



Ain Shams University
Faculty of Computer & information Sciences

A Computational Intelligent Technique for Biometric Recognition of Electrocardiograms (ECG)

A Thesis Submitted to
Department of Scientific Computing
In partial fulfillment of the requirements for
The PhD degree

By
Manal Mohsen Mohamed Tantawi
Assistant Lecturer
Scientific Computing Department
Faculty of Computer and Information Sciences
Ain Shams University

Under Supervision of
Prof. Dr. Mohamed Fahmy Tolba
Professor of Scientific Computing
Faculty of Computer and Information Sciences
Ain Shams University

Prof. Dr. Abdel-Badeeh M. Salem
Professor of Computer Science
Faculty of Computer and Information Sciences
Ain Shams University

Dr. Kenneth Revett
Associate Professor in Computer Science Department
Faculty of Informatics and Computer Science
British University in Egypt

August 2014

ACKNOWLEDGMENTS

First, thanks to God, the most beneficent and the merciful. Second, I wish to express my sincere thanks and appreciations to Prof. Dr.Mohamed Fahmy Tolba, Prof. Dr. Abdel-Badeeh Salem and Dr. Kenneth Revett for the continuous support and valuable comments for this thesis. Lastly, I cannot express my feelings toward my family and friends who supported me all over through my way and backed me up during my studies wishing me good luck throughout the preparation of this thesis.

TABLE OF CONTENT

LIST OF FIGURES.....	vi
LIST OF TABLES.....	ix
LIST OF ABBREVIATIONS.....	xi
ABSTRACT.....	xiii
1. INTRODUCTION.....	1
1.1 Thesis Motivations.....	2
1.2 Thesis Objectives.....	4
1.3 Thesis Achievements.....	5
1.4 Thesis Organization.....	6
2. BIOMETRIC FUNDAMENTALS.....	8
2.1 Biometric System.....	9
2.2 Biometric System Evaluation.....	11
2.3 Physiological Biometrics.....	13
2.3.1 Eye Biometrics.....	13
2.3.2 DNA Biometrics.....	14
2.3.3 Face Biometrics.....	15
2.3.4 Hand Biometrics.....	15
2.4 Behavioral Biometrics.....	18
2.4.1 Signature Biometrics.....	18
2.4.2 Voice Biometrics.....	19
2.4.3 Other Behavioral Biometrics.....	19
2.5 Evaluation of Existing Biometrics.....	20
2.6 Multimodal biometrics	22
2.6.1 Modes of Operation.....	23
2.6.2 Level of fusion.....	23
2.6.3 Different Scenarios for Multimodal Systems.....	25
2.6.4 Challenges for Multimodal Systems.....	26
2.7 Biometric Applications.....	27
3. ELECTROCARDIOGRAM (ECG) FUNDAMENTALS.....	30
3.1 Heart anatomy.....	31
3.2 Flow of Blood.....	31
3.3 Electrical Activity of the Heart.....	33
3.4 Electrocardiogram (ECG).....	34
3.5 ECG as a Biometrics: Motivation and challenges.....	35
3.5.1 Motivation.....	36
3.5.2 Challenges.....	36
4. ECG BASED BIOMETRIC SYSTEMS.....	38
4.1 ECG based Biometric System Architecture.....	39
4.2 Fiducial based Systems.....	40
4.3 Non-Fiducial based Systems.....	44
4.4 Combined Fiducial and Non-Fiducial Systems.....	50
4.5 Fiducial versus Non Fiducial Systems.....	52

