EEG-Based User Authentication Using Artifacts

Tien Pham, Wanli Ma, Dat Tran, Phuoc Nguyen, and Dinh Phung

Faculty of Education, Science, Technology and Mathematics,
University of Canberra, Australia
{tien.pham,wanli.ma,dat.tran,phuoc.nguyen,
dinh.phung}@canberra.edu.au

Abstract. Recently, electroencephalography (EEG) is considered as a new potential type of user authentication with many security advantages of being difficult to fake, impossible to observe or intercept, unique, and alive person recording require. The difficulty is that EEG signals are very weak and subject to the contamination from many artifact signals. However, for the applications in human health, true EEG signals, without the contamination, is highly desirable, but for the purposes of authentication, where stable and repeatable patterns from the source signals are critical, the origins of the signals are of less concern. In this paper, we propose an EEG-based authentication method, which is simple to implement and easy to use, by taking the advantage of EEG artifacts, generated by a number of purposely designed voluntary facial muscle movements. These tasks can be single or combined, depending on the level of security required. Our experiment showed that using EEG artifacts for user authentication in multilevel security systems is promising.

Keywords: EEG, authentication, security, biometrics, pattern recognition.

1 Introduction

Human electroencephalography (EEG) signals, which are a measurement of the generated electrical field when neurons are activated, were discovered in early 1900s, and they have been playing an important role in health and medical applications. Epileptic seizure detection is one of the most well-known applications. Another common usage of EEG signals in health is the study of sleep disorders. In additional, the relations between EEG signals and brain diseases have been investigated. Recording EEG signals is non-invasivewith a portable device, so EEG is also widely used in Brain Computer Interface (BCI) which can provide a link between the human subject and the computer without physical contact [16].

Recently, EEG emerges as a potential type of authentication with the advantages of being difficult to fake, impossible to observe or intercept, unique, un-intrusive, and alive person recording require [10] [16]. Many EEG modalities have been studied for person identification and verification such as motor imagery, mental tasks (e.g., mental multiplication), and responses to visual stimuli (i.e., Visual Evoked Potentials (VEPs)). In [10] [13] [17], the subjects were asked to imagine moving hand, finger,

J.G. de la Puerta et al. (eds.), *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14*, Advances in Intelligent Systems and Computing 299,
DOI: 10.1007/978-3-319-07995-0_34, © Springer International Publishing Switzerland 2014

foot or tongue while EEG data was recorded. In [10], nine subjects were recorded EEG data during to imagine generation of words beginning with the same random letter. In [14], the authors used five mental tasks including baseline, visual counting, geometric figure rotation, mental multiplication, and mental letter composing. In [1], a dataset was used in which the subjects were asked to look at black and white drawings of common objects when EEG signals were recording. However, these EEG modalities have their own disadvantages. Motor imagery and mental tasks are difficult to perform, and they require users have to train [1]. VEPs is slow and not universality since some users are visually impaired.

EEG signals are very weak and subject to the contamination from many artifact signals. Most of the current researches tried to separate the true EEG signals from the artifact interfaces. However, for the applications in human health, true EEG signals, without the contamination, are highly desirable, but for the purposes of authentication, where stable and repeatable patterns from the source signals are critical, the origins of the signals are of less concern. In addition, for an authentication system, it is desirable for the system to be nonintrusive, easy to implement and operate, and yet having different credentials for different levels of security.

This paper makes two contributions to EEG based biometrics: (i) proposing an EEG based authentication method, which is simple to implement and easy to use, by taking the advantage of EEG artifacts, generated by a number of purposely designed voluntary facial muscle movements; and (ii) introducing flexible EEG-based authentication policies for multilevel security systems by applying single artefact tasks when the system is of a lower security level or combined tasks if a high security level is required.

The rest of the paper is organized as follows. In Section 2, we study using EEG artifacts for authentication in multilevel security systems. Section 3 describes EEG features. Section 4 describes Support Vector Machine (SVM) modelling technique and hypothesis testing. Experiments and results are presented in Section 5. We conclude the paper with a discussion and our future work in Section 6.

