Measurement 45 (2012) 1927–1932

Contents lists available at SciVerse ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Increasing the accuracy of ECG based biometric analysis by data modelling

Saurabh Pal [a,*], Madhuchhanda Mitra [b]

[a] Dept. of Applied Electronics and Instrumentation Engg., Heritage Institute of Technology, Kolkata, India
[b] Dept. of Applied Physics, University of Calcutta, Kolkata, India

ARTICLE INFO

Article history:
Received 17 June 2011
Received in revised form 20 February 2012
Accepted 6 March 2012
Available online 21 March 2012

Keywords:
Authentication
Best fit
Biometry
ECG
PCA

ABSTRACT

Here an effort is made to use human electrocardiogram as a tool of biometric analysis for authentication. The proposed method is based on first accurate extraction of characteristic features from each ECG and then design of a suitable classification methodology to comment on the authenticity. As the feature matrix is a huge one, Principal Component Analysis (PCA) is applied to avoid handling of large amount of data. Next, the reduced features from PCA are fitted into a quadratic polynomial model by the method of least square. Then the fitted values for the allowed set of data is obtained and the range over which they vary, provides the signature matrix of a person. Finally the classification is done by a comparison based on nearest neighbor method. The method is tested on ECG of 20 individuals taken from PTB database. This method has accuracy more than 95% with the best fit modeling which becomes only 80% without data modeling proving the importance of best fit modeling of data before classification. This accuracy is comparable with conventional biometric techniques; moreover, ECG biometric can be used with other authentication scheme, with ECG providing liveliness proof.

1. Introduction

The electrocardiogram (ECG) is a record of time-varying bioelectric potential generated by electrical activity of the heart. It has become a fundamental tool of diagnosing different cardiac and blood circulation abnormalities [1]. The interpretation of ECG leads to a decision on electrical or functional abnormality of heart. In recent past, different studies have been conducted for the use of ECG in biometrics apart from its conventional usage. Biometric analysis and recognition provides security and restricted access to protected areas by identifying the persons using his/her physiological or behavioural features. Presently human fingerprint, face, iris or voice, anatomical traits and behavioral characteristics like signature characteristics and dynamics, etc. are the features that are being used singly or a fusion of more than one of them in biometric recognition systems. However, these biometrics modalities either cannot provide reliable performance in terms of recognition accuracy (e.g., gait, keystroke) or they are prone to falsification externally. Some examples of falsification are 3D fingerprint and face models or plastic surgery on these features, voice playback from a different source and contact lenses with different iris features printed on. In this regard, ECG has a great opportunity to be used as a non-falsifiable biometric parameter. As ECG is the electrical activity generated due to the auto-rhythm of pacemaker cells of cardiac dipole, it is almost impossible to modulate the signal from the outside world.

Recently some research has made to test the applicability of ECG as a biometric feature [2–4]. The possibility to use ECG as a biometric feature is supported by the fact that there is uniqueness in the individual ECG because of physiological and geometrical differences of the heart of different persons [5–7]. Biel et al. [8] have conducted the biometric experiment on ECG recorded from a group of 20 subjects. Twelve features have been selected from each record for identification of a person in a predefined group. Shen et al. [9] have investigated the feasibility of ECG as a new biometric for identity verification. The experiment has been conducted on 20 individuals on seven features,

* Corresponding author.
E-mail address: spal76@gmail.com (S. Pal).

0263-2241/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.measurement.2012.03.005

