Feature Selection, Cross-view Matching and Spoof Detection for Electrocardiogram-based Human Recognition

by

Majid Komeili

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

© Copyright 2017 by Majid Komeili
Abstract

Feature Selection, Cross-view Matching and Spoof Detection for Electrocardiogram-based Human Recognition

Majid Komeili
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2017

Biometric technology for identifying individuals has been extensively used all over the world. As this technology becomes increasingly popular, the potential of fooling or spoofing it is a rising concern. For example, fingerprint may be left behind whenever we touch a glass surface like a smart phone and from that some synthetic samples can be created that can be used to fool the biometric system. Iris images can also be captured from a few meters distance, not to mention face images that can be captured from a longer distance. Nowadays biometric spoofing and countermeasures is an active research area and there has been a lot of efforts towards a promising approach to ensure the presence of a real legitimate user.

Medical biometrics such as Electrocardiogram (ECG) have interesting advantages over conventional biometrics. ECG is a vital signal and it naturally provides liveness detection. In addition, compare to conventional biometrics, if not impossible it is far more difficult to steal someone’s ECG. Since ECG can be recorded from fingertips it is a natural choice for fusion with fingerprint. In this research work, we first provide a practical framework for fusion of ECG and fingerprint such that the resulting multimodal biometric system is more robust against novel spoof attacks.

Despite its advantages, ECG comes with its challenges. ECG is a non-stationary signal and factors like emotion, diet and hear rate affect its waveform especially when recorded in different sessions. It becomes more challenging when subjects are enrolled using only one session and tested on another session that could be a few weeks apart. To reduce the effect of such non-stationarity, we propose a feature selection method to select a subset of features that are more stable across different recording sessions.
Furthermore, to cope with the effect of heart rate on performance of ECG recognition, we propose a novel method for cross-view matching where the distance between two samples situated in different views can be computed. In this context view refers to different heart rates where samples in different views can be represented by feature sets that can be different in type and length.
Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Dimitris Hatzinakos, for his invaluable advice, guidance and continuous support of my PhD research. Thank you for being a supportive advisor and a wonderful mentor, and for giving me enough independence and confidence to explore my ideas.

I would also like to express my sincere gratitude to my defense committee members, Professor Ravi Adve, Professor Shahrokh Valaee, Professor Ashish Khisti and Professor Nasir Memon, for all their comments and feedback. I also would like to express my gratitude to my friends and colleagues who helped me go through my PhD study.

I wish to give my wholehearted thanks to my parents, Narjes and Alireza for their unconditional support and encouragement throughout my years of study without which this accomplishment would not have been possible. I am grateful to my sisters, Vahideh and Farzaneh for their care, empathy and love.

Last but not least, I owe my deepest thanks to my loving wife, Narges, for her unconditional love, brilliant ideas, patience, sacrifices, and encouragements. You were always there cheering me up and stood by me throughout the course of my studies.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Autocorrelation</td>
</tr>
<tr>
<td>BEMD</td>
<td>Bivariate Empirical Mode Decomposition</td>
</tr>
<tr>
<td>BSIF</td>
<td>Binarized Statistical Image Features</td>
</tr>
<tr>
<td>CWT</td>
<td>Continuous wavelet transform</td>
</tr>
<tr>
<td>ECG</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>EER</td>
<td>Equal Error Rate</td>
</tr>
<tr>
<td>FpR</td>
<td>Fingerprint Recognition</td>
</tr>
<tr>
<td>FpLD</td>
<td>Fingerprint Liveness Detection</td>
</tr>
<tr>
<td>GMA</td>
<td>Generalized Multiview Analysis</td>
</tr>
<tr>
<td>HC</td>
<td>Heartbeat Consistency</td>
</tr>
<tr>
<td>IMF</td>
<td>Intrinsic Mode Functions</td>
</tr>
<tr>
<td>ICA</td>
<td>Independent Component Analysis</td>
</tr>
<tr>
<td>IQA</td>
<td>Image Quality Assessment</td>
</tr>
<tr>
<td>LBP</td>
<td>Local binary patterns</td>
</tr>
<tr>
<td>LPQ</td>
<td>Local Phase Quantization</td>
</tr>
<tr>
<td>LDA</td>
<td>Linear Discriminant Analysis</td>
</tr>
<tr>
<td>LCPD</td>
<td>Local Contrast Phase Descriptor</td>
</tr>
<tr>
<td>MSFS</td>
<td>Multi-session Feature Selection</td>
</tr>
<tr>
<td>MSFS-RP</td>
<td>Multi-session Feature Selection-Random Projection</td>
</tr>
<tr>
<td>MvDA</td>
<td>Multi-veiw Discriminant Analysis</td>
</tr>
<tr>
<td>MmLD</td>
<td>Multimodal Liveness Detection</td>
</tr>
<tr>
<td>MmHR</td>
<td>Multimodal Human Recognition</td>
</tr>
<tr>
<td>PPG</td>
<td>Photoplethysmogram</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic</td>
</tr>
<tr>
<td>SVM</td>
<td>Support Vector Machine</td>
</tr>
<tr>
<td>STFT</td>
<td>Short-time Fourier Transform</td>
</tr>
<tr>
<td>SURF</td>
<td>Speeded Up Robust Features</td>
</tr>
<tr>
<td>TEOAE</td>
<td>Transient Evoked Otoacoustic Emissions</td>
</tr>
<tr>
<td>WLD</td>
<td>Weber Local Descriptor</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Characteristics of LiveDet2015 [1] datasets used in our experiments. .............. 23

2.2 EER (in percent) of MmLD (fusion of ECG and LCPD) on DigitalPersona dataset compared with 6 state-of-the-art liveness detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing methods. .......................................................... 32

2.3 EER (in percent) of MmLD (fusion of ECG and LCPD) on GreenBit dataset compared with 6 state-of-the-art liveness detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing methods. .......................................................... 33

2.4 EER (in percent) of MmLD (fusion of ECG and LCPD) on Biometrika dataset compared with 6 state-of-the-art liveness detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing methods. .......................................................... 34

2.5 EER (in percent) of MmHR (fusion of ECG and Bozorth) for different fusion rules and different values of HC threshold. .................................................. 36

2.6 Comparison with LivDet2015 liveness detection competition. HTER (in percent) is reported for all participants of LivDet2015 competition from [1]. EER (in percent) for the MmLD (fusion of ECG and LCPD) with template updating for different ECG lengths in the 3rd test session is also presented. To be consistent with LivDet2015 competition, only LiquidEcoflex and RTV are considered as unknown materials. .................................................. 40
2.7 ferrfake (in percent) at ferrlive=1%. Results for 3 participants of LivDet2015 is reported from [2]. ferrfake of the MmLD (fusion of ECG and LCPD) with template updating for different ECG lengths in the 3rd test session is also presented. To be consistent with LivDet2015 competition, only LiquidEcoflex and RTV are considered as unknown materials.

3.1 Across-session ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

3.2 Across-session TEOAE experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

3.3 Across-posture ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

3.4 Comparison of different feature selection algorithms: EER (in percent) and standard deviation (in percent) are reported for three experiments. Standard deviations are presented in parentheses. First and second rows are corresponded to across-session ECG and TEOAE experiments, respectively. Third row is corresponded to across-posture ECG experiment. Last row shows the average EER over the three experiments. Results without feature selection (No FS) is also appended.

3.5 Effect of changing the number of sessions used in feature selection.

3.6 CPU time (in seconds) of the MSFS and MSFS-RP algorithms. EER (in percent) is also presented in parentheses.

3.7 CPU time (in seconds) for our comparison feature selection algorithms. The symbols (*) and (†) denote algorithms are respectively implemented in MATLAB and C. The total time is reported in the last row. For convenience, CPU time of the proposed MSFS-RP method is repeated in the last column.
4.1 ECG recognition across different heart rates. EER (in percent) and standard
deviation (in percent) are reported for different methods. Standard deviations
are presented in parentheses. ........................................... 84

4.2 Near infrared versus visible light face recognition. Recognition rate (in percent)
for different methods are compared. ................................... 85

4.3 Cross pose face recognition. Recognition rate (in percent) for different methods
are compared. \( a_h \) refers to the two additional cameras (08-1 and 19-1) located
above the subject, simulating a typical surveillance camera view. ............... 86
List of Figures

1.1 P, R and T waves of a heart beat. ........................................ 1
1.2 Fake finger can be created using some fairly simple procedures. 3

2.1 Block diagram of the proposed method. ............................... 16
2.2 Our experimental setup and chimeric dataset of ECG and live/spoof fingerprint. 24
2.3 Average recording time versus HC. ........................................ 28
2.4 Recording time for different subjects for HC=0.92. For each subject 3 values corresponding to 3 test sessions are marked with circles. The average recording time is also shown as a solid red line. 29
2.5 ECG recognition performance versus the average recording time. EER of ECG recognition using the proposed stopping criteria HC to determine the length of each session is compared with ECG recognition without HC stopping criteria, i.e. same recording length for all sessions. EER is averaged over all test sessions. 30
2.6 EER of ECG and fingerprint recognition with and without template updating: (a) 2nd test session. (b) 3rd test session. Template updating improves both ECG recognition and fingerprint recognition performance. Results are averaged over 3 datasets. Solid and dashed lines denote with template updating (adaptive) and without template updating, respectively. 38
2.7 EER of the MmLD (fusion of ECG and LCPD) with template updating compared with conventional liveness detection methods in the third test session. 39
2.8 Comparison of recognition performance of the proposed system with Bozorth in the third test session. 42
3.1 Recording fingertip ECG and TEOAE signals................................. 45