2 Using EEG Artifacts for User Authentication in Multilevel Security Systems

From the point of view of human computer interaction (HCI), systems for a human being to use should be easy to use and natural to operate, with respect to human nature [5]. In [12], the authors stated that a good HCI system must try to adapt the intrinsic cognitive load, to reduce extraneous cognitive load, and to foster germane cognitive load of human users. Moreover, in [18], the author proposed the same ideas through perceptual user interface, which promises “natural, intuitive, adaptive, and unobtrusive” human-computer interaction. The “natural and intuitive” human-computer interaction is also suggested by Lenman et al [9].

Therefore, we propose an authentication system using EEG artifacts as illustrated in Figure 1.

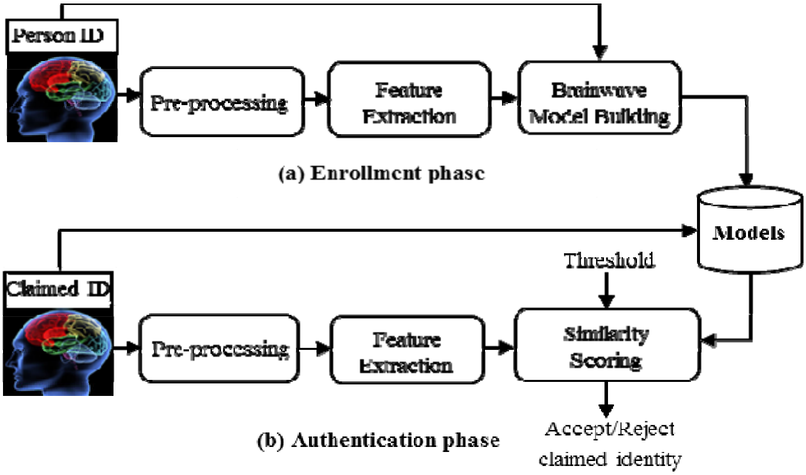

Fig. 1. EEG-based user authentication diagram

An EEG-based authentication system using artifacts has two phases: enrolment and verification. In the enrolment phase, a user is asked to do some facial muscle movement tasks, for example blinking left eye, blinking right eye, moving mouth to the left or the right, and EEG signals are recorded. The number of tasks can be flexible and depends on the security level of the system. After collecting the data, the EEG signals of each task corresponding to the user are pre-processed, extracted features, and then the features are used to train the model for this person, which is kept securely in a database.

In the verification phase, when a user wants to access the system, he or she has to provide EEG signals by repeating the tasks which he/she did in the enrolment phase. These input EEG data are processed in the same way as in the enrolment phase. The obtained features are then used to calculate a similarity score with the model of the individual who he or she claims to be. A threshold is also input to compare with the similarity score to accept or reject the individual.

The security systems can have a multiple security levels, depending on zones and resources with EEG based authentication because it can be adjusted by the number of matched tasks. If a system is of a lower security level, an individual may perform one task, and the system only requires that task is matched. If a system is of a high security level, a user has to perform some tasks and all those tasks in the sequence must be matched (AND case), so it helps to strength the security system.

3 EEG Features

3.1 Autoregressive (AR) Features

Autoregressive model can be used for a single-channel EEG signal. Each sample $s(n)$ in an AR model is considered to be linearly related, with respect to a number of its previous samples:

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + x(n) \tag{1}$$

where a_k, $k = 1, 2, \ldots, p$ are the linear parameters, n denotes the discrete sample time, and $x(n)$ is the noise input. The linear parameters of different EEG channel were taken as the features.

3.2 Power Spectral Density (PSD) Features

Power spectral density (PSD) of a signal is a positive real function of a frequency variable associated with a stationary stochastic process. The PSD is defined as the discrete time Fourier transform (DTFT) of the covariance sequence

$$\emptyset(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-i\omega k} \tag{2}$$

where the auto covariance sequence $r(k)$ is defined as

$$r(k) = E\{s(t) y^*(t-k)\} \tag{3}$$

and $s(t)$ is the discrete time signal $\{s(t); t = 0, \pm 1, \pm 2, \ldots\}$ assumed to be a sequence of random variables with zero mean.

In this paper, the Welch's method [19] using periodogram is used for estimating the power of a signal at different frequencies. Welch's method can reduce noise but also reduce the frequency resolution as compared to the standard Bartlett's method.