extracted from mainly QRS complex. Using the techniques of neural network and template matching the experiment of human identity verification has been performed. Irvine et al. [10] proposed a technique based on heart rate variability and some other temporal features with only a group of five subjects. A QRS complex based approach with its fourth order Legendre Polynomial as the signature is also proposed by Khalil et al. [11] with good accuracy. Wang et al. [12] has proposed methods using fiducial points and without fiducial points by AC/DCT technique for two groups with 13 subjects in each. Classification is done based on linear discrimination analysis and neural network based technique. In AC/DCT method similarity between the subjects is measured based on normalized Euclidian distance and a nearest neighbor is used as the classifier. But this ECG morphology based approach may fail when appearance of two ECGs are similar. Chan et al. [13] use a wavelet distance measurement technique for classification of 50 subjects with accuracy 89%. In [14] S.C. Fang et al. have proposed a technique for person identification by making the classification based on the similarities and dissimilarities on electrocardiogram phase space. A PCA/LDA based approach is also suggested by Baumbarov et al. [15]. A non-fiducial feature based technique is reported by J.L.C. Loong in [16] using spectral coefficients computed through linear predictive coding (LPC) and classification is done using neural network based approach. Double fold approach is proposed by Safie et al. in [17] where the parameters of the pulse active ratio (PAR) feature vector are represented by a four digit PIN number. Authentication is made for 20 subjects first by verifying the PIN number and finally by ECG feature vector matching. Most of the reported methods either suffers from lack of good accuracy or requires complicated mathematical procedures for processing and classification. Moreover, some of the previous works use 12 lead ECG for biometric authentication which is impractical to use in real biometric system in spite of good result.

In this paper we investigate the applicability of ECG as a biometric parameter with a requirement of fiducial detection for 20 subjects. Here 16 parameters including amplitude, temporal and angular features are extracted. Almost all works regarding ECG based biometric authentication, some data reduction technique is required to handle this large set of database. Principal Component Analysis (PCA) is a commonly used tool used for this purpose as used in the present work also. Additionally, here it is shown that classification accuracy greatly improves if the resulting data after PCA is modeled by a quadratic polynomial based curve fitting algorithm as then higher half of the reduced feature set is better discriminated. The steps of the proposed algorithm are shown in Fig. 1.

Not only for human identification, the technique of using ECG data in biometry offers some unique advantages. In automatic cardiac care units, it is required to monitor the patients continuously. This technique can provide identity of persons remotely without the requirement of any additional data processing. Thus it is advantageous to verify a patient's identity in medical records or prior to drug administration or other medical procedures from a remotely located control room.

2. Materials and methods

2.1. Signal pre-processing and feature extraction

In most of the cases recorded ECG data are corrupted by various high and low frequency noises arising from power line interference (for high frequency noise) and respiration, body movement, EMG, etc. (for low frequency noise). ECG filtering from any kind of above mentioned noise is mandatory from biometric analysis otherwise wrong estimation of features may lead to misclassification. In this work wavelet transform based filtering and feature extraction is performed. The method is taken from [18]. Wavelet transform is basically a convolution operation between the mother wavelet and the test signal as the mother wavelet translates along the test signal in time axis. Here db6 is chosen as mother wavelet due to its structural resemblance with the QRS complex of the ECG signal and decomposition is made up to level eight. Fig. 2 shows the decomposition of a typical ECG signal with db6 wavelet. The coefficient of level one may be considered as mostly noise with respect to the important high frequency parameters of ECG when the sampling frequency of the mother signal is 1000 Hz. According to the power spectra of the signal [19], it is clear that most energy of the QRS complex is concentrated at decomposition level 3, 4 and 5. The reconstructed wave with these coefficients is enhanced and the highest potential point is considered as R peak. Generally the Q and S waves are high frequency and low amplitude waves and their energies are mainly prominent at small scale. For that decomposition coefficients from d2 to d5 are retained and a five point differentiation is made to find out the point of inflections for Q and S points on either side of R peak. The energies of T and P waves are mainly at scale levels 6, 7 and 8. But, low frequency base-

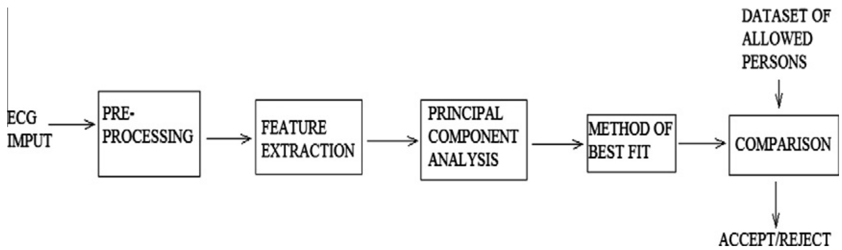

Fig. 1. Block diagram of the entire procedure.