3.2 Visualization of across-session training and testing on the synthetic dataset consists of four features, i.e., \( f_1 \) to \( f_4 \) with 5000 additional irrelevant features. Blue and red samples were used for training and testing respectively. Classification result using \( f_1 \) and \( f_2 \) is shown on the left and result for \( f_3 \) and \( f_4 \) is shown on the right. Classification boundaries are shown with different background colors.......... 57

3.3 Stabilization of TEOAE signal over time. (a) The first 10 samples. (b) The last 10 samples........................................................... 61

3.4 ROC curves for three experiments. (a) Across-session ECG. (b) Across-session TEOAE. (c) Across-posture ECG................................. 65

3.5 Sensitivity of weights to parameter \( \lambda \)........................................... 66

3.6 Sensitivity of weights to parameter \( \sigma \)........................................... 66

3.7 Percentage of positive margins over successive iterations for synthetic dataset.... 67

3.8 The CPU time (in seconds) taken by the proposed algorithm to perform feature selection on the synthetic dataset where the parameters \( \sigma \) and \( \lambda \) are set to 1. (a) CPU time vs number of features. The figure shows linear complexity of the proposed method with respect to feature dimensionality. Number of samples is set to 300. (b) CPU time vs number of samples. The figure shows the proposed algorithm has quadratic time with respect to number of samples. Dimensionality is set to 5000..................................................... 68

3.9 Approximation error in computing \( \bar{d}_{NM}^{(i)} \) in (3.5) for TEOAE dataset. Error is defined as \( \| \bar{d}_{ORG} - \bar{d}_{RP} \|_2 / \| \bar{d}_{ORG} \|_2 \) where \( \bar{d}_{ORG} \) corresponds to MSFS method and \( \bar{d}_{RP} \) corresponds to MSFS-RP method................................. 70

3.10 Percentage of common features shared between MSFS and MSFS-RP versus the number of selected features in TEOAE dataset for \( \gamma \in \{75\%, 85\%, 95\%, 99\%\} \)...... 70

3.11 Average error in the distance between members of a pair of samples versus \( K \). Error is averaged over 30 runs where in each run errors of 1000 pairs are computed. It can be seen that the error significantly increases when \( K \) is smaller than approximately 500........................................... 71
4.1 Illustration of the proposed approach. ........................................ 74
Chapter 1

Introduction

1.1 Background

1.1.1 Electrocardiogram

ECG describes the electrical activity of the heart over a period of time using electrodes placed on the skin. These electrodes detect the tiny electrical changes on the skin that arise due to the sequential depolarization and repolarization of the different heart’s muscles. A large amount of information about the structure of the heart and its electrical activity is encoded in ECG [3]. As shown in Figure 1.1, an ECG heartbeat consists of 5 main characteristic points, P-Q-R-S-T.

![ECG Waveform](image)

Figure 1.1: P, R and T waves of a heart beat.

Different configurations for electrode placement has been suggested. However, the most common configuration is the 12-lead ECG. It uses 10 electrodes, 6 of them are attached to the chest and the other 4 are attached to the limbs. In this configuration, 12 recordings are recorded...
from these 10 leads. Regarding the limbs, there are three main sets of lead orientations. The bipolar limb leads are usually denoted as I, II and III. In lead I the potential difference between the two arms is measured. Lead II, measure the potential difference between left leg and right hand. Finally, in lead III configuration, one electrode is attached to the left leg and the other one is attached to the left hand.

1.1.2 Biometric Recognition

Biometric systems operate in two modes: verification or identification. In the verification mode, the user provides their biometric and claims an identity. Here the task of the biometric system is to accept or reject the claimed identity. Hence, it is a 1-to-1 matching. In other words, it answers the questions: is s/he the subject who s/he claims to be? In the identification mode, the system answers the question: who is s/he? Hence, it is a 1-to-N matching where N is the number of subjects.

In our experiments on ECG, we consider verification mode of operation as is usual for ECG biometric. There are two types of error in verification mode: false rejection and false acceptance. The former occurs when a genuine identity request is mistakenly rejected by the system and the latter occurs when an intruder is mistakenly accepted by the system. Performance is often assessed based on false acceptance rate (FAR) and false rejection rate (FRR) defined as follows.

\[
FRR = \frac{n_{FN}}{n_G} \tag{1.1}
\]
\[
FAR = \frac{n_{FP}}{n_I} \tag{1.2}
\]

where \(n_G\) and \(n_I\) are total number of genuine and impostor attempts, respectively. \(n_{FP}\) is the number of falsely authenticated attempts and \(n_{FN}\) is the number of falsely rejected genuine attempts. A widely-used performance measure is equal error rate (EER) which is the operating point where \(FAR\) and \(FRR\) are equal.
1.2 Motivation

Biometric systems have been deployed around the world and have been used extensively in the past decade. Automatic identity verification is becoming more appealing in many aspects such as financial transactions, tele-medicine and access control. Physiological characteristics such as fingerprint, iris and facial structure and behavioral characteristics such as voice, gait and signature have been used for identity recognition. However, the choice of right biometric depends on characteristics of application environment and usually is a trade-off between factors such as performance, ease of collection, user acceptability and deployment cost.

With the wide deployment of biometric systems, the potential of fooling or spoofing this technology is widely admitted. For example, fingerprints may be left behind whenever you touch a glass surface like phone screen and fake fingerprints can be recreated by some simple procedures using plastic molds and gelatin – see Figure 1.2. Even iris images can be captured from few meters distance. Not to mention face images [4, 5] which can be captured from a longer distance. Voice trait is also vulnerable against replay attacks using pre-recorded voice playbacks. Such vulnerabilities have been motivating researchers towards novel biometrics such as Electrocardiogram (ECG) that are inherently robust to replay and falsification attacks [57].

Figure 1.2: Fake finger can be created using some fairly simple procedures. ¹

ECG is related to variation of electrical activity of heart. It has some advantages over conventional biometrics. ECG is a vital signal and presence of the ECG signal automatically

¹Images are from LivDet 2013
ensures the liveness. Moreover, it is difficult to obtain someone’s ECG without his or her permission. ECG is relatively an inexpensive technology because many mature data acquisition systems for collecting ECG signal is already developed in clinical applications. It is also a natural choice for human identification in medical care and tele-medicine application where ECG signal is collected primarily for diagnosis. Moreover, ECG is a continuous signal which allows a continuous authentication.

1.3 Problem Definition

Perhaps the most advantage of ECG as a biometric is its inherent liveness detection that makes it different from conventional biometrics. Also, ECG can be recorded from fingertips and this makes it a natural choice to be fused with fingerprint. Apart from ECG, there has been a lot of research on fingerprint spoof detection. A major drawback of existing approaches is that they are learned and tuned on a specific known attack. The main stream of these approaches use some training samples artificially produced via a certain spoofing procedure and materials and work well on test samples produced with the same procedure. But their performance on an unknown type of spoof is questionable. In practice, the way that a fake biometric is synthesized is unknown. There are some other spoof detection approaches that look for some evidence of liveness such as temperature, humidity or ECG pulses. Although such workarounds may work against naive spoofing attempts, a smart intruder may be able to use their own ECG to fool such systems. Temperature or humidity can also be intentionally manipulated to fool the system. Therefore, these approaches are seriously vulnerable to spoof attacks.

There are a few prior works in the literature on fusion of ECG and fingerprint that have combined ECG and fingerprint in a conventional way similar to combining any two biometrics and as one may expect identity verification rate improves due to fusion. However, the improved verification rate in the conventional fusion approach is because when input samples of one trait have poor quality and hence are less informative, the other trait will help the system to still identify the user. However, this opens up the possibility of spoofing because it may accept a fake fingerprint even if ECG does not match. One may design the fusion such that it only accepts if both traits show a good match with their corresponding templates. In that case, the
identification rate will degrade because the system will reject even if only one trait has poor quality. In fact, conventional combination of ECG and fingerprint is confined to a trade-off between identification rate and vulnerability to spoof attacks. None of the previous works have studied how ECG can be used in the context of spoof detection and how it compares with existing fake detection methods. We will discuss these issues in chapter 2.

Regarding the ECG as an individual trait, the main difficulty is that it is non-stationary. Through this study by non-stationary we mean that signals vary over different recording sessions or postures. One of the most challenging scenarios is the case that there is only one recording session for enrollment while testing happens in a different session. In fact, in this case, signals’ nonstationarity does not elicit during enrollment session and hence cannot be captured from only one recording session. Different feature extraction and classification techniques have been examined for ECG biometric recognition in the previous works, but none of them have suggested an explicit mechanism specially designed to consider non-stationarity of ECG to cope with across session variations while there is only samples of one session for enrollment. We will discuss these issues in chapter 3 and 4.

1.4 Prior Works on ECG Biometric Recognition

Since ECG is focus of this study, in this section we review prior works on ECG biometric recog- nition. Related works on other topics related to the subject of this research will be discussed later in the corresponding chapters. Previous works on ECG biometric can be categorized in fiducial based and nonfiducial based approaches. Fiducial based approaches relay on some points on a heart beat such as onset and end of each waves. P, R and T waves are required to be located. Features such as peaks, slope, radius of curvature and area is computed in a region surrounding each of P, R, and T waves.

On the other side, nonfiducial based approaches are holistic and consider ECG signal as a set of heart beats or just a time series without segmenting it to heart beats. Detection of fiducial points itself is a challenging process and it may not be possible to locate the onset and end of P, R and T waves due to noise and artifacts in real world scenarios.