4 Modelling Technique

4.1 Support Vector Machine (SVM)

The training data set obtained during the enrollment phase, is labeled as $\{x_i, y_i\}, i = 1, \ldots, l$, $y_i \in \{-1, 1\}$, $x_i \in R^d$. Support vector machine (SVM) using C-Support Vector Classification (C-SVC) algorithm will find the optimal hyperplane $f(x)$ [3]:

$$f(x) = w^T \Phi(x) + b \tag{4}$$

to separate the training data by solving the following optimization problem:

$$min \ \ \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} \xi_i \tag{5}$$

subject to

$$y_i \left[w^T \Phi(x_i) + b \right] \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \, , i = 1, \ldots, l \tag{6}$$

The optimization problem (5) will guarantee to maximize the hyperplane margin while minimizes the cost of error. $\xi_i, i = 1, \ldots, l$ are non-negative, and are being introduced to relax the constraints of separable data problem to the constraint (6) of non-separable data problem. For an error to occur the corresponding ξ_i must exceed unity, so $\sum_i \xi_i$ is an upper bound on the number of training errors. Hence an extra cost

$C\sum_i \xi_i$ for errors is added to the objective function where C is a parameter chosen by the user.

In test phase an SVM is used by computing the sign of

$$f(x) = \sum_i^{N_S} \alpha_i y_i \Phi(s_i)^T \Phi(x) + b = \sum_i^{N_S} \alpha_i y_i K(s_i, x) + b \tag{7}$$

where the S_i are the support vectors, N_S is the number of support vectors, K is kernel with $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$, Φ is a mapping to map the data to some other (possibly infinite dimensional) Euclidean space. One example is Radial Basis Function (RBF) kernel $K(x_i, x_j) = e^{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2}$.

4.2 Hypothesis Testing

The verification task can be stated as a hypothesis testing between the two hypotheses: the input is from the hypothesis person (H_0), or not from the hypothesis person (H_1).

Let λ_0 is the model of the claimed person and λ_1 is a model representing all other possible people, i.e. impostors. For a given input x and a claimed identity, the choice is between the hypothesis H_0: x is from the claimed person λ_0, and the alternative hypothesis H_1: x is from the impostor λ_1. A claimed person's score L(x) is computed to reject or accept the person claim satisfying the following rules

$$L(x) = \begin{cases} \geq \theta_L & \text{accept} \\ < \theta_L & \text{reject} \end{cases} \tag{8}$$

where θ_L is the decision threshold.

Let x be an EEG feature vector, the probability of x belonging to the class y is defined as $P(x|\theta_y) = ce^{yf(x)}$ where c is normalization factor and $f(x)$ is from (7).

If $x_1, .., x_k$ is a sequence of independent identical density feature vectors of class y, the probability of $x_1, .., x_k$belonging to the class y in the **AND case** is:

$$P(x_1, \dots, x_k|\theta_y) = \prod_{i=1}^{k} ce^{yf(x_i)} = c'e^{\sum_{i=1}^{k} f(x_i)} \tag{9}$$

Then the score $L(x)$ in (8) for SVM will become

$$L_{AND}(x) = P(x_1, \dots, x_k|\theta_y) = c'e^{\sum_{i=1}^{k} f(x_i)} \tag{10}$$

$$L'_{AND}(x) = \sum_{i=1}^{k} f(x_i) \tag{11}$$

5 Experiments and Results

5.1 Data Set

The data were collected from 3 healthy male subjects using Emotiv Epoc headset [6] which has 14 channels and 128Hz sampling rate. Neither training nor practice was conducted before the data collection. Each of subjects performed the actions listed below in a number of trials in two sessions on different days.

1. Blink the left eye
2. Blink the right eye
3. Raise the eyebrows
4. Move the mouth to left
5. Move the mouth to right
6. Move the tongue to left (inside of the mouth)
7. Move the tongue to right (inside of the mouth)
8. Roll the tongue up (inside of the mouth)

Subject 1 performed each of the actions 50 trials, and each of Subject 2 and 3 performed 28 trials. The data was collected by using open source software Experiment-Wizard [7]. Therefore, we can obtain the raw data of the signals. We tried to keep our experiments as simple and as plain as possible so that we can study the baseline recognition rates of these heavily artifacts-influenced EEG signals.