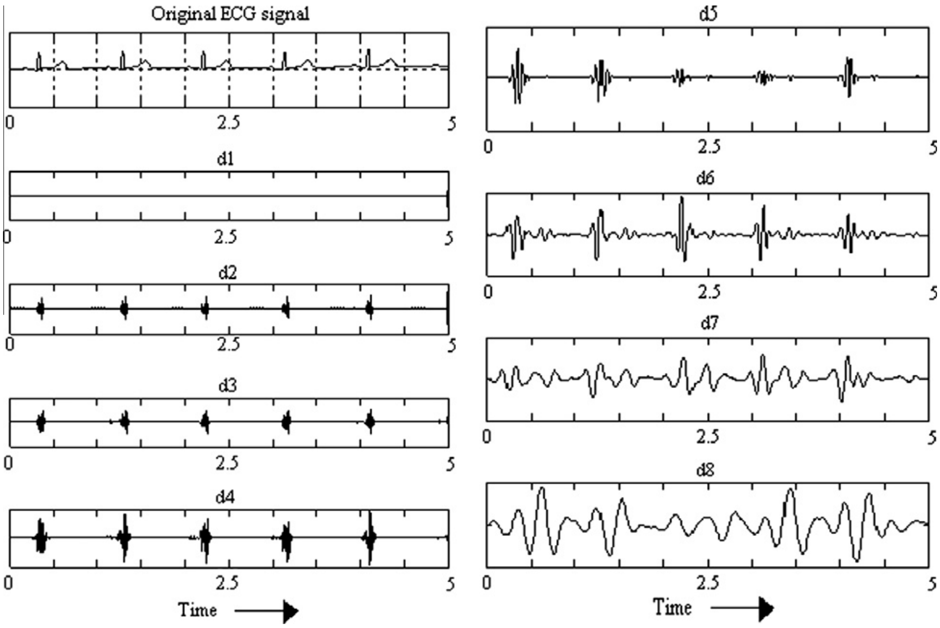

Fig. 2. Wavelet decomposition of ECG up to level eight.

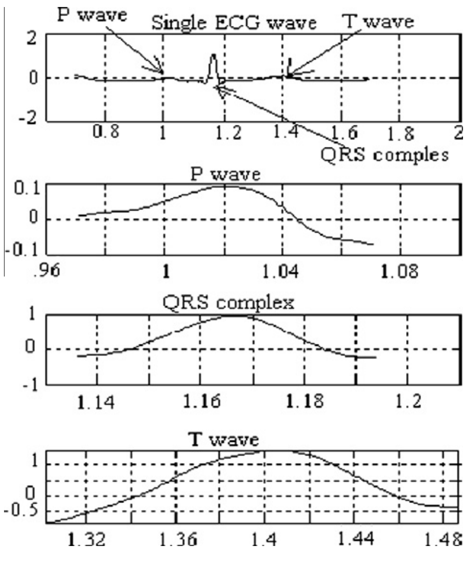

Fig. 3. Detection of fiducial points.

line drift may appear at scale 8, so reconstruction coefficients d6 and d7 are selected to detect T and P waves. Then the T peak is identified as the maxima after the detected S point within a predefined interval. As the T peak is pointed out, T onset and T offset is found out as the minimum potential crossing points on either side of the T peak. Fig. 3 shows the detected QRS complex, P and T wave for a typical ECG waveform. As it is not possible to define the features which differ from different persons in a group, maximum possible numbers of features are extracted for better accuracy of the analysis.

Figs. 4A–4C are the pictorial representation of the features used for biometric authentication which is tabulated in Table 1.