5. DATA SOURCES AND THE IMPLEMENTED TECHNIQUES.....	55
5.1 Data Sources.....	56
5.2 Preprocessing.....	57
5.2.1 Noise Removal.....	57
5.2.2 Techniques for Detection of Fiducial Points.....	58
5.3 Feature Extraction Techniques.....	63
5.3.1 Wavelets.....	63
5.3.2 Auto-correlation\Discrete cosine Transform (AC\DCT).....	68
5.4 Feature Reduction.....	70
5.4.1 Dimension Reduction techniques.....	70
5.4.2 Feature Selection Techniques.....	75
5.5 Classification Techniques.....	78
5.5.1 Multilayer Perceptron (MLP).....	78
5.5.2 Radial Basis Functions Neural Network (RBF).....	80
6 THE PROPOSED FIDUCIAL APPROACH FOR AN ECG BASED BIOMETRIC IDENTIFICATION.....	84
6.1 Stages of the Fiducial Approach.....	85
6.2 The Super Set.....	87
6.3 The PCA Reduced Set.....	89
6.4 The LDA Reduced Set.....	91
6.5 The RS Reduced Set.....	92
6.6 The IG Reduced Set.....	93
6.7 The Proposed PV Set.....	94
6.8 Performance Analysis and Evaluation.....	96
7. THE PROPOSED NON-FIDUCIAL APPROACH FOR AN ECG BASED BIOMETRIC IDENTIFICATION.....	101
7.1 Stages of the Non-Fiducial Approach.....	102
7.2 Results for RR Intervals.....	104
7.3 Results for QT and QRS Intervals.....	109
7.4 The Implementation of AC\DCT Approach.....	113
7.5 Performance Analysis and Evaluation.....	116
8. MINIMIZING THE IMPACT OF HEART RATE VARIABILITY FOR ECG BASED BIOMETRIC IDENTIFICATION.....	118
8.1 The Impact of HRV.....	119
8.2 Modifying the Fiducial Approach.....	120
8.2.1 The IG Set.....	121
8.2.2 The PV Set.....	123
8.3 Modifying the Non-Fiducial Approach.....	124
8.4 Performance Analysis and Evaluation.....	127
8.4.1 The Impact of HRV on Accuracy.....	127
8.4.2 Fiducial versus Non Fiducial ECG based Biometrics.....	129
9. CONCLUSIONS AND FUTURE WORK.....	135
9.1 Conclusions.....	136
9.2 Future Work.....	138

REFERENCES.....	139
LIST OF PUBLICATIONS.....	143
Appendix A: Error Back Propagation.....	145
Appendix B: The Orthogonal Least Squares (OLS).....	149

LIST OF FIGURES

Figure 2.1 A biometric System: Enrollment mode.....	9
Figure 2.2 A Biometric System: verification mode.....	10
Figure 2.3 A Biometric System: identification mode.....	11
Figure 2.4 Categories of Physiological Biometrics.....	13
Figure 2.5 Image of Iris [6].....	14
Figure 2.6 Image of Retina [6].....	14
Figure 2.7 Image of Finger print [6].....	16
Figure 2.8 Hand geometry scanner, acquired image of hand and the extracted geometric features [4, 6].....	17
Figure 2.9 Image of Palm print.....	18
Figure 2.10 Different Behavioral Biometrics.....	18
Figure 2.11 An image of an individual's signature [6].....	19
Figure 2.12 An Example of a Multimodal biometric system [5].....	23
Figure 2.13 A bimodal biometric system showing the three levels of fusion (FU: fusion module, MM: matching module, DM: decision module) [7].....	24
Figure 2.14 Various scenarios in a multimodal biometric system [5].....	26
Figure 3.1 The heart anatomy.....	31
Figure 3.2 A heart diagram that illustrates the flow of blood in and out of the heart.....	32
Figure 3.3 The electrical activity of the heart.....	33
Figure 3.4 An ECG trace and a magnified labeled heartbeat.....	34
Figure 3.5 The interpretation for the three complexes of the ECG heartbeat.....	35
Figure 4.1 General ECG based system architecture.....	39
Figure 4.2 The fiducial points extracted from each ECG heartbeat.....	40
Figure 5.1 Reduction of noise a) an original ECG segment and b) the ECG segment after filtering.....	58
Figure 5.2 The R peak detection algorithm steps for a normal ECG segment a) original signal b) after successive low & high pass filtering. c) Differentiated signal. d) Output of squaring process. e) Results of moving-window integration. f) position of R peaks after the thresholding step and delay removal [57].....	59
Figure 5.3 The lower trace is the QRS signal and upper trace is the resulting envelope [58].....	60
Figure 5.4 Window locations a) for P wave and b) for T wave.....	61
Figure 5.5 'wings' functions a) for negative T-wave, b) for positive T-wave [32].....	62
Figure 5.6 Radius of Curvature and the resulting computation of curvature [10].....	62
Figure 5.7 Comparing Sine Wave and Wavelet [60].....	64
Figure 5.8 Multi-Resolution Wavelet Decomposition. h = low-pass decomposition filter; g = high-pass decomposition filter; $\downarrow 2$ = down-sampling operation. $A_1(t)$, $A_2(t)$ are the approximated coefficient of the original signal at levels 1, 2 etc. $D_1(t)$, $D_2(t)$ are the detailed coefficient at levels 1, 2 [60].....	66
Figure 5.9 Multi-Resolution Wavelet Reconstruction. h' = low pass reconstruction filter; g' = high pass reconstruction filter; $\uparrow 2$ = up-sampling operation. $A_1(t)'$, $A_2(t)'$ are the processed or non-processed approximated coefficient of the original signal at levels 1, 2 etc. $D_1(t)'$, $D_2(t)'$ are the processed or non processed detailed coefficient at levels 1, 2 [60].	67