After collecting data, the facial muscle actions were classified using LibSVM [4] wrapped in WEKA [8] to find out the stable and high accuracy tasks. The result as seen in Table 1 showed that the recognition rates for the actions and the subjects vary greatly. The results reflect the great degrees of freedom in performing some of the actions, due to no training or practicing beforehand. It seems that the ambiguity in the action description contributes to these low rates, because the physical actions may not be performed consistently. Actions 1 and 2 have consistent good recognition rates, yet more accurate descriptions of the actions, for example, “blinking the left eye firmly” etc., can also further decrease the ambiguity of the physical actions, and therefore may increase the recognition rates.

To sum up, Blink the left eye (task 1) and Blink the right eye (task 2) gave high and stable classification rates. As a result, the data of these two tasks were chosen for user authentication.

Table 1. Classification rate of 8 EEG artifact tasks using facial muscle movement

Task Subject	Task1	Task2	Task3	Task4	Task5	Task6	Task7	Task8
S01	85.7%	92.9%	85.7%	57.1%	57.1%	35.7%	71.4%	85.7%
S02	85.7%	85.7%	42.9%	57.1%	42.9%	42.9%	28.6%	42.9%
S03	100%	100%	71.4%	42.9%	28.6%	14.3%	85.7%	14.3%

5.2 Feature Extraction

The collected EEG signals were cut into segments of 5 seconds. Each segment contains an artifact. The signals from electrodes F3, F4, C3, C4, P3, P4, O1 and O2 for investigation on the frontal, central, parietal and occipital sites. The channel signals in each segment were used to extract features and these features were merged together to make a single feature vector.

The power spectral density (PSD) in the band 8-30 Hz was estimated. The Welch's averaged modified periodogram method was used for spectral estimation. Hamming window was 1 second 50% overlap. There were 12 power components extracted.

Besides PSD features, autoregressive (AR) model parameters were extracted. Burg's lattice-based method was used with the AR model order 11th.

5.3 Results

The SVM method was used with separate training set and test set. The RBF kernel function $K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}$ was used. The parameters for SVM training are γ and c. The parameter γ was searched in $\{2k: k = -4, -3, \ldots, 1\}$. The parameter c was searched in $\{2k: k = -1, -2, \ldots, 3\}$. The best parameters found are $c = 8, \gamma = 0.25$.

Due to the levels of security, the matching policy can be single task matched or a combination of tasks T_1, T_2, T_3, and T_4 matched in cases $\mathrm{AND}(\wedge)$. For example all of tasks in the right order, e.g. $(T_1 \wedge T_2 \wedge T_3 \wedge T_4)$.

Table 2 and Figure 2 present the authentication results when users doing different single facial muscle movement tasks as well as combined tasks. Table 3 is a comparison of present work with some recent EEG based authentication studies.

Table 2. Equal Error Rate (EER) in authentication of 3 persons S01-S03 using the facial muscle movement Blink left eye and Blink right eye

Task EER	Blink Left Eye (Task 1)	Blink Right Eye (Task 2)	Task 1 AND Task 2 (Task 1 $\wedge$ Task 2)
EER	1.87%	1.45%	1.15%

Table 3. Some EEG based person authentication results and other biometric system extracted from cited literature and the our work

Study	# Subjects	Experimental Modality	EER
Marcel and Millán 2007 [10]	9	EEG mental tasks, 32 channels	6.6%
Safont et al. 2012 [17]	50	EEG resting state, 2 channels	2.4%
Nguyen et al. 2013 [13]	9	EEG motor imagery, 3 channels	2.21%
Nakagawa et al. 2012 [11]	35	Voice	0.72%
Zhao et al. 2012 [20]	100	Fingerprint	0.20%
Our work	3	EEG artifact tasks, 8 channels	1.15%