2.2. Data transformation by principal component analysis

Principal Component Analysis (PCA) [20] is one of the oldest and most widely used data transformation techniques for multivariable analysis. The dimension of input dataset is reduced using this technique. PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Let X be the N dimensional dataset of length L where each column represents a specific parameter extracted during multiple observation. The empirical mean of the dataset is calculated as

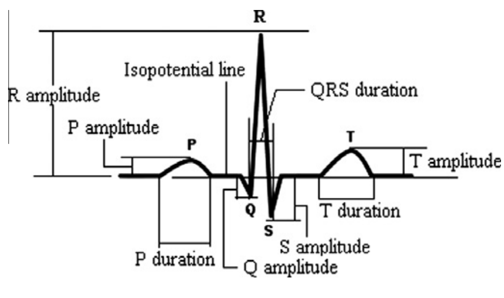

Fig. 4A. Amplitude and duration features.

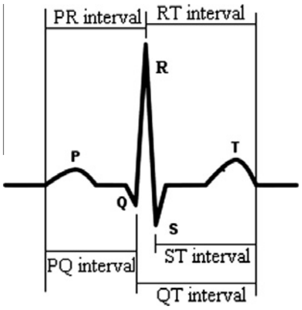

Fig. 4B. Interval features.

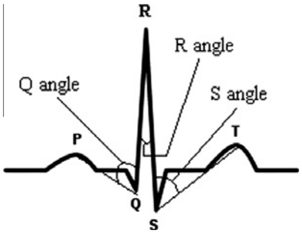

Fig. 4C. Angle features.

Table 1
Features extracted for biometric analysis.

P amplitude	T amplitude	PQ interval	QT_c interval
Q amplitude	P duration	PR interval	Q angle
R amplitude	QRS duration	RT interval	R angle
S amplitude	T duration	ST segment	S angle

$$E(m) = \frac{1}{N}\sum_{n=1}^{N} X(m,n) \tag{1}$$

The deviation from the mean is given by

$$D = X - Eu \tag{2}$$

where $u(n) = 1$ for $n = 1, \ldots, N$. Then the covariance matrix is calculated as

$$C = \frac{1}{N}\sum D \cdot D^* \tag{3}$$

where D^* stands for the transpose matrix of D. Next the Eigen vector matrix G is calculated which diagonalizes the covariance matrix. Hence,

$$G^{-1}CG = V \tag{4}$$

where V is the diagonal matrix of Eigen values of C. Next, the diagonal matrix V is arranged in descending order and a specific subset of it is selected as basis vector. A plot of the basis vector for a typical record is shown in Fig. 5.

2.3. Data fitting

The output of the data reduction technique shows that the reduced feature values decrease appreciably in such a way that it becomes difficult to classify the data for authentication especially in the upper half of the data set

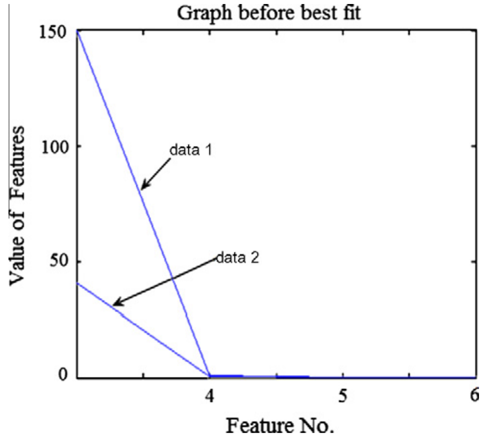

Fig. 5. Basis vector before best fit.

which results in less accuracy for classification as indicated in the next sections. Hence the resulting data set is statistically modeled based on explicit mathematical function.

The standard form of the statistical model is denoted by,

$$y = f(\vec{x}, \vec{a}) + k \tag{5}$$

where y is the response or output, $\vec{x}$ stands for the collection of all the predictor variables and $\vec{a}$ is the collection of all parameters in short. k is the random error.

