IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
21-23 September 2009, Rende (Cosenza), Italy

ECG Personal Identification in Subspaces using Radial Basis Neural Networks

Ognian Boumbarov, Yuliyan Velchev, Strahil Sokolov
Technical University of Sofia, 8, Kl. Ohridski str., Sofia 1000, Bulgaria,
olb@tu-sofia.bg, julian_velchev@abv.bg, s_sokolov@tu-sofia.bg

***Abstract* - In this paper an approach for personal biometric identification is presented based on extraction of ECG features and classification with RBFNN. We perform denoising and segmentation on the input signal, after which we realize dimensionality reduction and feature extraction based on PCA transform. The separability of the selected features is improved by applying LDA. The final stage of the proposed approach is classification and recognition of the extracted features with classifier score fusion.**

***Keywords* – ECG personal identification, neural networks, PCA, LDA, HMM**

I. INTRODUCTION

The importance of personal biometric identification has been growing over the last 20 years. There are several biometrics that have proven to be robust and are used in major identification tasks – such as fingerprint, signature, voice and face [7, 8]. However, these modalities can't provide enough reliability, especially for systems with high security demands.

It is proven that physiological differences in human heart for different individuals reflect in ECG as unique time intervals, amplitudes and shapes [14]. Also ECG analysis is attractive for human identification purposes because the ECG acquisition is relatively easy for implementation. In recent years efforts have been made to improve existing technology and to introduce new biometrics for identification – such as ECG [2, 18, 19]. An ECG person identification method using a set of temporal and amplitude features is described in [2]. The system has been tested with a database of about 20 individuals. The achieved recognition performance is near 100%, but the described system is not fully automatic. On the other hand, the values of amplitude features depend very strongly on ECG acquisition hardware, placement of the electrodes and current condition of the recognized individual. So, in our approach the amplitude parameters of the ECG signal are completely rejected as features for personal identification. Another method for ECG personal identification is given in [17]. It uses a combination of autocorrelation with Discrete Cosine Transform (DCT). It refers as the AC/DCT method. The major advantage is avoiding the ECG segmentation, which is one of the most complicated tasks in automatic ECG analysis. This method involves the following stages: signal windowing in order to extract a portion with two or more cardiac cycles; normalized autocorrelation function (ACF) estimation for each window; performing DCT over each ACF; classification on significant DCT coefficients. One of the method's drawbacks is that the proper window size selection for ECG segmentation is critical and strongly depends on current heart rate.

The rest of the paper is organized as follows: in section II we give a detailed overview of the proposed ECG identification system. We describe the main methods for signal acquisition, denoising and segmentation in point II-A. In the next section II-B we present our approach for dimensionality reduction and feature selection from the segmented ECG signal. In section II-C an improvement of the class-separability of the features is described. Section II-D provides details on the neural network classifiers that we used for recognition. In section III we have shown some experimental results. Finally, in section IV we conclude and give a future roadmap of our research.

II. DESCRIPTION OF THE PROPOSED APPROACH

A. ECG Signal Acquisition, Denoising and Segmentation

The ECG signals are acquired with one channel standard electrocardiograph. The used lead is RA – LA from Einthoven triangle [5]. The signals are sampled with standard sampling rate of 128 Hz and 12 bit resolution. This digital ECG signal is transmitted via interface to a personal computer and stored in binary file. The acquired signals are high-pass filtered with zero-phase digital filter with cut-off frequency of 0.5 Hz in order to remove the possible baseline drift. The dominating noise in real-world ECG signals is muscle artifacts. This noise is additive and often modeled with Gaussian white noise [13]. For this reason, a Wavelet denoising with soft threshold of the Wavelet coefficients is realized.

