Computer Science Department
Faculty of Computer & Information Sciences
Ain Shams University

Biometric Intelligent System based on Heart Signals

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences.

By

Mahmoud Mohamed Bassiouni EL-Sayed

Demonstrator at Computer Science Department,
Faculty of Computer and Information Sciences,
Egyptian E-learning University

Under Supervision of

Prof. Dr. Abdel-Badeeh Mohamed Salem

Professor of Computer Science,
Faculty of Computer and Information Sciences,
Ain Shams University

Prof. Dr. El-Sayed A. El-Dahshan

Professor of Scientific Computing,
Faculty of Sciences-Physics Department,
Ain Shams University

Dr. Wael Hamdy Khalifa

Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

December 2016

Acknowledgment

I would like express my gratitude and a special thanks to my supervisor Prof. Dr. Abdel-Badeeh Mohamed Salem for his guidance, advice and constant support throughout my thesis. I would like to thank him for being my advisor. I want to express my gratitude to Prof. Dr. El-Sayed A. El-dahshan for his full support, encouragement, knowledge, teachings and his expected help. I want to thank Dr. Wael khalefa for his support in my thesis. I want to thank my friends and my workmates for all the thoughtful and motivating discussions. I am especially grateful to my parents for their love and support and would like to thank my parents for raising me in a way to believe that I can achieve anything in life with hard work and patience.

Abstract

Recent reported research proved that Electrocardiogram (ECG) and Phonocardiogram (PCG) can be used as a biometric. In this work, we built and developed two different biometric identification systems one for (ECG) and the other for (PCG).

In the first system, we presented an ECG divided mainly into four stages namely, data acquisition, preprocessing, feature extraction and classification. First stage, data acquisition stage, data sets were collected from two different databases, ECG-ID and MIT-BIH Arrhythmia database. Second stage noise reduction of ECG signals using wavelet transform and series of filters used for de-noising. Third stage obtains the features using three different techniques a non-fiducial, fiducial and a fusion approach between them. In the last stage, the classification stage, three classifiers have been developed to classify subjects. The first classifier is based on Artificial Neural Network (ANN), the second classifier is based on Euclidean distance (ED) and the last classifier is sequential minimal optimization (SMO) algorithm for training a support vector machine (SVM) using polynomial kernel classifier. Classification accuracy of 95% for ANN, 99 % for ED and 99% for SVM on the ECG-ID database, while 100% for ANN, ED, SVM on MIT-BIH database.

In the second system we presented a machine learning approach based on feature level fusion for person identification using phonocardiogram (PCG). The proposed approach consists of five stages; starting with data acquisition, pre-processing, segmentation, feature extraction, and classification. Firstly data set were collected from the HSCT-11 database working on 60 subjects. Secondly process is concerned with noise reduction by removing noise from PCG signal

using wavelets. Thirdly, segmentation process was done by applying Shannon energy envelope on the filtered signal to detect the positions of S1 and S2. Then a new combination of features is investigated for robust biometric PCG identification. Mel Frequency Cepstral Coefficients (MFCC) is efficient for PCG identification in clean speech while Wavelet features are robust for noisy environments. Therefore, combining both features together is better than taking each one individually the fusion is done using canonical correlation analysis (CCA). Finally, artificial neural network (ANN) and sequential minimal optimization (SMO) algorithm for training a support vector machine (SVM) using polynomial kernel classifier has been applied to classify subjects. The result shows a classification ratio 98.33%.

The results obtained from those two systems (ECG) and (PCG) were encouraging to show how robust our machine learning techniques used are. A comparison is already made with the previous work to ensure the efficiency of our methods.

List of Publications

1- BASSIOUNI M., KHALEFA W., El-DAHSHAN E.S.A., SALEM A.B.M. ,A Machine Learning Technique for Person Identification using ECG Signals, International journal of Applied physics, Vol 1 2016 (pp. 37-41).

2- Bassiouni M., Khalefa W., El-Dahshan E.S.A. and Salem A.B.M., A study on the Intelligent Techniques of the ECG-based Biometric Systems. *RECENT ADVANCES in ELECTRICAL ENGINEERING*, 2015, (pp.26-31).

3- Bassiouni, M., Khalifa, W., El Dahshan, E.S.A. and Salam, A.B.M., A study on PCG as a biometric approach. In *2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, 2015, December, (pp. 161-166).

4- Bassiouni, M., Khalifa, W., El Dahshan, E.S.A. and Salam, A.B.M., An Intelligent Approach for Person Identification Using Phonocardiogram Signals, International Journal of Applications of Fuzzy Sets and Artificial Intelligence (ISSN 2241-1240), Vol. 6, 2016, (pp.103 -117).