Polynomial models are most frequently used for mathematical modeling. A general polynomial model is given by,

$$y = a_n x^n + a_{n-1}x_{n-1} + \cdots + a_2 x_2 + a_1 x_1 + a_0 \tag{6}$$

where n is a non-negative integer denoting the degree of the polynomial. Besides its simplicity and flexibility, polynomial model have the advantage that the mapping retains the input structure of the data. Observation of the PCA output enables to select the quadratic polynomial model as,

$$y = a_0 + a_1 x + a_2 x^2 \tag{7}$$

where a_0, a_1, a_2 are modeling coefficients. Modeling coefficients are calculated by least square fitting method by least

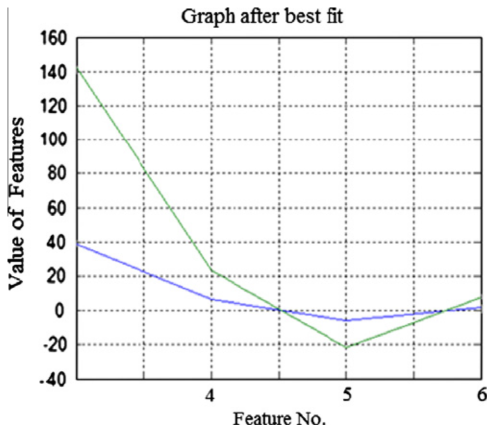

Fig. 6. Signature matrix after data modeling.

square estimation. Mathematically the least square criteria is given by,

$$\varepsilon = \sum_{i=1}^{n}[y_i - f(\vec{x}, \vec{a})]^2 \quad (8)$$

In this technique the unknown values of the parameters are so calculated that the sum of the squared deviations between the input and the functional portion of the model is minimized. The output after best fit is shown in Fig. 6. It is clear from this figure that now it will be easier for any classifier to discriminate the patterns.

3. Comparison for authentication

As per the previous steps, the ECG beat pattern for each person is transformed into a set of six best fit values providing the biometric signature for each person. For a classifying system for authentication, it is required to accept a predetermined set of signatures and reject the others. Here the classification is done using a simple nearest neighbor method. It marks a new entry as the class corresponding to a stored signature that gives the minimum distance for each best fit value. Any test signature matrix element value is considered similar to a stored data if the test value falls within ±2% of the stored one.

4. Result and analysis

The algorithm is tested against 20 databases taken from PTB diagnostic data from physionet data bank [21]. All beats are normal and having 1 kHz sampling frequency. 16 features are extracted from each wave as stated earlier.

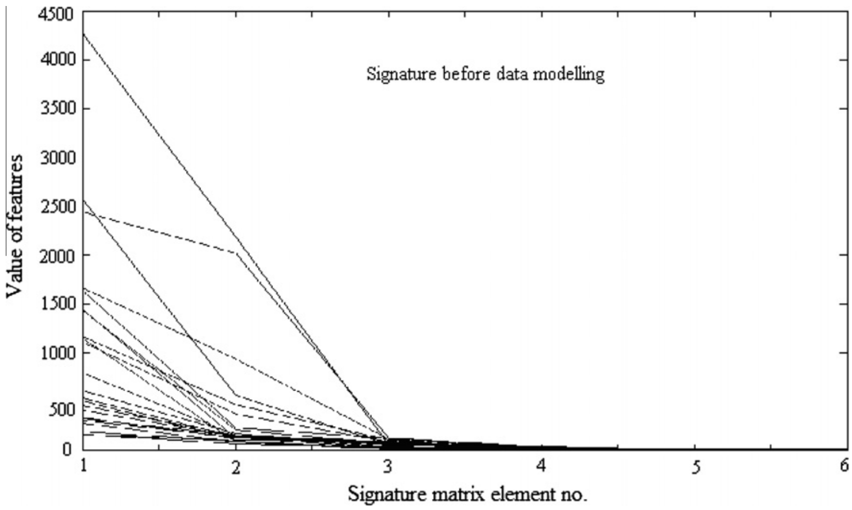

Fig. 7. All data graph of basis vector before data modeling.