In normal ECG there are several standard regions of interest defined named waves, complexes and segments (P wave, QRS complex, T wave, etc.). These components are related to distinct states of the cardiac muscle [5] (Fig. 1). The residual part from ECG is called isoelectric line (baseline). The ECG segmentation (onset/offset determination of the ECG components) is a very important stage after which dimensionality reduction techniques can be applied, such as PCA or LDA. In this

project ECG is segmented into complete cardiac cycles. It is desirable to perform full segmentation (boundary detection of all standard ECG waves and complexes). The baseline between P wave and next T wave is excluded from segmented signal portions because this segment is influenced by heart rate variations.

The ECG signal represents a random process and the signal has strong cyclic recurrence (for healthy persons). Due to these reasons, in our approach we use sequential probabilistic model such as Hidden Markov Model (HMM) [1, 6, 12]. The aim of the segmentation process is to assign each signal sample to a given class - wave, complex, etc. – so that classification and boundaries detection are performed within a single procedure.

The HMM is a probabilistic model which describes the statistical dependency between an observation sequence O and a sequence of hidden states S. A HMM λ is denoted with:

$$\lambda = (\mathbf{A}, \mathbf{B}, \mathbf{q}), \tag{1}$$

where $\mathbf{A}$ is transition probability matrix, $\mathbf{B}$ is observation probability distribution matrix and $\mathbf{q}$ is initial distribution vector. In this work the first order Markov process (Markov chain) is used:

$$P(s_{t+1} \mid s_t, s_{t-1}, s_{t-2}, ...) = P(s_{t+1} \mid s_t), \tag{2}$$

where s_t is the state (ECG component) for time t.

The purpose of ECG segmentation is to find the optimal state sequence S' given the HMM parameters λ and observation sequence O:

$$\begin{aligned} S' &= \underset{S}{argmax}\{P(S \mid O, \lambda)\} = \\ &= \underset{S}{argmax}\left\{\frac{p(S, O \mid \lambda)}{p(O \mid \lambda)}\right\} = \\ &= \underset{S}{argmax}\{p(S, O \mid \lambda)\} \end{aligned} \tag{3}$$

The optimal state sequence S' in (3) is determined as the most probable state sequence given the observation. The final result is achieved using Bayes' rule. An effective way to calculate (3) is to use the Viterbi algorithm [4].

HMM learning (finding the optimal HMM parameters λ) is a task, which can be done using supervised or unsupervised techniques. Unsupervised learning is related to maximum likelihood:

$$\lambda' = \underset{\lambda}{argmax}\{p(\mathbf{O} \mid \lambda)\}, \tag{4}$$

where λ' represents HMM parameters, which maximize the probability and $\mathbf{O} = \{O_1, O_2, ..., O_N\}$ are the observation sequences.

The model's architecture can be seen in Fig. 1. It is according to the standard ECG components and it is valid for most common ECG morphology with positive deflection of the standard waves.

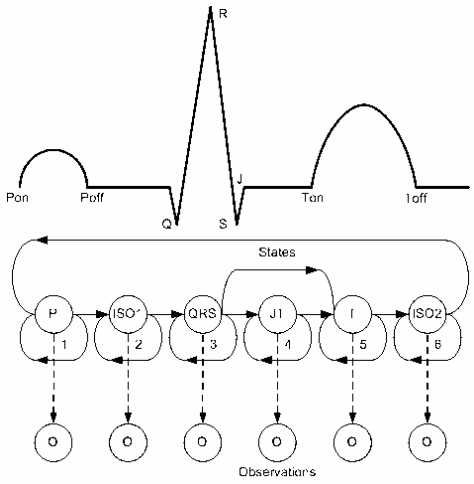


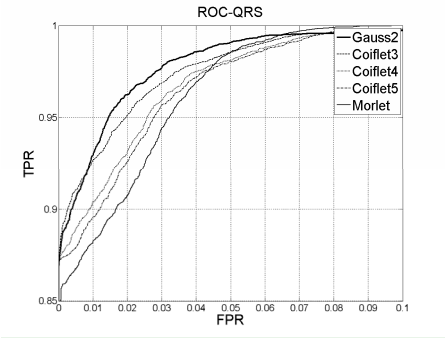
Fig. 1. Architecture of the sequential probabilistic models for ECG segmentation.