1.4.1 Fiducial based ECG analysis

Biel et al. [6] extracted a set of 30 features from the fiducial points. In a 12-lead rest ECG recording configuration, a total of 360 ($12 \times 30$) features were extracted. They manually selected 12 features out of 30. Besides, various lead placements were examined and their results shows only one lead is enough for identification. A PCA based model for each subject was trained. Their method lacks an automatic feature selection. A data set consists of 20 subjects was used in their experiments.

Kyoso et al. [7] used four features i.e., the P, PQ, QRS and QT duration. A threshold base approach is applied on the second order derivative and correlation and Mahalanobis distance were used for classification. Their results shows that QRS and QT intervals provide a better performance compare to other regions.

Shen et al. [8] extracted seven features related to QRS complex in an one-lead setup. Their classification consisted of two steps: template matching and decision-based neural network (DBNN). In template matching step, two signals were considered correlated if their correlation coefficient is bigger than a threshold. Correlated signals were passed to DBNN for final decision.

Isard et al. [9] extracted nine features related to duration of different intervals. It was suggested that such temporal features are more robust against sensor misplacement compare to amplitude based features. A band-pass filter 1.1- 40 Hz was used for preprocessing. Variations due to lead placement and stress were also investigated. Linear Discriminant Analysis (LDA) was used for classification.

Palaniappan et al. [10] proposed to use a form factor which is a measure of signal complexity. Six features related to QRS complex were extracted. For classification, they examined several architectures for neural network and in average a multilayer perceptron using back propagation (MLP-BP) with one hidden layer achieved the best performance. Their data set consisted of 10 subjects.

Kim [11] suggested not to consider P wave for feature extraction because of difficulty of detecting it specially when subjects perform physical activities. Classification were done using Mahalanobis distance in the same way as [7]. A database of 10 subjects were used in this work.

In method proposed by Saechia [12], heat beats were segmented. Each heart beat was
divided into three parts related to P, QRS and T waves. Fourier transform of each part along with the heart beat itself were passed to a neural network. Their results suggested that by considering aforementioned three parts significant improvement can be achieved compare to the case that whole heart beat is passed to a neural network. Hear rate variation is addressed by normalizing each hear beat to have 750 millisecond length.

In Wang et al. [13] 15 temporal features from fiducial points were extracted similar to [14] and [9]. In addition, six amplitude features were extracted from fiducial points. Preprocessing were done by bandpass filtering of raw signal between 1 - 40 Hz. Heart beats were segmented via considering a 800 millisecond window around the R peaks. R peaks were located using a QRS detector in [15] and [16]. Each heart beat were normalized with respect to P’T’ duration to reduce the effect of heart rate variation. A thresholding approach were used to remove outliers. In addition to fiducial based features, holistic features using PCA and LDA were also examined. Their data set consisted of 13 subjects.

In methodology proposed by Irvin et al. [14], raw signal is filtered using a bandpass filter 0.2 - 40 Hz. To locate R peaks, they computed variance over a 0.2 second interval and looked at the maximum variance among overlapping 0.2 windows. P, Q, R, and S points as well as T’, S’ P’ and L’ points were localized and fiducial based features were extracted similar to [9]. A sequential sampling procedure is used for decision making.

Ting et al. [17] proposed to use a nonlinear dynamic model to represent ECG signals in a state space. Extended Kalman filter is used to infer the posterior states and Log-likelihood scoring is used for classification. Their data set consisted of 13 subjects.

In Venkatesh et al. [18] nine fiducial based features were extracted. They used only features related to QRS complex. Dynamic time warping (DTW) and fisher linear discriminant analysis (FLDA) followed by k-NN were examined for classification. Their data set consisted of 12 subjects.

In Tawfik et al. [19] heart rate variation is compensated using Framingham correction formula that linearly normalize the QT interval to 60 beats per minutes heart rate. They suggested not to use P wave as it changes by hear rate and also because it is difficult to detect P wave and delineate it. Raw signals were bandpass filtered between 1- 40 Hz. Classification is done by passing DCT coefficients of normalized signals to a neural network. They examined
two cases: identification based on QT interval and identification based on QRS complex only. The best performance obtained by using QRS complex only. Their data set consisted of 20 subjects.

In Tantawi et al. [20] QRS, P and T waves were localized and 19 temporal, 6 amplitude and 3 angle fiducial based features were extracted similar to [21]. Amplitude features were normalized with respect to R peak amplitude. Heart rate variability is compensated by normalizing each heartbeat with respect to full heart beat duration. For dimension reduction PCA, LDA, information-gain ratio (IGR), and Rough Sets (RS) were examined. Classification was done by a radial basis functions neural network. Raw signals were bandpass filtered between 0.5 - 40 Hz. It was suggested that temporal features are not as discriminant as amplitude and angle features due to inaccurate localization of fiducial points.

In Kaveh et al. [22] temporal and spectral features in a single lead ECG framework were considered. Fiducial points and QRS, P and T waves were localized using DCT and wavelet. PCA was used for dimension reduction. Classification was done using SVM with a RBF kernel. SVM was used in an one-versus-all approach such that one SVM was trained for each subject.

1.4.2 Nonfiducial based ECG analysis

Unlike the fiducial point based approaches that use characteristic points to compute features that represent the signal, nonfiducial based approaches don’t use characteristic points for computing features. However, in some works, characteristic points were used for preprocessing i.e. segmentation and signal alignment. We consider them as nonfiducial based approaches.

The work in [23] proposed to use Dynamic Time Warping (DTW) to compensate the effect of heart rate variation. ECG signals were segmented and aligned with respect to R peaks. ECG signals were collected from two electrodes placed on both wrists. Their dataset consists of ECG signals of 10 subjects recorded in 5 sessions distributed over four weeks. Training and testing samples were selected randomly from all 5 sessions and hence samples from all sessions were used for training. According to the experimental results, DTW is more effective when samples of both low and high heart rates are available during enrollment. So that the heart rate of query is between low and high heart rates seen during the enrollment.

In Plataniotis et al. [24] input signals were divided into several windows and for each
window autocorrelation (AC) coefficients were computed. Discrete Cosine Transform (DCT) were used for dimension reduction and classification were done by normalized Gaussian log likelihood and Euclidean distance. Input signals were bandpass filtered between 0.5 - 40 Hz. All autocorrelations were normalized by the autocorrelation value at lag zero.

In Agrafioti et al. [25], similar to [24] autocorrelation features were extracted from a set of overlapping windows and normalized by their maximum value. DCT and LDA were examined for dimension reduction and LDA outperformed DCT. Nearest Neighborhood with Euclidean distance was used for classification.

In Chan et al. [26] a wavelet-based distance was used for evaluation of differences between two ECG signals. Signals were recorded from fingertips. 50 subjects were recorded for three sessions with a minimum of one day between sessions and each session was 90 seconds. In their work, PQRST complexes were detected and then aligned using cross-correlation.

The work in [27] applied principal component analysis (PCA) to raw ECG signals segmented with respect to the R peaks to extract features. Their data set consisted of 39 subjects where one session is recorded for each subject and training and testing signals selected from the same session. Classification was done by nearest neighborhood with Euclidean distance.

In Fatemian et al. [28] Discrete Wavelet Transform (DWT) were used to denoise and delineate the ECG signals. QRS, T, P waves were detected. To reduce the effect of heart rate variation, each heart beat was normalized such that the region related to T wave was resampled to have a length of 120 milliseconds. Then the resulted heart beat was resampled to have a fix length of 850 milliseconds. As a result of the proposed heart beat normalization, the length of each heat beat is fix and invariant to heart rate variation. Normalized heart beats of each subject were averaged to build the final template. A correlation analysis were used for classification.

The work proposed by Fang et al. [29] and [30] is based on phase space reconstruction. ECG signals were segmented and aligned with respect to R peaks and averaged to form an averaged waveform. The average heart beat was embedded in a time-series of three-dimensional vectors with a constant time delay. The reconstructed phase space was used directly as features representation of input signal. PCA was used for dimensionality reduction and classification was done using a Neural-Network. Preprocessing is done via a band-pass filter between 2-50Hz.
Their data set consists of ECG signals of 100 subjects collected from chest area in two sessions. Second session signals were collected 30 minutes after the first session.

Odinaka et al. [31] proposed to use Short-time Fourier Transform (STFT) for feature extraction. They used a feature selection based on symmetric Kullback-Leibler (KL) distance. Each subject was modeled by an axis aligned Gaussian distribution as a generative model and Log-likelihood ratio (LLR) were used for classification. ECG signals of 269 subjects were collected from chest area for 3 sessions. This study shows the effectiveness of feature selection to improve the performance.

In Ye et al. [32] Discrete Wavelet Transform (DWT) and independent component analysis (ICA) were applied on each heart beat to extract features. PCA were used for dimension reduction. SVM with RBF kernel were used for classification. This work used the ECG signals of two leads and a decision level strategy were used to fuse the results related to two leads. For each heart beat if the decision based on two leads were different, the whole heart beat is discarded.

In Li et al. [33] cepstral information extracted from the ECG signals and modeled by Gaussian mixture modeling (GMM). Dimentially reduction is done by heteroscedastic linear discriminant analysis [34]. Besides, hermite polynomial expansion (HPE) was used to extract features in time domain and support vector machine (SVM) was used for classification of HPE features. Both time domain and cepstral information were fused is score level to improve the overall performance. For preprocessing, a band-pass filter between 0.5-40Hz was used.