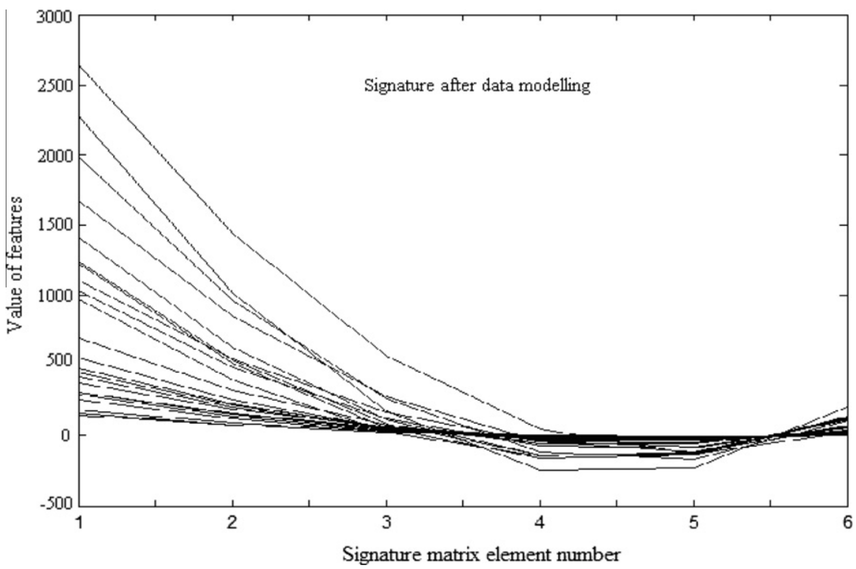

Fig. 8. All data graph of signature matrix after data modeling.

To make the result more convenient, all the features are detected nine times from nine different waves of same record taken in some arbitrary intervals thus generating a 16×9 feature matrix. Then the use of PCA generates the signature matrix with dimension 1×6. The resulting output is shown in Fig. 6. It shows the overlapping nature of the matrix mainly in the last three elements which makes classification less accurate.

Quadratic polynomial based best fit modeling modifies the signature matrix making the higher order attributes more significant as shown in Fig. 7. From Figs. 7 and 8 it is seen that the discrimination of the signature matrix for same element especially in the higher order is better if the data is modeled with the best fit algorithm as discussed earlier. As the spread of the signature is more, it is easier to classify using nearest neighbor method. The comparison of each element of the test signature is made with the stored patterns and the signature with minimum distance with most of the elements is identified as equivalent to the test entry. Without the best fit modeling, the accuracy was 80% whereas it becomes 95% with best fit modeling.

5. Discussion

In this work the applicability of ECG for biometric authentication is studied. ECG has a good potential to be a biometric parameter as the signal source is beyond the scope of falsification.

Moreover, it can be used as a liveliness detector along with the biometric application. There are basically two types of methods used for ECG based biometric analysis – one is feature detection based and depending on overall texture of the signal is the other. Most of the existing works uses feature extraction followed by a classification algorithm based on ANN or other. In this technique, normally for better accuracy of biometric signature, a number of features are detected and each feature is measured several times from different signals of the same person to make it robust and reliable. Hence the resulting data for each person is a huge one and needs to be dimensionally reduced retaining the essence of parameters. Principal Component Analysis is a well adopted tool for the same. Here the classification is made by the measurement of nearest neighbor from the stored set of data of signature matrix by comparison with the new entry. It is seen that the accuracy level is 80% for person identification. It is noticed that the higher half of the signature matrix for most of the persons are quite similar leading to misclassification in some cases. Hence a novel quadratic curve fitting algorithm is proposed based on least square fit of the modeled data. This best fit data shows greater discrimination at the lower half of the signature matrix for each person and the accuracy level goes up to 95% for person identification. This method is also useful for tracking the identity of the patients remotely without any additional data processing. Thus automatic monitoring or other medical procedures may be done without any human intervention. Other physiological or behavioral biometric parameters may be fused with the ECG based technique to make the authentication more reliable.