There are two types of errors: False Acceptance and False Rejection. False Acceptance error occurs when the system accepts an impostor, and False Rejection error occurs when the system rejects a true client. Performance of a system is evaluated by Decision Error Trade-off (DET) curve, which is a plot of False Acceptance Rate (FAR) on y-axis versus False Rejection Rate (FRR) on x-axis. To compare the accuracy of systems with different DET curves, researchers use Equal Error Rate (EER) that is a point on a DET curve where FAR and FRR are equal. The lower DET curve, as well as smaller EER, is considered as the better authentication system. Figure 2 illustrates the FRR and FAR when using single and combination of different artifact tasks. We can see that different single artifacts have different authentication accuracies. Moreover, the results in Table 2 and the DET curves confirm that errors are significantly reduced when tasks are combined together in multiple matched policy (AND(∧) task combination). In addition, with multiple matched policy, it is much more difficult for an imposter to access system that means the security is considerably strengthened. To sum up, EEG-based user authentication using artifact is not only easy to implement, but also suitable for multilevel security systems.

Table 3 shows that the performance of EEG based authentication system using artifact is interesting as other EEG based modalities and other biometric system such as fingerprint and voice, so it can be expected to reach the level security of other authentication systems.

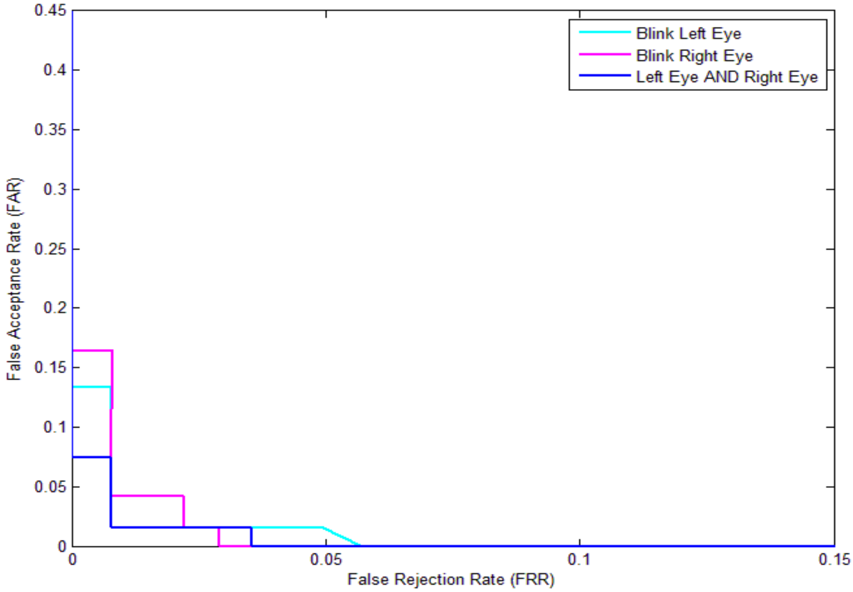

Fig. 2. DET curves of user authentication using EEG artifact signal of Blink left eye (Task 1), Blink right eye (Task 2)

6 Discussion and Future Work

EEG signals are biometric information of individuals. In additional, brain patterns correspond to particular tasks, and they be regarded as individualized passwords. As the result, EEG based authentication has all the benefits of password based and biometrics based authentication, yet without their vulnerabilities since EEG signals are difficult to fake, impossible to observe or intercept, and alive person recording require.

In this paper, we propose to take the advantage of EEG artifacts, rather than try to remove them, for an EEG based authentication system, which is simple to implement and easy to use, yet being ubiquitous without restrictions on the surrounding environments. Moreover, it can provide multilevel security systems and users a flexible authentication mechanism with different single as well as combined artifact tasks policies. Different from existing Brain Computer Interface (BCI) systems, our proposed system using EEG signals generated from well purposely designed facial muscle movements treats artefacts as an information carrier about human intentions. EEG artifacts also can be used in BCI systems for disabled people to issue simple commands to control devices, such as artificial limbs, wheelchairs etc., and communicate with other people via a computerized device. Moreover, people without disabilities can use EEG artifacts as an extra means of human machine interaction in various real world applications, such as electronic gaming, communication with wearable computers.

The dataset in this paper is small in terms of the number of subjects involved with 3 subjects; however, the experiments were designed and conducted considerately. The number of trials is quite large with 50 trials for subject 1 and 28 trials for each subject 2 and subject 3. Moreover, data were recorded in two sessions in different days. The data from one session was used for training and the other one for testing. This ensures that data in the same trial or the same session were not used in both the training and testing datasets, so the validation is acceptable.