In Pathoumvanh et al. [35], ECG signal is segmented to heart beats. Continuous wavelet transform (CWT) of each heart beat is used for feature extraction. Total energy of wavelet coefficients for P, QRS, and T intervals are computed and Fisher Linear Discriminant Analysis (FLDA) is used for dimensionality reduction. Classification is done by 1-NN with Euclidean distance. Experiments were done on a dataset consists of 10 subjects. Only one session for each subject was available on their dataset.

In Zhao et al. [36] ensemble empirical mode decomposition (EEMD) which is a variation of empirical mode decomposition (EMD) is used for feature extraction. ECG signal is decomposed into a number of intrinsic mode functions (IMFs) and PCA is used for dimensional reduction. Classification is done by using k-NN. Experiments are done on MIT-BIH and PTB datasets.
The effect of heart rate variation is compensated via a heartbeat normalization strategy similar to [28] based on resampling PQ, T and ST intervals to have a fixed predefined length. Similar to [37] periodicity transform (PT) were used as a quality measure of heart beats.

In Wang et al. [38] ECG signal is divided into a set of overlapping windows. Each window is normalized to have a unit length. A dictionary is learned under sparsity constraint. All samples were represented by linear combination of atoms of the dictionary under $l_1$ norm constraint using LARS algorithm. Sparse coefficients of each recording were mixed using a max-pooling approach. Classification is done by 1-Nearest Neighbor (1-NN). 100 subjects from PTB dataset were selected and recognition rate of 99.48% were reported while test and train samples were selected from the same session.

1.5 Contributions

This thesis introduces algorithms and systems to address the problems discussed in section 1.3. The main contributions of this work can be categorized as follows:

- **Liveness Detection and Automatic Template Updating using Fusion of ECG and Fingerprint**: In chapter 2, a unified approach towards human recognition using ECG and fingerprint with emphasis on liveness detection is presented. Different from conventional biometrics such as fingerprint, iris and face, ECG is a vital signal and motivated by its inherent liveness, we propose to fuse it with a fingerprint liveness detection algorithm. ECG signals are collected from fingertips that naturally makes sense to be fused with fingerprint. In the proposed approach, ECG is incorporated to improve both liveness detection and biometric recognition rate of a fingerprint system. In addition, the resulting system is capable of updating templates and hence manual re-enrollment is no longer necessary for maintaining the performance of the biometric system in long term. Different from previous ECG recognition works that usually limit the length of sessions to a fixed value, we propose a stopping criteria that takes into account the fact that some subjects have a very repetitive ECG and does not need as many heartbeats as other subjects. We have performed comprehensive experiments on LivDet2015 database which is presently the latest available liveness detection database and compare the proposed
system with state-of-the-art fingerprint liveness detection methods.

- **Feature Selection for Non-stationary Data:** In chapter 3 we present a novel feature selection method for biometric recognition using medical signals such as ECG. Such signals are time-dependent and this makes them hard to deal with in across-session human recognition scenario where only one session is available for enrollment. The proposed method is based on an auxiliary dataset with multiple sessions where it selects a subset of features that are more persistent across different sessions. It uses local information in terms of sample margins while enforcing an across-session measure. This makes it a perfect fit for aforementioned biometric recognition problem. Extensive experiments on ECG variability due to time lapse and body posture are done. In addition to ECG, we have shown the generalization of the proposed algorithm on Transient Evoked Otoacoustic Emissions (TEOAE) signals which is a medical signal collected from ear. Performance of the proposed method is compared against seven state-of-the-art feature selection algorithms as well as another six approaches in the area of ECG and TEOAE biometric recognition.

- **Cross View Matching for Biometric Recognition:** To cope with the effect of heart rate on performance of ECG recognition, we have developed a novel method for cross-view matching where distance between two samples situated in different views can be computed. In this context view refers to different heart rate where samples in different views can be represented by feature sets that can be different in type and length. Hence, direct comparison of samples does not help. We have shown the generalization of the algorithm on other scenarios including cross-pose face recognition and near infrared versus visible light face recognition.
1.6 List of publications

The publications, including journal papers, conference papers during my PhD studies at the University of Toronto are listed below.

• Journal Papers


• Conference Papers


• Patents

1.7 Thesis Outline

The rest of this thesis is organized as follows. Chapters 2, 3 and 4 include the three main contributions of the thesis as explained above. Finally, the thesis summary and concluding remarks are presented in Chapter 5.
Chapter 2

Fusion of ECG and Fingerprint for Spoof Detection

2.1 Introduction

Nowadays biometric spoofing and countermeasures is an active research area and there has been a lot of efforts towards a promising approach to ensure the presence of a real legitimate user. There are different attack points in a biometric system. The first vulnerable point is the sensor used in the biometric system. Biometric systems and in particular fingerprint can be spoofed by presenting synthetic samples to the sensor e.g. gummy fingers that have fingerprint impressions. There has been a huge literature on other vulnerabilities that for example bypass feature extraction or matcher, or manipulate database or communication channel. However, in such cases, some information about the system such as feature extraction, matcher, database and/or physical access to some of those components is necessary. In contrast, fooling the sensor using a fake biometric sample does not need any specific information about internal mechanism of the biometric system. In addition, sensor level attacks are in analog domain and hence many solutions such as cryptography and watermarking that are in digital domain are not useful. This highlights the importance of developing biometric spoofing countermeasures to classify an input sample as live or fake that is focus of this chapter. Investigating other types of attacks is out of scope of this study.
Liveness detection has been an active area in the past decade and numerous approaches have been proposed in the literature to solve this problem. Considering the results reported in liveness detection competitions including LivDet2009 [39], LivDet2011 [40], LivDet2013 [41] and LivDet2015 [1], liveness detection is still an open problem and performance of the existing approaches does not satisfy requirements of many practical applications [42]. The main stream of the current approaches use some training samples artificially created via certain spoofing process and work well on test samples created by the same process involved in training, but their performance on a novel type of spoof is questionable. In practice, the way that a fake biometric is fabricated is unknown.

Our comparative analysis of 6 state-of-the-art liveness detection algorithms on LivDet2015 which is the latest database in this field shows that the best method has an average EER of %10.9 that confirms vulnerability of the current approaches to spoof attacks. Further, performance greatly varies from one spoof material to another. For example, in our experiments on DigitalPersona dataset, the average EER of the 6 state-of-the-art liveness detection algorithms on Latex as an unknown spoof is %10.2 but the average performance on WoodGlue degrades to %36.5. This again highlights the vulnerability of the current fingerprint recognition systems to novel spoof attacks.
Considering the aforementioned shortcomings, in this study we approach the problem from a different perspective by incorporating electrocardiogram (ECG). ECG is among the newer additions to the biometric family and unlike the conventional biometrics such as fingerprint, iris and face, ECG is a vital signal and presence of the ECG automatically ensures the liveness [26,27,29,30,32,43,44]. In addition, conventional biometrics can be easily stolen from people. For example, fingerprints may be left behind whenever you touch something and it can be recreated by using some simple procedures. Even iris images can be captured from a few meters distance. Not to mention face images which can be captured from a longer distance. However, compared to conventional biometrics, if not impossible it is far more difficult to steal someone’s ECG. Besides, ECG is relatively an inexpensive technology because it has been widely used in health care applications for diagnosing cardiac disorders in the past few decades and many mature and inexpensive ECG acquisition devices have been developed. Beside its advantages, accuracy of ECG is not as good as some other mature biometrics such as fingerprint [45,46]. Therefore, we seek to fuse ECG and fingerprint to improve the recognition accuracy and more importantly benefiting from ECG’s inherent liveness detection.

Fusion of ECG and fingerprint is not a new idea. In [47] and [48] it has been suggested to combine ECG with other biometrics such as face and fingerprint to get a better recognition rate. But, all these works were restricted to analyzing recognition rate and have failed to consider spoof attacks. While fusion can improve the recognition rate, it may forfeit the security of the system. The improved recognition rate in the conventional fusion approach is because when input samples of one trait have poor quality and hence less informative, the other trait will help the system to still identify the user. However, this opens up the possibility of spoofing because it may accept a fake fingerprint even if ECG does not match. Therefore, a multimodal system based on conventional fusion of ECG and fingerprint forfeits the ECG’s inherent liveness detection for a better recognition rate. This contradicts the main motivation to use ECG, i.e. liveness detection. This issue has been overlooked in the previous works [47,48].

In this chapter, we offer an alternative to the conventional fusion of ECG and fingerprint by proposing to fuse ECG with fingerprint for liveness detection. To this end, we combine ECG recognition score with fingerprint liveness detection score instead of fingerprint recognition score. This greatly improves the accuracy of liveness detection task, however unlike conventional
fusion of ECG with fingerprint, it does not forfeit the security because a liveness detection
algorithm by definition is supposed to be robust against spoofing. Figure 2.1 shows an overview
of the proposed method which will be explained in detail in section 2.3. In this context, we use
the terms “fingerprint liveness detection” and “fingerprint recognition” to indicate two blocks
in the proposed system as shown in Figure 2.1. Our comprehensive experiments on 6 different
spoofs demonstrate that the proposed method not only improves the average liveness detection
performance but also significantly reduce the performance variation among different spoofs.
The ECG signals used in this study are collected from fingertips. Majority of previous works
on ECG recognition has been based on signals collected from chest area [9, 20, 22, 25, 38] or
lower rib cage [31] and only a few works has been done based on fingertip ECG signals, e.g.
[49, 50]. Fingertip ECG has two advantages: First, it eliminates the need for user to undress for
electrode placement; Second, it makes the fingerprint a natural choice to be fused with ECG.
In this study, we explore the fusion from two different perspectives: Improving the liveness
detection performance of conventional liveness detection methods, and improving recognition
rate of conventional fingerprint recognition methods. The main contributions of this chapter
are as follows:

- In order to get the most out of ECG, we fuse it with a fingerprint liveness detection method
  for liveness detection purpose and also fuse it with a fingerprint recognition method for
  recognition purpose. Although the latter has been previously investigated, e.g. in [47]
  and [48], to the best of our knowledge, the former has not been explored in the previous
  literature.