References

[1] P. Kligfield, The centennial of the einthoven electrocardiogram, Journal of Elctrocardiology 35 (2002) 123–129.
[2] M. Kyoso, A. Uchiyama, Development of an ECG identification system, in: Proceedings of 23rd IEEE Engineering Medicine and Biology Conference, 2001, vol. 4, pp. 3721–3723.
[3] M. Kyoso, A technique for avoiding false acceptance in ECG identification, in: Proceedings IEEE EMBS Asian–Pacific Conference Biomedical Engineering, 2003, pp. 190–191.
[4] S.A. Israel, J.M. Irvine, A. Cheng, M.D. Wiederhold, B.K. Wiederhold, ECG to identify individuals, Pattern Recognition 38 (1) (2005) 133–142.
[5] R. Hoekema, G.J.H. Uijen, A. van Oosterom, Geometrical aspect of the interindividual variability of multilead ECG recordings, IEEE Transactions Biomedical Engineering 48 (2001) 551–559.
[6] C.C. Chiu, C.M. Chuang, C.-Y. Hsu, A novel personal identity verification approach using a discrete wavelet transform of the ECG signal, in: International Conference on Multimedia and Ubiquitous Engineering, Busan, Korea, 2008, pp. 201–206.
[7] Y. Singh and P. Gupta. ECG to individual identification, in: 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, Washington, DC, 2008, pp. 1–8.
[8] L. Biel, O. Pettersson, P. Lennart, W. Peter, ECG analysis: a new approach in human identification, IEEE Transaction on Instrumentation and Measurement 50 (3) (2001) 808–812.
[9] T.W. Shen, W.J. Tompkins, Y.H. Hu, One-lead ECG for identity verification, in: Proceedings of the Second Joint EMBS/BMES Conference, 2002, pp. 62–63.
[10] J.M. Irvine, B.K. Wiederhold, L.W. Gavshon, S. Israel, S.B. McGehee, R. Meyer, M.D. Wiederhold, Heart rate variability: a new biometric for human identification, in: Proceedings International Conference Artificial Intelligence, 2001, pp. 1106–1111.
[11] Ibrahim Khalil, Fahim Sufi, Legendre polynomials based biometric authentication using QRS complex of ECG, ISSNIP, 2008, pp. 297–302.
[12] Yongjin Wang, Foteini Agrafioti, Dimitrios Hatzinakos, Konstantinos N. Plataniotis, Analysis of human electrocardiogram for biometric recognition, EURASIP Journal on Advances in Signal Processing, vol. 2008, 2008 (Article ID 148658).
[13] Adrian D.C. Chan, Mohyeldin M. Hamdy, Armin Badre, Vesal Badee, Wavelet distance measure for person identification using electrocardiograms, Transactions on Instrumentation and Measurement 57 (2) (February 2008) 203–248.
[14] S.C. Fang, H.L. Chan, Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space, Pattern Recognition 42 (9) (2009) 1824–1831.
[15] O. Boumbarov, Y. Velchev, S. Sokolov, ECG personal identification in subspaces using radial basis neural networks, in: IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2009, pp. 446–451.
[16] J.L. Cheang Loong, K.S. Subari, R. Besar, M.K. Abdullah, A new approach to ECG biometric systems: a comparative study between LPC and WPD systems, World Academy of Science Engineering and Technology (68) (2010).
[17] S.I. Safie, J.J. Soraghan, L. Pertopoulakis, ECG based biometric for doubly secure authentication, in: 19th Europian Signal Processing Conference (EUSIPCO 2011), 2011, pp. 2274–2278.
[18] Saurabh Pal, Madhuchhanda Mitra, Detection of ECG characteristic points using multiresolution wavelet analysis based selective coefficient method, International Journal of Measurement, Elsevier Science Publication, vol. 43, February, 2010, pp. 205–261.
[19] N.V. Thakor, J.G. Webster, W.J. Tompkins, Estimation of QRS complex power spectra for design of a QRS filter, IEEE Transactions on Biomedical Engineering BME 31 (1984) 702–705.
[20] I.T. Jolliffe, Principal Component Analysis, Second., Springer-Verlag, New York, 2002.
[21] www.physionet.org.