Figure 5.10 Haar mother wavelet [61].....	67
Figure 5.11 the Daubechies wavelet family [61].....	68
Figure 5.12 Six wavelets from the biorthogonal family [61].....	68
Figure 5.13 MLP with one hidden layer.....	79
Figure 6.1 General block diagram for the proposed fiducial approach.....	85
Figure 6.2: The super set of 36 features.....	88
Figure 6.3 The ROC curve for the superset.....	89
Figure 6.4 The HB accuracy achieved by Test set 1 (upper solid line) and Test set2 (lower dashed line) as a function of number of principle components selected.....	90
Figure 6.5 The ROC curve for the PCA set.....	90
Figure 6.6 The HR accuracy achieved by Test set 1 (upper solid line) and Test set2 (lower dashed line) as a function of number of linear discriminant components selected.....	91
Figure 6.7 The ROC curve for the LDA set.....	92
Figure 6.8 The ROC for the RS set.....	93
Figure 6.9 The ROC for the IG set.....	94
Figure 6.10 The ROC for the PV set.....	96
Figure 6.11 The HR accuracy for the two test sets achieved by the super set and its four reduced versions.....	98
Figure 6.12 The ROC curves of the super set and its reduced versions.....	98
Figure 7.1 General block diagram for the proposed non-fiducial approach.....	103
Figure 7.2 An example for an RR cycle.....	103
Figure 7.3 The considered RR, QT and QRS intervals from the ECG trace for wavelet decomposition.	104
Figure 7.4 The discrete wavelet decomposition levels and the formation of the resulted wavelet coefficient structure (190 coefficients from 6 parts).....	105
Figure 7.5 Decomposition of RR cycles (a) an original averaged RR cycle and (b) its wavelet coefficient structure (the bounds of each part are defined by dashed vertical lines).....	105
Figure 7.6 The mean of the HR accuracy for each of the resulted reduced set from the first phase (a) test set 1 and (b) test set 2.....	106
Figure 7.7 The mean of the HR accuracy for each of the resulted reduced set from the first phase for (a) PTB and (b) Fantasia.....	108
Figure 7.8 The ROC for the proposed wavelet method with the RR intervals.....	109
Figure 7.9 The achieved SI and HR accuracies for test set 2 with a) RR intervals, b) QT intervals and c) QRS intervals.....	112
Figure 7.10 The ROC curves for the proposed method with the three considered ECG heartbeat representations: QRS, QT and RR.....	113
Figure 7.11 General block diagram for the proposed AC\DCT approach.....	114
Figure 7.12 The AC\DCT algorithm a) 200 AC Coefficients of an ECG window and b) their corresponding DCT coefficients.....	114
Figure 7.13 The ROC for the AC\DCT method.....	116
Figure 7.14 The ROC curves for the proposed method with RR representation and the AC\DCT methods.....	117
Figure 8.1 The impact of QT correction a) 10 heartbeats belong to the same subject are normalized to the same length, b) the same heartbeats after QT correction.	119
Figure 8.2 The ROC for IG set with and without QT correction.....	122