- In addition, the proposed system is capable of automatically adapting ECG and fingerprint
  templates to operational data. Since ECG is time dependent and might be affected by
  factors like diet and emotion, template updating is crucial to maintain the performance of
  the system in long term without requiring to re-enroll or retrain the system from scratch
  and to the best of our knowledge this has not been investigated in the literature.

- Another shortcoming of the previous works on ECG [27, 31, 51–57] is lack of a proper
  stopping criteria to limit the length of recording sessions. Previous works usually limits the
  length of sessions by fixing the number of recorded heartbeats to a predefined threshold.
Therefore, the number of recorded heartbeats is the same for all subjects which does not consider the fact that some subjects have a very stable ECG and don’t need as many samples as other subjects with less stable ECG. In this study, we present an easy-to-compute yet effective criterion based on local averaging and correlation that measures heartbeat consistency (HC) in successive heartbeats. This reduces the average recording time.

The rest of this chapter is organized as follows: Section 2.2 briefly reviews the previous works on fusion of ECG and fingerprint as well as fingerprint liveness detection. The proposed method on fusion of ECG and fingerprint is presented in section 2.3. Details of automatic template updating are explained in section 2.4. Comparison with other methods as well as participants of LivDet2015 is presented in section 2.5 and finally section 2.6 concludes the chapter.

2.2 Related work

2.2.1 Fusion of ECG and Fingerprint

While multimodal biometric systems based on conventional traits such as face and fingerprint has been extensively investigated in the literature [58, 59], there exist only a few works about a multimodal biometric system that includes ECG. In [48] the idea of securing handheld devices and fingerprint readers with ECG biometrics is pointed out and a biometric system based on ECG signals collected from fingertips is investigated but no experimental result on fusion of ECG and fingerprint was reported.

In [47], fusion of face, fingerprint and ECG is studied. They used an ECG dataset of 78 subjects acquired from European ST-T Database, MIT-BIH Normal Sinus Rhythm Database, MIT-BIH Arrhythmia Database and QT Database of PhysioBank [60]. A fiducial based method previously presented in [61] were used for ECG recognition. It involves extracting fiducial features related to various intervals, amplitude and angles. For fingerprint and face recognition, match scores provided in NIST-BSSR1 [62] were used. BSSR1 has match scores for two face and one fingerprint matcher. Since ECG signals used in their work were acquired from chest area, its fusion with fingerprint is of little value in real world scenarios because the user needs
to undress for electrode placement. Further, the ECG recognition method that were used is fiducial-based and as suggested in [51] is not appropriate for real-world scenarios where accurate detection of characteristic points may not be possible due to noise. Moreover, they performed fusion only to achieve a better recognition rate and indeed they failed to benefit from the main advantage of ECG, i.e. its inherent liveness detection.

However, our work doesn’t have aforementioned limitations. Our ECG signals were collected from fingertips. More importantly, in addition to fusion of ECG and fingerprint for human recognition, for the first time, we perform fusion of ECG and fingerprint for liveness detection. In the following section, we briefly explain the previous works on fingerprint liveness detection.

### 2.2.2 Fingerprint Liveness Detection

Liveness detection methods can be divided into two main categories: hardware or software-based. Hardware-based techniques add some specific components that look for some evidence of liveness such as temperature, humidity or ECG pulses. Although such workarounds may work against naive spoofing attempts, a smart intruder may be able to use their own ECG to fool such systems. Temperature or humidity can also be intentionally manipulated to fool the system.

In software-based solutions, detection is performed by processing the obtained image. In [63], morphology-based and perspiration-based features are jointly considered. Pore perspiration is investigated in [64–66]. In [67] the difference between quality of fake and real samples is used to detect fake samples. This is realized through analyzing 25 image quality measures extracted from samples. There are other works that have successfully applied local descriptors for liveness detection. Local Binary Pattern (LBP) and its extensions has been used for liveness detection in [68,69]. LBP is a texture descriptor based on the intensity difference between neighboring pixels and a center pixel of a patch of input image [70]. In [71], a method based on Local Phase Quantization (LPQ) was proposed. Similar to LBP, LPQ works on patches but instead of gradient it computes phase information by computing short time Fourier transform. Phase information are then decorrelated and uniformly quantized. Another descriptor that has been used for liveness detection is Weber Local Descriptor WLD [72]. Binarized Statistical image Features (BSIF) [73] was also tested for fingerprint liveness detection. It is based on binarizing
the response to a set of filters. Filters are not fixed and are learnt using independent component analysis (ICA). Recently a method known as Local Contrast Phase Descriptor (LCPD) is presented in [74] that takes into account a spatial-domain component inspired by WLD and a phase component inspired by LPQ. To improve the performance of liveness detection, there has been a few works on fusion of liveness scores with conventional biometrics such as fingerprint and face [63,75]. Since these biometrics are easy to spoof, when involved in the fusion process, make the security of the resulting system questionable. In the 2-dimensional space defined by liveness and recognition scores, distribution of samples from different spoofs can greatly vary [75]. Therefore, any decision boundary in the aforementioned 2-dimensional space may fail when encountering a novel spoof with a different distribution. Recently, a method based on 1-median filtering is presented in [76]. As an alternative to sum rule as a conventional fusion rules, 1-median filtering is used on a mutibiometric problem with 5 traits. However, 1-median filtering requires enough number of traits to determine the median and it cannot be used in our scenario that has 2 traits, i.e. ECG and fingerprint.

In spite of all efforts in this field, accuracy of current liveness detection methods does not satisfy the requirements of many practical applications. All these methods suffer from significant performance variation when encountering different spoofs especially when encountering novel spoofs that has not been seen during training [42]. A detailed survey of liveness detection approaches can be found in [77] and [78].

2.3 Proposed Method

2.3.1 Overview

Multimodal Liveness Detection and Multimodal Human Recognition

In general, a fingerprint biometric system has a fingerprint liveness detection (FpLD) module and a fingerprint recognition (FpR) module. A desired system should have a good recognition rate as well as a good liveness detection performance. In the proposed system, both fingerprint and ECG are collected from fingertips and ECG is incorporated to improve both liveness detection and recognition rate of a fingerprint biometric system. Block diagram of the proposed
system is shown in Figure 2.1.

FpLD is a binary classifier, i.e. live/fake that given a fingerprint sample, provides a liveness score. ECG recognition can also be treated as a binary classifier, i.e. genuine/impostor that given an ECG signal, provides a score that can be further fused with FpLD score to form a Multimodal Liveness Detection (MmLD) block. The use of ECG for fingerprint liveness detection is motivated by the assumption that ECG is hard to spoof compare to conventional biometrics and presence of an authentic ECG inherently implies liveness of the subject. Another block is FpR that can also be treated as a binary classifier, i.e. genuine/impostor that given a fingerprint sample, provides a score that can be further fused with ECG score to form a Multimodal Human Recognition (MmHR). While the MmLD block does the liveness detection task and aims to reject the spoof attempts, the second block, i.e. MmHR, does the recognition task and aims to reject impostor attempts. In this way, ECG is used to improve both liveness detection and biometric recognition aspects of the system.

Adaptive Templates

ECG and fingerprint samples that their scores are greater than some updating thresholds can be added to previously existing training samples to update the templates. This process is known as template updating or adaption [79]. The updating thresholds usually are different than the acceptance thresholds and are set to zero false acceptance rate point to ensure that only live and genuine samples are selected for adaption. This solves the problem of long term maintenance of the biometric system through automatic updating of ECG and fingerprint templates without being susceptible to adaptation to impostor/spoof samples.

2.3.2 Datasets and Experimental Setup

We use ECG database collected in our lab (BioSec) in University of Toronto [80]. ECG signals were recorded using Vernier EKG sensor and Go!Link interface [81] with 12-bit resolution and sampling rate of 200 Hz using three dry AgCl electrodes from fingertips. There are 82 subjects that have 2 or more ECG recordings in sitting posture. 46 out of 82 subjects have exactly 5 sessions. Follow-up sessions are collected over a six-month period. We divide the dataset into 2 parts. Aforementioned 46 subjects are used for enrollment and testing and the remaining 36
Table 2.1: Characteristics of LiveDet2015 [1] datasets used in our experiments.

<table>
<thead>
<tr>
<th>Scanner</th>
<th>Model</th>
<th>Resolu. (dpi)</th>
<th>Size (px)</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>DigitalPersona U.are.U5160</td>
<td>500</td>
<td>252 × 324</td>
<td>PNG</td>
<td></td>
</tr>
<tr>
<td>GreenBit DactyScan26</td>
<td>500</td>
<td>500 × 500</td>
<td>PNG</td>
<td></td>
</tr>
<tr>
<td>Biometrika HiScan-PRO</td>
<td>1000</td>
<td>1000 × 1000</td>
<td>BMP</td>
<td></td>
</tr>
</tbody>
</table>

subjects are used as an auxiliary dataset for feature selection that will be described in section 2.3.3.