Sometimes the JT segment is hard to be distinguished, so the model allows direct transition from QRS complex to T wave.

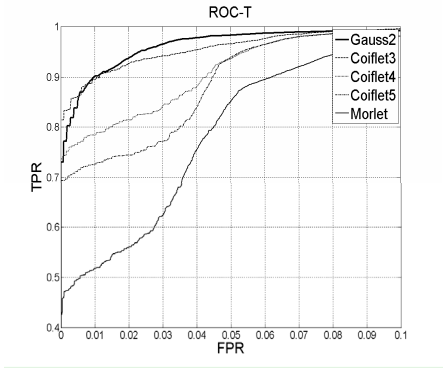
Although it is possible to use signal samples as observations, the better approach is to use observation vectors which are built from a mulitiscale linear transform such as wavelet transform:

$$W_x(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - \tau}{s}\right) dt, \tag{5}$$

where $x(t)$ is the ECG signal and ψ is the wavelet function [9]. The coefficients from wavelet transform W serve as similarity measure between signal and a wavelet function at different scales s and time shifts τ as parameters. The wavelet function must satisfy several requirements: maximum similarity between ψ and ECG components; minimal effective width; it is also desirable for ψ to be symmetric (linear phase response of the filters). The wavelet functions, which can fulfill these requirements, are: second derivative of the Gaussian function (Mexican hat), Morlet and the last three wavelets from Coiflet family. The optimal parameters for wavelet transform (wavelet function and scales) are chosen according to segmentation performance evaluation with different wavelet functions (Fig. 2a and Fig. 2b) and different scales.



a)



b)

Fig. 2. The performance of HMM – SGM based ECG segmentation with different wavelet functions for QRS complex (a) and T wave (b).

As can be seen from ROC curves, the best suited wavelet function for the purpose is second derivative of Gaussian. The optimal scales are dyadic from 2 up to 64. The features (observation vectors) for ECG segmentation are according to:

$$\mathbf{O}'_{\tau} = \left[W_x\left(2^k, \tau\right) \right]^T, \quad k = 1,...,6, \tau = 0,...,N-1, \tag{6}$$

where N is the number of samples in the ECG signal.

Because the features in a given class are not normally distributed, the used HMM has Gaussian mixture model core (HMM-GMM). Some experiments have been made by using HMM with single Gaussian model (HMM-SGM) but with considerably lower performance. We have also experimented with another sequential probabilistic model (Conditional Random Field – CRF) for ECG segmentation purposes [15]. As can be seen from the evaluation of the experimental results, the achieved performance is not satisfying.

B. Feature Extraction from ECG using Linear Subspace Projection

The Principal Component Analysis (PCA) is a well-known technique for dimensionality reduction that concentrates the discriminative information into low number of coefficients [3]. The previous stage of our approach provides noise reduction and segmentation of the ECG signal. We consider each segmented and labeled cardiac cycle a row vector with a finite number of elements N. The matrix $\mathbf{X}$ is formed from the vectors for each cardiac cycle:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \dots & \dots & \dots & \dots \\ x_{M1} & x_{M2} & \dots & x_{MN} \end{pmatrix}, \tag{7}$$

where M is the number of cardiac cycles. For the purposes of our experiment we have selected also a fixed number M of training cycles per person that are strongly correlated. The input data $\mathbf{X}$ is sorted according to its correlation matrix and the first 20 vectors are used for PCA, thus avoiding incorrectly segmented cardiac cycles (Fig. 3).