Figure 8.3 The ROC for PV set with and without QT correction.	124
Figure 8.4 The heartbeat representations a) a RR interval, b) its corresponding QQ interval	124
Figure 8.5 The impact of QT correction on QQ intervals a) 10 Q-Q intervals for the same subject interpolated to the same length b) shows the same QQ intervals but with QT interval correction.....	125
Figure 8.6 The discrete wavelet decomposition levels and the formation of the resulted wavelet coefficient structure (201 coefficients from 6 parts).....	125
Figure 8.7 The corresponding wavelet structure for the QQ interval in figure 8.4 (b).....	126
Figure 8.8 The ROC for the modified non-fiducial approach with & without correction.....	127
Figure 8.9 The SI accuracy versus the percentage of heart rate variability. The solid (dash) line represents the SI accuracy when the duration of the RR intervals decreased (increased) by 0-50% from the reference point.....	129
Figure 8.10 The HR accuracy achieved for the two test sets with the considered modified feature sets.....	130
Figure 8.11 The ROC curves for the considered modified feature sets.....	130
Figure 8.12 The ROC curves achieved by the three features sets using PTB database a) with RBF and b) with MLP classifiers.....	132
Figure 8.13 The ROC curves achieved by the three features sets using our data set a) with RBF and b) with MLP classifiers.....	134

LIST OF TABLES

Table 2.1 Comparison of various biometric technologies based provided by [4, 5]. High, Medium, and Low are denoted by H, M, and L, respectively.....	21
Table 2.2 Evaluation of some of the biometric traits according to EER, FAR and FRR measures [6].....	22
Table 4.1 A summary of the key existing fiducial based systems discussed in this study.....	43
Table 4.2 A summary of the key existing non-fiducial based systems discussed in this study.....	49
Table 4.3 A summary of the key existing combined fiducial and non-fiducial systems discussed in this study.....	52
Table 4.4 The consideration of the mentioned critical issues of ECG by the existing systems in the literature.....	54
Table 6.1 The labels for the 36 fiducial features of the super set.....	88
Table 6.2 The SI and HR results of the super set.....	89
Table 6.3 The SI and HR results of the PCA set.....	90
Table 6.4 The SI and HR results of the LDA set.....	91
Table 6.5 The ranked list of features selected RS.....	92
Table 6.6 The SI and HR results of the RS set.....	93
Table 6.7 The ranked list of features selected IG.....	94
Table 6.8 The SI and HR results of the IG set.....	94
Table 6.9 The list of features belongs to the PV set.....	95
Table 6.10 The SI and HR results of the PV set.....	95
Table 6.11 A Summary of the mean 'm' and standard deviation 's' of the error in the location for each of the peaks and valleys points as a result of the level of noise added to each of the four generated noisy versions of test set 1. Results are given in milliseconds in the form of $m \pm s$	99
Table 6.12 A Summary of the mean 'm' and standard deviation 's' of the error for each of the onsets and offsets caused by noise added to each of the four generated noisy versions of test set 1. Results are given in milliseconds in the form of $m \pm s$	99
Table 6.13 A Summary of the SI and HR accuracies achieved for the four noisy versions of test set1 by the IG and PV sets. (m/n) means m of n subjects were successfully identified.....	100
Table 7.1 The achieved SI and HR accuracies.....	107
Table 7.2 The SI and HR achieved by the final reduced set of 66 coefficients.....	109
Table 7.3 The achieved SI and HR accuracies using QT intervals.....	110
Table 7.4 The achieved SI and HR accuracies using QRS intervals.....	111
Table 7.5 A summary of the achieved SI and HR accuracies using the proposed wavelet method and the AC\DCT method.....	117
Table 8.1 The ranked list of features selected by IG with different correction formulas and common features are in bold.....	121
Table 8.2 The SI and HR results of the IG set with & without QT correction.....	122
Table 8.3 The SI and HR results of the PV set.....	123
Table 8.4 The SI and HR results achieved by the modified non-fiducial approach with and	

without correction.....	127
Table 8.5 A summary of the SI and HR accuracies for the three test sets along with the optimum parameters for the RBF classifier with each feature set.....	131
Table 8.6 A summary of the SI and HR accuracies for the three test sets along with the optimum parameters for the MLP classifier with each feature set.....	131
Table 8.7 A summary of the SI and HR accuracies for our data test set along with the optimum parameters for the RBF classifier with each feature set.....	133
Table 8.8 A summary of the SI and HR accuracies for our data test set along with the optimum parameters for the MLP classifier with each feature set.....	133

