain. In: The Joint Conference of the 4th International Conference on Information, Communications and Signal Processing, vol. 3, pp. 1442–1445 (2003)
6. Nakanishi, I., Baba, S., Miyamoto, C.: EEG based biometric authentication using new spectral features. In: International Symposium on Intelligent Signal Processing and Communication Systems, pp. 651–654 (2009)

7. Paranjape, R.B., Mahovsky, J., Benedicenti, L., Koles, Z.: The electroencephalogram as a biometric. In: Electrical and Computer Engineering (2001)
8. Poulos, M., Rangoussi, M., Alexandris, N., Evangelou, A.: Person identification from the EEG using nonlinear signal classification. Methods of Information in Medicine 41(1), 64–75 (2002)
9. Palaniappan, R.: Electroencephalogram signals from imagined activities: a novel biometric identifier for a small population. In: Belli, F., Radermacher, F.J. (eds.) IEA/AIE 1992. LNCS, vol. 604, pp. 604–611. Springer, Heidelberg (1992)
10. Lawhern, V., David Hairston, W., McDowell, K., Westerfield, M., Robbins, K.: Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. Journal of Neuroscience Methods 208, 181–189 (2012)
11. Rosenblum, M.G., Pikovsky, A.S., Kurths, J.: Phase synchronization of chaotic oscillators. Phys. Rev. Lett. 76, 1804–1807 (1996)
12. Rosenblum, M., Pikovsky, A., Kurths, J., Schafer, C., Tass, P.A.: Phase synchronization: from theory to data analysis. In: Moss, F., Gielen, S. (eds.) Handbook of Biological Physics, vol. 4, pp. 279–321
13. Sazonov, A.V., Ho, C.K., Bergmans, J.W.M., Arends, J.B.A.M., Griep, P.A.M., Verbitskiy, E.A., Cluitmans, P.J.M., Boon, P.A.J.M.: An investigation of the phase locking index for measuring of interdependency of cortical source signals recorded in the EEG. Biological Cybernetics 100, 129–146 (2009)
14. Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F.J.: Measuring phase synchrony in brain signals. Human Brain Mapping 8, 194–208 (1999)
15. Jain, S., Deshpande, G.: Parametric modeling of brain signals. In: Proceedings of Technology for Life: North Carolina Symposium on Biotechnology and Bioinformatics, pp. 85–91 (2004)
16. Lal, T.N., Schröder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., Schölkopf, B.: Support vector channel selection in BCI. IEEE Transations on Biomedical Engineering 51(6), 1003–1010 (2004)
17. Übeylï, E.D.: Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert System with Applications 37, 233–239 (2010)
18. Lawhern, V., David Hairston, W., McDowell, K., Westerfield, M., Robbins, K.: Detection and classification of subject-generated artifacts in EEG signals using autoregressive models. Journal of Neuroscience Methods 208, 181–189 (2012)