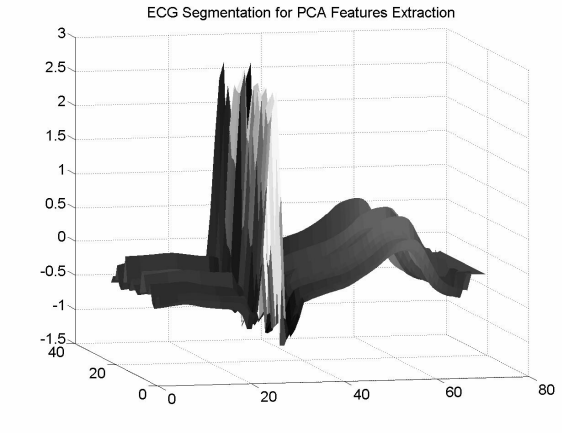


Fig. 3. ECG segmentation example for PCA features extraction.

In order to prepare the transformation in PCA subspace, we subtract the mean vector from the data. The covariance matrix of the input data $\mathbf{\Sigma}_{\mathbf{X}}$ is according to:

$$\mathbf{\Sigma}_{\mathbf{X}} = E\{\mathbf{X}\mathbf{X}^T\} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T. \tag{8}$$

The PCA is a linear transform:

$$\mathbf{Y} = \mathbf{P}^T\mathbf{X}, \tag{9}$$

where $\mathbf{Y}$ is the transformed data and $\mathbf{P}$ is the matrix for linear transformation. The goal is to diagonalize the covariance matrix of the output data:

$$\mathbf{\Sigma}_{\mathbf{Y}} = diag\{\lambda_1, \lambda_2, ..., \lambda_M\}. \tag{10}$$

For $\mathbf{\Sigma_Y}$, the following is valid:

$$\mathbf{\Sigma_Y} = \frac{1}{N-1}\mathbf{Y}\mathbf{Y}^T = \frac{1}{N-1}(\mathbf{P}^T\mathbf{X}\mathbf{X}^T\mathbf{P}) \quad = \\ = \mathbf{P}^T\mathbf{\Sigma_X}\mathbf{P} \tag{11}$$

The columns of $\mathbf{P}$ consist of the eigenvectors of $\mathbf{\Sigma_X}$. Let $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal is formed from the eigenvalues of $\mathbf{\Sigma_X}$. The eigenvalue and eigenvector decomposition is according to:

$$\mathbf{\Sigma_X}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}. \tag{12}$$

$\mathbf{\Sigma_X}$ is symmetric so $\mathbf{P}$ is orthonormal matrix and the eigenvalues of $\mathbf{\Sigma_X}$ are real numbers. For the orthonormal matrix $\mathbf{P}$:

$$\mathbf{P}^T\mathbf{\Sigma_X}\mathbf{P} = \mathbf{P}^T\mathbf{P}\mathbf{\Lambda} \tag{13}$$

and

$$\mathbf{P}^T\mathbf{\Sigma_X}\mathbf{P} = \mathbf{\Lambda}. \tag{14}$$

Obviously $\mathbf{\Sigma_Y} = \mathbf{\Lambda}$. The eigenvalues are solutions of a polynomial with degree M. The final step is to find the eigenvectors corresponding to the eigenvalues and sort them in descending-energy order

The original basis is with dimension M x M. The new basis is with reduced size M x L ($L << M$). The components with insignificant energy are omitted.

PCA is an optimal method which delivers the opportunity to perform back-projection to signal space - the mean square error (MSE) between the restored signal and original signal is minimized.

C. Improving Features Class-Separability by LDA

The Linear Discriminant Analysis (LDA) is a method closely connected to the statistical classification [10]. LDA is a linear transform with matrix equation according to:

$$\mathbf{y} = \mathbf{W}^T\mathbf{x}. \tag{15}$$

The key task is to find $\mathbf{W}$. The projection over this subspace gives the best class separability. The LDA is defined as linear transform in which the between-class variances increase and within-class variances decrease. The problem formulation is as follows:

$$J(\mathbf{W}) = \underset{\mathbf{w}}{argmax}\frac{|\mathbf{W}^T\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_W\mathbf{W}|} \tag{16}$$