We use LivDet2015 [1] as fingerprint database that includes 4 datasets corresponded to 4 different scanners: CrossMatch, DigitalPersona, GreenBit and Biometrika. Among them, CrossMatch is excluded from our experiments due to small number of spoof samples per finger that makes it unsuitable for template updating. Characteristics of these datasets are shown in Table 2.1.

The three datasets that are considered in this study contain live and fake samples that are constructed through 6 different spoofs including Ecoflex, Gelatine, Latex, WoodGlue, Liquid Ecoflex and RTV. Each dataset originally comes in two sets: training and testing. The training set is the same as testing set but does not include Liquid Ecoflex and RTV spoofs. In this study, to emphasize on the more challenging case of detecting unknown spoofs, one experiment is performed for every spoof such that samples of that spoof are omitted from the training set. Thus, type of spoof is always unknown in all 6 experiments of each scanner.

A chimeric dataset is constructed by combining the above ECG and fingerprint datasets. Each subject in ECG dataset is paired with a subject in the fingerprint dataset in a random way. In this way, a chimeric dataset of 500 unique subjects is generated. All experiments are repeated 5 times to cope with the randomness in constructing the chimeric dataset and the average results are reported. We consider 2 types of attack in construction of the chimeric dataset: zero-effort attack and spoof attack. In the zero-effort attack, intruder provides his own fingerprint sample but claims to be someone else, hence the term zero-effort. In the spoof attack, intruder provides a fake fingerprint sample of the claimed identity. In both cases, ECG signal is also recorded. 3 fingerprint datasets that are considered in this study have 4 fake samples per finger. Hereinafter, we consider each finger as a subject due to the limited number of actual subjects. Note that not all fingers have fake samples. Also, not all subjects have
samples of all 10 fingers. For each subject 4 live and 3 fake samples are used: 1 live sample for enrollment and 3 live and 3 fake samples for subsequent test sessions. The number of test sessions, 3, is due to restriction of the number of fake and live samples that belong to the same subject in LivDet2015 database.

As explained above, the ECG dataset has 5 sessions. But, the fingerprint datasets have only 4 sessions. We pick the 2nd session for training and 3rd, 4th and 5th sessions for testing. We ignore the first session and train on the second session because the second session is less noisy that makes it a better choice for enrollment and constructing templates.

Number of subjects with the aforementioned number of live and fake samples varies among the different datasets and spoofs and ranges between 60-100. Although the maximum number of positive trials can be $46 \times 60 = 2760$, for computational simplicity we consider 500 positive trials. As shown in Figure 2.2, in each test session, there are 500 positive trials and 500 negative trials. Half of the negative trials are corresponded to spoof attacks and the other half is corresponded to zero-effort attacks. Spoof and zero-effort attacks are to be detected by MmLD and MmHR blocks respectively.

Figure 2.2 shows details of trials used in our experiments. The first column is trial index.
Trials are divided into 3 sections that are described in the second column. Forth column is index of chimeric templates, i.e. the claimed identity and the third column describes the relation between query and claimed identity.

Note that in the forth column, claimed identities are repeated from positive trials to negative trials and because of that there is one positive and one negative trial for each template. This is important from template updating point of view. If there was a claimed identity in the negative trials with no positive counterpart, it will never be updated by a positive sample because it will never be compared against a positive sample. Likewise, if there was a claimed identity in the positive trials without any negative counterpart, it will never be evaluated against any impostor. If n-th template is selected to be updated due to the outcome of the n-th positive trial, then in next iteration the updated template will be used for both n-th positive and n-th negative trials.

An important characteristic of this experimental design is that it keeps updating of templates independent of each other. Chimeric templates consist of two parts: ECG and fingerprint. Different subjects may have at most one part in common. Consider two chimeric subjects that may have either ECG or fingerprint in common and only one of them is selected to be updated. In the next test session, both ECG and fingerprint parts of the selected subject’s template are updated and subsequently will be used in one positive and one negative trial associated with that template. This will not affect any other trials that may or may not have the same ECG or fingerprint part. Therefore, having templates with the same ECG or fingerprint part can be used along with the adaption process described above and it has the advantage of increasing the number of chimeric subjects instead of restricting it to the minimum of the number of subjects in ECG and fingerprint datasets.

2.3.3 ECG Recognition

We form a comprehensive pool of different types of features and a subset of more persistent features are selected by looking into an auxiliary dataset that contains multiple sessions. Having a subset of more persistent features, we enroll the biometric system’s users using a single session based on only a subset of selected features and perform testing on another session which is at least 1 week apart. Auxiliary dataset is dedicated to feature selection and there is no overlap
between subjects of auxiliary dataset and actual biometric system’s users involved in enrollment and testing. We also present a stopping criterion to decide when to stop ECG recording in each session.

**Description of Features**

Different types of features have been employed in the literature to represent ECG signals. Unlike many of previous works that rely on only one type of feature, we concatenate different types of features and form a fairly comprehensive feature vector. We use short-time Fourier transform with Hamming window of the length 16 with step size of 13 computed over a 1 second window centered at R peak. Continuous wavelet transform with 32 scales and Daubechies 5 as mother wavelet is computed on a 1 second window centered at R peak. Mean of power, standard deviation of power, maximum amplitude, standard deviation of amplitude, kurtosis and skewness are computed on a 2-second window centered at R peak on the following frequency bands: 8-13Hz, 13-18Hz, 18-25Hz, 25-30Hz, 30-35Hz, 35-50Hz. The signal’s amplitude itself is also considered. This gives a total of 7198 features. We use z-score normalization for every feature, so that features have zero mean and unit variance after normalization.

**Feature Selection**

In this section, we propose a feature selection algorithm that selects a subset of features such that class separability is maximized and within class variation across different sessions is minimized. We will revisit the feature selection problem in chapter 3 where we explore this area more in depth. Let \( \mathcal{X} = \{X_1, \ldots, X_N\} \) be the auxiliary data set consists of \( N \) subjects and for each subject \( S_i \) different sessions are available i.e. \( X_i = \{X_{i,1}, \ldots, X_{i,S_i}\} \), where \( X_{i,j} \) consists of samples of \( j \)-th session of \( i \)-th subject. Weight of \( l \)-th feature has two parts. The first part \( w_1(l) \) encourages class separability and is defined as follows:

\[
w_1(l) = \frac{1}{N} \sum_{i=1}^{N} d(f(X_i(l)), f(\mathcal{X}(l)))
\]

where \( d(\cdot) \) is the symmetric Kullback-Leibler divergence and \( f(\cdot) \) denotes probability density function (pdf). \( f(X_i(l)) \) is pdf of \( l \)-th feature computed over all samples of \( i \)-th subject and
\( f(X(l)) \) is pdf of \( l \)-th feature computed over all samples. We consider normal distribution and use maximum likelihood estimates, i.e. sample means and variances to estimate the symmetric Kullback-Leibler divergence as follows:

\[
d(f_1, f_2) = \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} + \frac{\sigma_2^2 + (\mu_1 - \mu_2)^2}{2\sigma_1^2} - 1
\] (2.2)

where \( f_1 = N(\mu_1, \sigma_1^2) \) and \( f_2 = N(\mu_2, \sigma_2^2) \) are two distributions whose distance is to be computed. \( w_1 \) is large when the overall distribution of every subject is different from the overall distribution of all subjects. However, \( w_1 \) only considers overall class distribution of a subject and disregards the session distributions of that subject. To address this limitation, we define the second part, \( w_2 \), as follows:

\[
w_2(l) = \frac{1}{\sum_{i=1}^{N} M_i} \sum_{i=1}^{N} \sum_{j=1}^{M_i} d(f(X_{i,j}(l)), f(X_i(l)))
\] (2.3)

The first summation is over all subjects and the second summation is over all sessions. \( M_i \) denotes the number of \( i \)-th subject’s available sessions and denominator is a normalization factor. \( w_2 \) encourages the stability across multiple sessions and is smaller when a session can represent the class distribution. Finally, considering both \( w_1 \) and \( w_2 \), weight of the \( l \)-th feature, \( w(l) \), is defined as follows:

\[
w(l) = \lambda w_1(l) - (1 - \lambda) w_2(l)
\] (2.4)

where \( \lambda \) is a constant that controls the trade off between class separability and across session stability. Features can be selected by comparing their weights with a threshold. We can also sort features according to their weights and pick from top features. A proper value for \( \lambda \) can be determined using cross validation. In our experiments, \( \lambda \) and number of selected features are set to 0.3 and 400 respectively.

**Stopping Criteria**

Previous works on ECG [31, 51–57] suffer from being restricted to a fixed recording length. Motivated by the fact that some subjects have a very repetitive ECG and does not need as many heartbeats as some others, we define heartbeat consistency (HC) as a stopping criterion.
Figure 2.3: Average recording time versus HC.

as follows:

\[ HC(n) = corr \left( (b_n + b_{n+1}), (b_{n+2} + b_{n+3}) \right) \]  

(2.5)

where \( corr \) is correlation function and \( b_n \) is \( n^{th} \) heartbeat amplitude. Since the subjects are in the rest condition during recording, a longer recording time provide more time for the subject to rest and allows their ECG to better stabilized. We keep recording until HC exceeds a predefined threshold or the recording time reaches 30 seconds, whichever satisfied first. Therefore, a session takes at most 30 seconds. As shown in Figure 2.3, increasing HC threshold increases the average recording time because a larger HC threshold is harder to satisfy. For example, recording time for different sessions of subjects for HC=0.92 is shown in Figure 2.4. This is corresponded to an average recording time of about 10 seconds. For each subject, recording time of 3 test sessions are marked with circles. The average length of sessions are also provided as a solid red line. Note that although the average recording time is about 10 seconds, majority of sessions have a significantly smaller length (about 5 seconds) and only a small number of sessions require a longer recording time. Any R peak appeared during the first second of recording cannot be used because some of the aforementioned features require a 2-second window centered at R peak. Moreover, by definition HC needs at least 4 full heartbeats and that limits the minimum recording time to about 5 seconds.