Table of Contents

List of Figures

List of Tables

`

List of Algorithms

List of Abbreviations

ANN	**Artificial Neural Network**
AC	**Auto Correlation**
BP	**Back Propagation**
BFCC	**Bark Frequency Cepstral Coefficient**
CRR	**Correct Recognition Rate**
CCA	**Canonical Correlation Analysis**
CCORR	**Correlation Coefficient Recognition Rate**
DNA	**Deoxyribo Nucleic acid**
DTW	**Dynamic Time Wrapping**
DCT	**Discrete Cosine Transform**
DWT	**Discrete Wavelet Transform**
DNN-BT	**Deep Neural Network Breaking Tires**
DBNN	**Decision Based Neural Network**
ECG	**Electrocardiogram**
ED	**Euclidean Distance**
ECG-ID	**Electrocardiogram – Identification**
FFT	**Fast Fourier Transform**
FPR	**False Positive Rate**
GMM	**Gaussian Mixture Models**
HHT	**Huang Hilbert Transform**
HMM	**Hidden Marchov Models**
HSCT-11	**Heart Sound Catania 2011**
ISOM	**Increment Self Organization Map**
KLDA	**Kernel Linear Discernment Analysis**
KNN	**K-Nearest Neighbor**
LBFC	**Linear Band Frequency Cepstra**
LPCC	**Linear Predication Cepstral Coefficient**
LDA	**Linear Discernment Analysis**
MIT-BIH	**Massachusetts Institute of Technology**
MSE	**Mean Square Error**
MSDWT	**Multi Scale Discrete Wavelet Transform**
MLP	**Multi-Layer Perceptron**
MVC	**Majority Vote Classifier**
MFCC	**Mel Frequency Cepstral Coefficient**
NN	**Neural Network**
PCG	**Phonocardiogram**
PBNN	**Probability Based Neural Network**
PNN	**Probabilistic Discernment Analysis**
PSA	**Power Spectral Analysis**
PRD	**Percent Residual Difference**
PSA	**Power Spectral Analysis**
PTB	**Physikalisch Technische Bundesanstalt**
QNN	**Quantum Neural Network**
QP	**Quadratic Programming**
PCA	**Principle Component Analysis**
RBFNN	**Radial Bias Function Neural Network**
RS	**Rough Set**
SMO	**Sequential Minimal Optimization**
SVM	**Support Vector Machine**
SACA	**Statistical and Coherence Analysis**
SOM	**Self-Organization Map**
STDFT	**Short Time Discrete Fourier Transform**
TM	**Template Matching**
TC	**Template Correlation**
TPR	**True Positive Rate**
WT	**Wavelet Transform**
WDIST	**Wavelet Distance**
WPT	**Wavelet Packet Transform**
VQ	**Vector Quantization**

Chapter 1

Introduction

Introduction

1.1 Overview

Due to identity fraud, loss of privacy, friendly thief, terrorism the identification security is considered one of the important solutions to these problems. Although computers help human and provide an important services to him but also they does not deliver enough security solutions to identify the individual. Systems based on computer for identification are defined as biometric systems. Those systems are used a lot now a days and in a lot of applications [1].

1.2 Motivations for ECG and PCG

Some people are still reluctant to engage in ecommerce or conduct other network transactions having misgivings about well-founded systems that will protect their privacy and prevent their identity from being stolen or misused. The loss of personal privacy, fraudulent funds transfers, outright theft, and abuse of identity in network transactions leads to identity fraud and identity theft. Identity theft is defined as the personal information is accessed by third parties without obvious authorization from the owner. Identity fraud takes place when a criminal gets an illegally-obtained personal information and uses it for his financial gain. Those personal information may comprise social insurance number, bank or credit card account numbers, passwords, telephone calling card number, birth date, name, address and so on.

Another type of identity theft is known as online identity theft such as phishing, hacking and spyware. Also friendly theft that occurs within friends, family or in-home employees who obtain private data for their personal gain. Due to all those theft problems researchers were thinking and researching for a lot of solutions. This motivated researchers to seek for a dependable and exact substitute solution for identity identification over the traditional password/ ID card based systems and any traditional techniques used [2].

Biometrics, which are physiological or behavioral characteristics extracted from human subjects, have emerged to be a new set of technologies that promise an effective solution for this problem. So they cannot be lost, stolen, forged, or subject to failure. In the past few periods, biometrics has been lengthily used for law execution such as criminal investigations, fatherhood determination, and forensics [3].

On one hand the traditional techniques using username and password, also using keys and identification cards can be stolen and lost and also the traditional biometrics such as face, Signature, DNA, Speech, Fingerprint, face and iris they all can be falsified, forged, fooled and faked. On the other hand PCG and ECG biometric are expected to address some deficiencies thanks to their universality, distinctiveness, permanence, small storage requirements, speed, and their stability over a sufficiently long period of time as well as their uniqueness for individuals. Those signals are difficult to disguise, forged, falsified or faked.

1.3 Objectives

- The Main Aims of the thesis is the analysis of two relatively new medical biometric attributes, the electro-cardiogram (ECG), and the heart sound (PCG),

  The primary objectives of the current research are to address the following issues:

  - **Data collection**: Searching for a dataset recently in ECG and PCG for patients of in different ages, currently searching for a combined database for both ECG and PCG for the same patient.
  - **Design of a robust Identification system based on ECG signal:** To date, two major approaches have been proposed for ECG recognition, namely fiducially-based and non-fiducially-based methods, that achieve good performance in an almost ideal scenario in terms of noise and heart rate.
  - **Design of a robust Identification system based on PCG signal:** Human recognition based on the PCG signals is a new area only lately. Having an acoustic nature, the use of the PCG signal in recognition context is very challenging due to wave characteristic in the presence of noise components.