and

$$\mathbf{S}_W = \sum_{i=1}^{C}\mathbf{S}_i, \tag{17}$$

where $\mathbf{S}_W$ is the within-class scatter matrix and $\mathbf{S}_i = \sum_{\mathbf{x}\in C_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T$. The matrix of between-class variances is $\mathbf{S}_B = \sum_{i=1}^{C} n_i(\mathbf{m}_i-\mathbf{m})(\mathbf{m}_i-\mathbf{m})^T$, where $\mathbf{m}$ is the mean of all vectors, $\mathbf{m}_i$ is the mean for the *i*-th class and n_i is the number of feature vectors in *i*-th class. The solution is:

$$\mathbf{S}_B\mathbf{w}_i = \lambda_i\mathbf{S}_W\mathbf{w}_i. \tag{18}$$

where λ_i are the eigenvalues. After determining λ_i the last step is to calculate the eigenvectors $\mathbf{w}_i$.

Over the extracted PCA features we apply LDA in order to obtain further dimensionality reduction and to improve the class discriminativity.

D. Classification Based on RBFNN

In this phase of our method, we perform classification of the transformed data. After the projection with PCA and LDA, in the corresponding subspace there are the vectors $\mathbf{Y}_i^j$ for each of the classes from the training set. Each of them is obtained via projecting the training input ECG sequence $\mathbf{X}_i$ for each person onto the feature space:

$$\mathbf{Y}_i^k = \mathbf{P}^T\left(\mathbf{X}_i^k - \hat{\boldsymbol{\mu}}\right); i = 1,...,M; k = 1,...,C, \tag{19}$$

where M is the number of training ECG sequences, C is the number of classes from the training set, and $\hat{\boldsymbol{\mu}}$ is the global average of all training sequences.

The method for ECG feature recognition that we propose involves the usage of Radial Basis Function Neural Networks (RBFNN). The RBFNN are a widely used type of neural networks. It differs from the ordinary MLP by the output function of each neuron and the purposes of application. The typical RBFNN contains: a set of input nodes; hidden neuron layer in which each neuron has a special type of output radial basis function, centered at the mean vector of a cluster or sub-cluster in feature space; output layer in which the outputs of the hidden neurons are summed using a linear output function.

The combined classification that we propose can increase accuracy using two parallel modules (distributed system). A distributed system is where two or more modules work in a parallel manner so that the error is distributed. The reliability of a distributed system R_s is [16, 11]:

$$R_s = 1 - \prod_{i=1}^{2} (1 - R_i), \quad (20)$$

where R_i is the score from each classifier, i =1, 2 and $R_s, R_i \in [0,...,1]$.

The result is correct if the following condition is fulfilled: $R_s \geq max(R_i), i = 1...2$. The R_1 classifier is trained with and classifies the PCA-projections of the original ECG sequences, R_2 classifier works respectively with the LDA-projections of the PCA-coefficients. Both classifiers R_1 and R_2 are realized with neural networks with radial basis functions.

III. Experimental Results

A. Segmentation

The segmentation performance is evaluated by using synthetic ECG signal composed from Gaussians with random parameters. This approach ensures that ECG signal is noise free and eliminates the subjective criterion in terms to define onset and offset in ECG waves. Finally, the model performance is evaluated by segmenting the same synthetic signal corrupted with additive white Gaussian Noise (AWGN) (Fig. 4). The AWGN is often used to model the muscle artifacts (EMG noise).

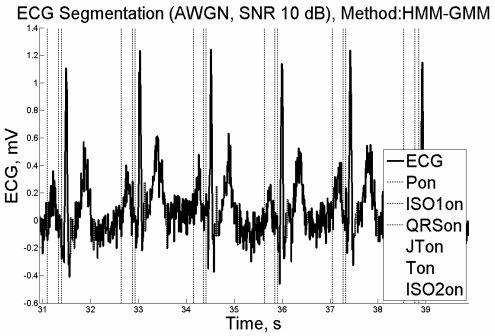


Fig. 4. ECG segmentation result with HMM-GMM and AWGN influence (SNR=10 dB).