LIST OF ABBREVIATIONS

AC	Auto-Correlation
AFIS	Automatic Fingerprint Identification System
APC	Atrial Premature Contraction
AR	Auto-Regression
BIR	Biometric Information Record
BMP	Beat per Minute
CWT	Continuous Wavelet Transform
DBSS	Decision Based Support System
DCT	Discrete Cosine Transform
DET	Detection Error Tradeoff
DNA	Deoxyribo Nucleic Acid
DWT	Discrete\Dyadic Wavelet Transform
ECG	Electrocardiogram
EEG	Electroencephalogram
EER	Equal Error Rate
EMD	Empirical Mode Decomposition
EOG	Electrooculogram
FAR	False Acceptance Rate
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Rejection Rate
GMM	Gaussian Mixture Model
GSV	GMM Super Vectors
HMM	Hidden Markov Model
HPE	Hermite Polynomial Expansion
HR	Heartbeat Recognition
HRV	Heart Rate Variability
HTK	HMM ToolKit
IMF	Intrinsic Mode Functions
IG	Information Gain
KLT	Karhunen Loeve Transform
KNN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LS	Linear Squares
MLP	Multilayer Perceptron
MRA	Multi-Resolution Analysis
MSE	Mean Square Error
NN	Neural Network
OLS	Orthogonal Least Squares
PASH	Parameterized Averaged Support Heuristic
PCA	Principle Component Analysis
PDA	Personal Digital Assistant

PDM	Polynomial Distance Measure
PSD	Power Spectral Density
PV	Peaks & Valleys
PVC	Premature Ventricular Contraction
RBF	Radial Basis Function
ROC	Relative Operating Characteristic
RS	Rough Sets
SFA	Simplified Fuzzy ARTMap
SI	Subject Identification
SSE	Sum Squared Error
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
WR	Window Recognition
WSA	Welch Spectral Analysis

ABSTRACT

Biometric systems have become integrated into the fabric of everyday life – deployed where and whenever secure access to a trusted instrument is required. Since the discovery of fingerprints over 100 years ago, a variety of approaches to person identification\verification have been devised. Recently, electrocardiogram (ECG) has been introduced as a new biometric trait. It distinguishes itself by being a liveliness indicator without any further processing and a difficult to falsify biometric trait.

The existing ECG based biometric systems can be generally categorized according to the utilized features to fiducial and non-fiducial systems. The derivation of fiducial features significantly relies on the accuracy of detecting 11 fiducial points, which is a very challenging task by itself. On the other hand, non-fiducial approaches relax the detection process to include only the sharpest point or sometimes no fiducial detection is needed. However, they usually result in high dimension feature space.

This work presents a systematic study that contributes to ECG based individual identification. A fiducial based approach that utilizes a super set of features is first introduced. Reduction of this set has been investigated using different reduction techniques like principle component analysis (PCA), linear discriminant analysis (LDA), rough sets (RS) and information gain (IG). Furthermore, in order to relax the fiducial detection process, another fiducial feature set namely PV set that is derived from only five peaks and valleys fiducial points has been also introduced. The results showed that the IG overwhelms the other considered reduction techniques, while the PV set preserves the Subject Identification (SI) accuracy with slight decrease in the Heartbeat Recognition (HR) accuracy.

Moreover, a non-fiducial wavelet based approach is proposed. To avoid the high dimensionality of the resultant wavelet coefficient structure, the structure has been investigated using a proposed two-phase reduction process which has resulted in excluding roughly 65% of the structure. In addition, the proposed non-fiducial approach has been applied to different heartbeat representations such as RR, QT, QRS intervals. The results revealed the deficiencies of utilizing QT and QRS intervals only for representing heartbeats.

In addition, a QT correction stage is introduced before feature extraction for both approaches to resolve the impact of heart rate variability (HRV). Finally, the proposed fiducial and non-fiducial approaches have been examined by both Radial Basis Functions (RBF) and Multilayer

Perceptron (MLP) neural network classifiers on the basis of stability over time, rejection of intruders, generalization and scalability issues.

Experimentation was conducted using Physionet databases and our own established datasets. The results revealed the significance of the suggested QT correction stage and the superiority of the proposed non-fiducial wavelet approach with respect to all the considered critical issues. It has achieved 100% SI accuracy for all considered datasets and approximately 100% and 96% HR accuracy for Physionet databases and our collected dataset respectively, in addition, to the best ROC curve in comparison with the fiducial approach.

CHAPTER 1

INTRODUCTION