Although the preliminary results are encouraging, there are still questions need to be answered.

The common facial muscles movements can be performed in a very similar manner by everybody without the need of special training. Among these actions, which ones can produce stable and repeatable patterns? For these actions which can be easily performed, yet produce stable and repeatable patterns, are the patterns only valid for each individual or the whole population? If the former, the system must be calibrated by a training phase before it can be used.

EEG signals are heavily influenced by the artifacts when a subject doing the facial muscle movements, so do we still need to collect the signals from the scalp of a subject? Can we find other alternative spots on the face, with only a few electrodes and may be somewhere without hair. If possible, attaching electrodes becomes a very simple task, yet with a high level of operation accuracy.

The artifact tasks can be combined together flexibly during authentication depending on the level security of system. The more combined tasks can provide the better security to the system, but how many tasks are enough and what are they for very high security level? This need to be more investigated with larger dataset.

The representing features and the machine learning algorithms used in our experiments so far are these of commonly used in processing EEG signals. Are they the best choices for the EEG signals with the strong presence of artifacts? More experiments on a much larger dataset are required.

In the future, collecting more data from more individuals and in many different environments is our top priority task. After conducting more experiments on the data, we can then answer the above questions.

References

1. Brigham, K., Kumar, B.V.K.V.: Subject Identification from Electroencephalogram (EEG) Signals During Imagined Speech. In: Proc. IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems (BTAS 2010) (2010)
2. Brown, L.: Computer Security: Principles and Practice. William Stallings (2008)
3. Burges, J.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998)
4. Chang, C., Lin, C.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 27 (2011)
5. Dix, A.: Human–computer interaction: A stable discipline, a nascent science, and the growth of the long tail. Interacting with Computers 22(1), 13–27 (2010)
6. Emotiv EPOC headset, `http://www.emotiv.com/`
7. Experiment Wizard software tool, `http://code.google.com/p/experiment-wizard/`
8. Hall, M., et al.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
9. Lenman, S., Bretzner, L., Thuresson, B.: Using marking menus to develop command sets for computer vision based hand gesture interfaces. In: Proceedings of the Second Nordic Conference on Human-Computer Interaction 2002, pp. 239–242. ACM, Aarhus (2002)
10. Marcel, S., Millán, J.R.: Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 743–752 (2007)
11. Nakagawa, S., Wang, L., Ohtsuka, S.: Speaker Identification and Verification by Combining MFCC and Phase Information. IEEE Transactions on Audio, Speech, and Language Processing 20, 1085–1095 (2012)
12. Nina, H., et al.: Integrating cognitive load theory and concepts of human–computer interaction. Computers in Human Behavior 26(6), 1278–1288 (2010)
13. Nguyen, P., Tran, D., Huang, X., Ma, W.: Motor Imagery EEG-Based Person Verification. In: Rojas, I., Joya, G., Cabestany, J. (eds.) IWANN 2013, Part II. LNCS, vol. 7903, pp. 430–438. Springer, Heidelberg (2013)
14. Palaniappan, R.: Two-stage biometric authentication method using thought activity brain waves. International Journal of Neural Systems 18 (2008)
15. Safont, G., Salazar, A., Soriano, A., Vergara, L.: Combination of multiple detectors for EEG based biometric identification/authentication. In: 2012 IEEE International Carnahan Conference on Security Technology (ICCST), pp. 230–236 (2012)
16. Sanei, S., Chambers, J.: EEG signal processing. Wiley-Interscience (2007)

17. Sun, S.: Multitask learning for EEG-based biometrics. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4 (2008)
18. Turk, M., Robertson, G.: Perceptual user interfaces. Communications of the ACM 43(3) (2000)
19. Welch, P.: The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodogram. IEEE Trans. Audio Electroacoustics, 70–73 (1967)
20. Zhao, W., Zhang, H.: Secure Fingerprint Recognition Based on Frobenius Norm. In: 2012 International Conference on Computer Science and Electronics Engineering, vol. 1, pp. 388–391 (2012)