As can be seen from Fig. 5 the comparative analysis for ECG segmentation performance confirms that HMM-GMM (8 components) outperforms HMM-SGM and CRF methods.

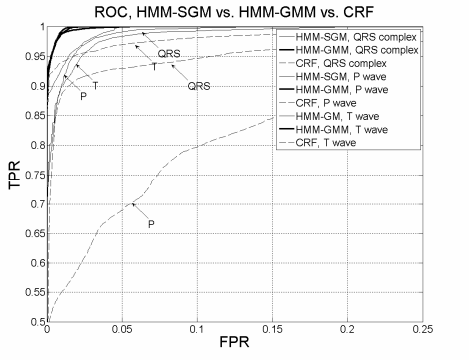


Fig. 5. Segmentation performance comparison between HMM-GMM (8 components), HMM-SGM and CRF.

The ECG segmentation performance for HMM-GMM is additionally evaluated by using mean value μ_e and standard deviation σ_e of the absolute error between expert annotated and automatically detected ECG characteristic point (onset and offset of a given ECG wave or complex). The segmentation performance results for noise free synthetic ECG and synthetic ECG with AWGN can be seen in Table I.

Table I

	P_{on}	$ISO1_{on}$	QRS_{on}	T_{on}
	μ_e, ms σ_e, ms	μ_e, ms σ_e, ms	μ_e, ms σ_e, ms	μ_e, ms σ_e, ms
Noise free ECG	-3.8 6.1	5.8 4.2	-11.1 5.1	-6.9 7.2
ECG with AWGN (SNR = 10 dB)	7.5 48.1	-4.8 91.6	-16.2 7.1	4.4 12.3

B. Feature Space Projection

In our experiments we have used real-world ECG data from 9 persons, which is acquired with industry-standard ECG hardware, equipped with PC-interface. The ECG-cycles projections for two different persons form clusters in PCA subspace which can be seen in Fig. 6.

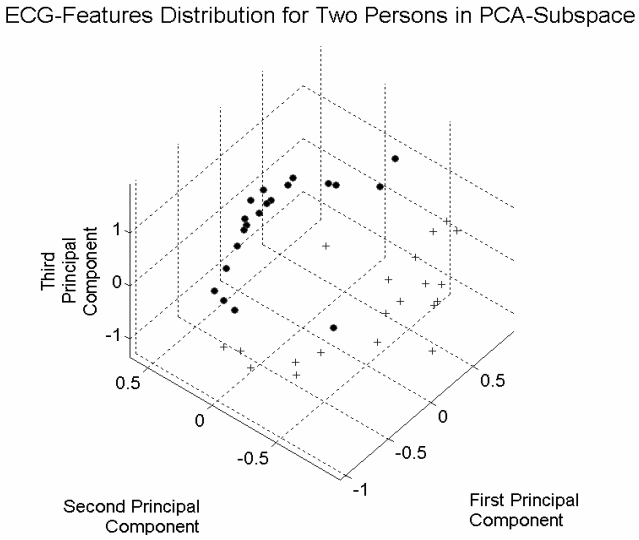


Fig. 6. ECG-Features Distribution for two persons (o and +) according to the first 3 principal components in PCA subspace.

C. Classification

In the classification experiments we have used ECG training set from 9 persons with 20 complete cardiac cycles for each. The RBFNN-based classifier has been tested with other test ECG records acquired from the same persons at different time. In Table II can be seen the recognition rate for PCA features. Also, this table shows the classification results when LDA is applied on PCA coefficients and the dimensionality of the features is reduced from 20 down to 9.

