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

LIST OF TABLES

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	Karhumen 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