Classification

We use linear SVM as classifier and since it is a binary classifier, a one-versus-all strategy is adopted. The score corresponded to a session (i.e. \( s \)) is determined as weighted sum of its
heartbeats scores (i.e. $s_i$) as follows:

$$s = \sum_{i=1}^{K} \frac{w_is_i}{\sum_{j=1}^{K} w_j}$$

(2.6)

Note that length of sessions are determined by HC and $K$ is not fixed and above summation is indeed over all heartbeats of each session. Denominator is just a normalization factor and weights $w_i$ are defined as:

$$w_i = \frac{1}{1 + \exp(-\frac{s_i - \mu}{\sigma})}$$

(2.7)

where $s_i$ is i-th heartbeat’s score and $\mu$ and $\sigma$ are sample mean and standard deviation of heartbeat’s score in the corresponding session. This definition of weights can be interpreted as probabilities that are assigned to each heartbeat. So that we assign higher probabilities to heartbeats with larger scores. The final decision for a session can be made by comparing its score against an acceptance threshold. Such definition of weights has the advantage that if even after an artifact removal stage, there are a few artifact in a session, their effect will be mitigated because they will get small weights. Note that artifact are usually irregular beats that do not look like normal heartbeats and hence usually get small scores.

We force the enrollment session to include a minimum of 22 heartbeats. This sets the minimum length of an enrollment session to about 20 seconds. Regarding the ECG template updating, we set the minimum length of the first test session 12 heartbeats, i.e. about 10 seconds. This allows a more effective template updating specially when HC threshold is set

![Figure 2.4: Recording time for different subjects for HC=0.92. For each subjects 3 values corresponding to 3 test sessions are marked with circles. The average recording time is also shown as a solid red line.](image-url)
to a small value. However, the length of the subsequent test sessions, i.e. second and third sessions, are determined solely by HC threshold and a 30-second upper limit. Heartbeats with very small HC are usually very irregular due to muscle or electrode movements. Therefore, heartbeats with HC below 0.5 are considered as outlier and discarded.

Figure 2.5 shows EER of ECG versus average length of ECG recording. Different recording length is achieved through changing the HC threshold as suggested in Figure 2.3. It can be seen that increasing the recording time improves the performance of ECG recognition. By using the proposed stopping criterion different sessions can have different length. For comparison, the EER of the conventional approach that has a fixed session length is also presented in Figure 2.5. It can be seen that the proposed stopping criterion effectively reduces the recording time needed to achieve a desired EER. As the recording time approaches 30 seconds, HC criteria is dominated by the 30-second limit and converges to the conventional approach. Recording time beyond 30 seconds is not considered in our experiments because it is of little value in a real world scenario. However, increasing the maximum recording time to 1 minute leads to an average EER of 3.75% and further increasing to 2 minutes does not improve the performance anymore. Since not all subjects have more than 2 minutes recording, we did not go beyond 2 minutes.

Figure 2.5: ECG recognition performance versus the average recording time. EER of ECG recognition using the proposed stopping criteria HC to determine the length of each session is compared with ECG recognition without HC stopping criteria, i.e. same recording length for all sessions. EER is averaged over all test sessions.
2.3.4 Fingerprint Liveness Detection (FpLD)

We examined 6 different fingerprint liveness detection algorithms including LCPD\(^1\) [74], BSIF\(^2\) [73], IQA\(^3\) [67], LBP\(^4\) [70], LPQ\(^5\) [71] and WLD\(^6\) [72]. Performance of these methods are provided in the right side of Table 2.2, 2.3 and 2.4 corresponding to 3 different scanners. One may observe that on average the performance of the existing methods are not acceptable for many practical applications. Moreover, all these methods suffer from having a large performance variation when encountering different novel spoofs. It can be seen that on average LCPD outperforms the other liveness detection methods. Therefore, we use it in the rest of experiments on fusion and template updating which will be explained later.

\(^1\)http://www.grip.unina.it/web-download.html
\(^2\)http://www.ee.oulu.fi/jkannala/bsif/bsif.html
\(^3\)Codes for IQA method was obtained directly from the author.
\(^4\)http://www.cse.oulu.fi/wsgi/CMV/Downloads/LBPSoftware
\(^5\)http://www.cse.oulu.fi/wsgi/CMV/Downloads/LPQMatlab
\(^6\)http://www.cse.oulu.fi/wsgi/CMV/Research/
Table 2.2: EER (in percent) of MmLD (fusion of ECG and LCPD) on DigitalPersona dataset compared with 6 state-of-the-art liveness detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing methods.

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>Fusion of ECG and LCPD</th>
<th>No Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum</td>
<td>Multiplication</td>
</tr>
<tr>
<td>Ecoflex</td>
<td>3.6</td>
<td>6.4</td>
</tr>
<tr>
<td>Gelatine</td>
<td>6.6</td>
<td>8.7</td>
</tr>
<tr>
<td>Latex</td>
<td>3.7</td>
<td>3.9</td>
</tr>
<tr>
<td>LiqEcoflex</td>
<td>7.3</td>
<td>8.2</td>
</tr>
<tr>
<td>RTV</td>
<td>4.8</td>
<td>6.5</td>
</tr>
<tr>
<td>WoodGlue</td>
<td>9.9</td>
<td>10.3</td>
</tr>
<tr>
<td>Average</td>
<td>6.0</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Table 2.3: EER (in percent) of MmLD (fusion of ECG and LCPD) on GreenBit dataset compared with 6 state-of-the-art liveness
detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing
methods.

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>Fusion of ECG and LCPD</th>
<th>No Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum</td>
<td>Multiplication</td>
</tr>
<tr>
<td>Ecoflex</td>
<td>1.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Gelatine</td>
<td>3.9</td>
<td>7.4</td>
</tr>
<tr>
<td>Latex</td>
<td>1.9</td>
<td>3.6</td>
</tr>
<tr>
<td>LiqEcoflex</td>
<td>4.1</td>
<td>8.1</td>
</tr>
<tr>
<td>RTV</td>
<td>2.3</td>
<td>5.0</td>
</tr>
<tr>
<td>WoodGlue</td>
<td>8.5</td>
<td>10.1</td>
</tr>
<tr>
<td>Average</td>
<td>3.8</td>
<td>6.3</td>
</tr>
</tbody>
</table>
Table 2.4: EER (in percent) of MmLD (fusion of ECG and LCPD) on Biometrika dataset compared with 6 state-of-the-art liveness detection algorithms. Results for 3 different ECG recording times are provided. The last row shows the average EER over the 6 spoofing methods.

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>Fusion of ECG and LCPD</th>
<th>No Fusion</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sum</td>
<td>Multiplication</td>
<td>Max</td>
</tr>
<tr>
<td>Ecoflex</td>
<td>2.5 2.1 1.7</td>
<td>5.2 4.1 2.9</td>
<td>3.1 3.0 3.0</td>
</tr>
<tr>
<td>Gelatine</td>
<td>4.1 3.6 3.3</td>
<td>7.8 5.8 5.0</td>
<td>9.1 9.2 8.9</td>
</tr>
<tr>
<td>Latex</td>
<td>1.6 1.3 1.1</td>
<td>3.0 2.0 1.3</td>
<td>1.7 1.7 1.8</td>
</tr>
<tr>
<td>LiqEcoflex</td>
<td>3.9 3.3 2.7</td>
<td>7.3 5.4 4.6</td>
<td>7.6 7.7 7.1</td>
</tr>
<tr>
<td>RTV</td>
<td>3.8 3.2 2.8</td>
<td>7.9 5.3 4.5</td>
<td>9.0 9.0 8.8</td>
</tr>
<tr>
<td>WoodGlue</td>
<td>7.8 6.1 4.2</td>
<td>9.6 6.8 4.6</td>
<td>14.4 14.5 14.6</td>
</tr>
<tr>
<td>Average</td>
<td><strong>3.9 3.3 2.6</strong></td>
<td><strong>6.8 4.9 3.8</strong></td>
<td><strong>7.5 7.5 7.4</strong></td>
</tr>
</tbody>
</table>
2.3.5 Multimodal Liveness Detection (MmLD)

ECG has an inherent liveness detection characteristic. While an intruder can easily provide the system with a fake fingerprint, they cannot provide an authentic ECG sample. We use LCPD as the state-of-the-art method in liveness detection and fuse it with ECG. The resulting multimodal liveness detection module, referred as MmLD, can be realized through different fusion rules. We consider weighted sum, multiplication and maximum rules. In the weighted sum rule, weights can be determined using an evaluation set. However, a simple approach to determine the weights is to choose them proportional to the EER of the individual traits. In our experiments, since EER of ECG is roughly half of the EER of LCPD, we set the weight of ECG and LCPD to $\frac{2}{3}$ and $\frac{1}{3}$, respectively. As shown in Table 2.2, 2.3 and 2.4, the sum rule outperforms the other fusion rules so we use it for template updating experiments that will be discussed in section 2.4. In addition to having a better average EER over different spoofs, the proposed method has the advantage of having a relatively small performance variation among different spoofs. Therefore, compared to conventional liveness detection methods, the proposed method has a better generalization on novel spoofs.