Table II

	PCA	LDA
Person	Recognition rate, %	Recognition rate, %
1	94	94
2	93	93
3	86	83
4	94	90
5	93	74
6	72	82
7	87	63
8	94	94
9	62	77

IV. CONCLUSION

In this work we have presented a new approach for personal biometric identification based on ECG features classification. We denoise the input signal and segment it into sequences. After that we perform dimensionality reduction of the input sequence and extract the PCA features. The separability of the features is improved by LDA. Then we apply RBFNN classifiers for recognition.

The future development of the method will involve a combination with another biometric modality, such as face or fingerprint. We will also research other methods for classification.

ACKNOWLEDGMENT

This work was sponsored by the Ministry of Education and Science of Bulgaria, NSF Grant ВУ-ТН-202/2006 – "Multimodal biometric analysis: methods and algorithms".

REFERENCES

[1] R.Andreão, J. Boudy, "Combining Wavelet Transform and Hidden Markov Models for ECG Segmentation", *EURASIP Journal on Advances in Signal Processing*, vol. 2007 (1), pp. 95-95, 2007.

[2] L. Biel, O. Pettersson, L. Philipson, P. Wide, "ECG Analysis: A New Approach in Human Identification", *IEEE Transactions on Instrumentation and Measurement*, vol. 50 (3), pp. 308-312, 2001.

[3] F.Castells, "Principal Component Analysis in ECG Signal Processing", *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[4] G. D. Forney, "The Viterbi algorithm", *Proceedings of the IEEE*, vol. 61, pp. 268-278, 1973.

[5] K. Grauer, "A Practical Guide to ECG Interpretation", *Elsevier Health Sciences*, Oxford, UK, 1998.

[6] N.P. Hughes, L. Tarassenko, S. J. Roberts, "Markov Models for Automated ECG Interval Analysis", *In Advances in Neural Information Processing Systems*, vol. 16, NIPS, 2003.

[7] A.K. Jain, A. Ross, "Multibiometric Systems", *Appeared in Communication of the ACM*, Special Issue on Multimodal Interfaces, vol. 47 (1), pp. 34-40, 2004.

[8] A.K. Jain , A. Ross, S. Prabhakar, "An Introduction to Biometric Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14 (1), pp. 4–20, 2004.

[9] S. Mallat, "A Wavelet Tour of Signal Processing", *Academic Press*, 2nd edition, 1999.

[10] A.M. Martinez, A. C. Kak, "PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 228 – 233, 2001.

[11] N. Poh, S. Bengio, J. Korczak, "A multi-sample multi-source model for biometric authentication", *Proceedings of the 2002 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 375-384, 2002.

[12] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, vol. 77 (2), pp. 257-286, 1989.

[13] A. Reisiner, G. D. Clifford, R. G. Mark, "Advanced Methods and Tools for ECG Data Analysis", *Artech House*, Boston, 2006.

[14] Y.N. Singh, P. Gupta, "ECG to Individual Identification", *2nd IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS 2008)*, pp. 1-8, 2008.

[15] C. Sutton, A. McCallum, "An Introduction to Conditional Random Fields for Relational Learning", *Introduction to Statistical Relational Learning*, MIT Press, 2007.

[16] N. Ueda, "Optimal linear combination of neural networks for improving classification performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22 (2), pp. 207-215, 2000.

[17] Y. Wang, F. Agrafioti, D. Hatzinakos, K. N. Plataniotis, "Analysis of Human Electrocardiogram for Biometric Recognition", *Journal on Advances in Signal Processing (EURASIP)*, vol. 2008 (1), 2008.

[18] Yogendra Narain Singh, P. Gupta, "Biometrics Method for Human Identification Using Electrocardiogram", *ICB 2009*, *LNCS 5558*, pp. 1270-1279, 2009.

[19] Zhaomin Zhang, Daming Wei, "A New ECG Identification Method Using Bayes' Theorem", *2006 IEEE Region 10 Conference TENCON 2006*, pp. 1-4, 2006.