2.3.6 Fingerprint Recognition (FpR)

We use the fingerprint recognition software from the NIST Biometric Image Software (NBIS 5.0.0) [82]. NBIS detects minutiae by an algorithm called MINDTCT and computes fingerprint matching scores using a matcher called Bozorth3. MINDTCT performs image binarization, minutiae detection, false minutiae removal, neighbor ridges counting and minutiae quality assessment, and generates a list consists of location, orientation, type and quality of the detected minutiae to be used by Bozorth. Bozorth is a minutiae-based fingerprint matching system invariant to rotation and translation. It considers the location and orientation of the top 150 high quality minutiae and computes the matching score. EER of Bozorth on 3 different datasets is shown in the last column of Table 2.5. Note that while the Bozorth’ EER on Greenbit dataset is small, i.e. %1.4, it performs poorly on DigitalPersona and Biometrica datasets. This can be attributed to low quality of samples in those datasets and that they are constructed primarily for liveness detection rather than fingerprint recognition.
Table 2.5: EER (in percent) of MmHR (fusion of ECG and Bozorth) for different fusion rules and different values of HC threshold.

<table>
<thead>
<tr>
<th>Rec. Time</th>
<th>Sum</th>
<th>Multiplication</th>
<th>Max</th>
<th>Bozorth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.5</td>
<td>10</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>DigitalPersona</td>
<td>3.7</td>
<td>3.2</td>
<td>1.7</td>
<td>7.3</td>
</tr>
<tr>
<td>GreenBit</td>
<td>2.4</td>
<td>2.0</td>
<td>1.1</td>
<td>5.9</td>
</tr>
<tr>
<td>Biometrika</td>
<td>4.7</td>
<td>3.8</td>
<td>2.4</td>
<td>8.0</td>
</tr>
</tbody>
</table>

2.3.7 Multimodal Human Recognition (MmHR)

We consider three different rules for fusion of ECG and fingerprint recognition, i.e. sum, multiplication and maximum rules and results are presented in Table 2.5. Since the performance of Bozorth and ECG recognition modules are on average in the same range, we use equal weights for both traits in sum rule. Comparing the results of different scanners in Table 2.5, one may observe that while Bozorth gives an EER as good as %1.4 on GreenBit scanner, its performance degrades to %15.7 on Biometrika scanner. On the other side, performance of multimodal system is less sensitive to the scanner, i.e. a better generalization on different scanners. We use sum rule for fusion of ECG and Bozorth in the subsequent experiments on template updating due to its better performance compared to multiplication and max rules.

2.4 Automatic Updating of ECG and Fingerprint Templates

2.4.1 Overview

Automatically updating the ECG templates allows to model the intra-class variation of ECG over different sessions. So the biometric system can adapt to temporal variations of ECG signal across different sessions. If the system ensures the liveness and genuineness of a query, then the provided ECG and fingerprint samples can be safely added to the corresponding templates. To avoid adaption using impostor samples, ideally zero false acceptance rate for both liveness detection and human recognition tasks is necessary. This can be achieved in two different ways. First, the fusion score of MmLD and MmHR blocks, i.e. $S_{MmLD}$ and $S_{MmHR}$ can be compared against updating thresholds:

$$(S_{MmLD} > T_{MmLD}^U) \& (S_{MmHR} > T_{MmHR}^U)$$
where $T_{MmLD}^U$ and $T_{MmHR}^U$ are thresholds corresponding to zero false acceptance rate operating point of MmLD and MmHR blocks, respectively and $\&$ is logical AND operator. Those subjects that pass the MmLD block are fed to the MmHR block and those that pass that block too are selected for updating.

In the second approach, each of the ECG, FpLD and FpR blocks are set to zero false acceptance operating point. Considering that inside MmLD (MmHR) block, ECG and FpLD (FpR) are in parallel configuration, the following criterion can be used to achieve the overall zero false acceptance rate:

$$((S_{ECG} > T_{ECG}^U) | (S_{FpLD} > T_{FpLD}^U)) \& ((S_{ECG} > T_{ECG}^U) | (S_{FpR} > T_{FpR}^U))$$

where $S_{ECG}$, $S_{FpLD}$ and $S_{FpR}$ are the scores and $T_{ECG}^U$, $T_{FpLD}^U$ and $T_{FpR}^U$ are the updating thresholds corresponding to zero false acceptance rate operating point of ECG, FpLD and FpR blocks, respectively and $|$ is logical OR operator. Note that while in our experiments we use zero false acceptance rate operating point for template updating, if in some applications the number of subjects is very large, the updating thresholds may be set to a very small value instead of zero.

Both approaches ensure that fake samples as well as impostor attempts are rejected and will not contribute to template updating process. We use the second criterion because it is independent of the fusion rule inside MmLD and MmHR blocks. Queries that satisfy above criterion are selected for template updating. For every selected subject, ECG template is updated by adding the new ECG samples to the previously existing training samples of that subject. Since Bozorth is a matcher, fingerprint template is updated by adding the current test sample to the previously existing gallery samples.

Note that if i-th chimeric subject is selected for updating, and in the next session we are testing against the i-th chimeric subject, only the ECG and fingerprint templates involved in the i-th chimeric subject are updated. For ECG trait, only the positive class of the one-versus-all SVM classifiers is updated and the negative class that includes all other subjects only contains samples of one session i.e. no updating for negative class. This is to ensure that the process for i-th chimeric subject is independent of other chimeric subjects that may have either ECG or fingerprint in common. In other words, updating process is personalized for each subject. Fingerprint template updating is similar. The only difference is that since Bozorth is a matcher,
similarity of the new test sample to gallery samples are computed and averaged to form a single score corresponding to that test fingerprint sample.

### 2.4.2 Effect of Adaption on ECG and Fingerprint Recognition

At the end of the first test session, on average about 82% of chimeric dataset’s subjects are selected for template updating. Likewise, at the end of the second test session about 80% are selected. In Figure 2.6, the effect of template updating on ECG and fingerprint recognition is shown for 2nd and 3rd test sessions. It can be seen that the performance of both ECG recognition, and fingerprint recognition using Bozorth improves due to template updating.

![Chart](image1)

![Chart](image2)

**Figure 2.6:** EER of ECG and fingerprint recognition with and without template updating: (a) 2nd test session. (b) 3rd test session. Template updating improves both ECG recognition and fingerprint recognition performance. Results are averaged over 3 datasets. Solid and dashed lines denote with template updating (adaptive) and without template updating, respectively.
Chapter 2. Fusion of ECG and Fingerprint for Spoof Detection

2.5 Comparison with Other Methods

2.5.1 Comparison with Conventional Liveness Detection Approaches

We compare the liveness detection performance of the proposed system with 6 state-of-the-art liveness detection algorithms. Figure 2.7 shows the EER of the proposed method in the third test session after performing automatic template updating in the first and second test sessions. EER of our comparison methods are also shown as horizontal lines. It can be seen that the proposed approach performs significantly better for a wide range of ECG length.

2.5.2 Comparison with Participants of LivDet2015 Competition

We also compare the performance of the proposed method with participants of LivDet2015 liveness detection competition [1]. LivDet2015 test set includes 6 different spoofs, among them LiquidEcoflex and RTV are not in the training set. Therefore they can be considered as unknown spoofs that have similar training set as our experiments. For the sake of fair comparison, throughout this section we consider only LiquidEcoflex and RTV as unknown spoofs for testing. Table 2.6 shows the half total error rate (HTER) for participants of LivDet2015 competition on aforementioned unknown spoofs as well as EER of the proposed method for various length of ECG recording. For example ECGFP-5 denotes the proposed method with 5 seconds of ECG. Results for LivDet2015 participants in Table 2.6 are reported from [1].

Among the 6 liveness detection algorithms implemented in this research, LCPD with EER of 15.3%, 9.3% and 8.1% on the 3 datasets has the best performance which is in the range of top
Chapter 2. Fusion of ECG and Fingerprint for Spoof Detection

Table 2.6: Comparison with LivDet2015 liveness detection competition. HTER (in percent) is reported for all participants of LivDet2015 competition from [1]. EER (in percent) for the MmLD (fusion of ECG and LCPD) with template updating for different ECG lengths in the 3rd test session is also presented. To be consistent with LivDet2015 competition, only LiquidEcoflex and RTV are considered as unknown materials.

<table>
<thead>
<tr>
<th>LivDet2015 Competition</th>
<th>Dig.Persona</th>
<th>GreenBit</th>
<th>Biometrika</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPILHA</td>
<td>24.2</td>
<td>30.6</td>
<td>32.7</td>
<td>29.1</td>
</tr>
<tr>
<td>CSI</td>
<td>26.2</td>
<td>20.3</td>
<td>17.1</td>
<td>21.2</td>
</tr>
<tr>
<td>CSI MM</td>
<td>26.4</td>
<td>14.4</td>
<td>11.2</td>
<td>17.3</td>
</tr>
<tr>
<td>hbirkholz</td>
<td>13.7</td>
<td>11.0</td>
<td>7.1</td>
<td>10.6</td>
</tr>
<tr>
<td>hector</td>
<td>19.3</td>
<td>12.0</td>
<td>15.0</td>
<td>15.4</td>
</tr>
<tr>
<td>anonym</td>
<td>18.5</td>
<td>11.4</td>
<td>10.8</td>
<td>13.5</td>
</tr>
<tr>
<td>Jinglian</td>
<td>15.1</td>
<td>12.0</td>
<td>5.3</td>
<td>9.3</td>
</tr>
<tr>
<td>UFPE I</td>
<td>23.7</td>
<td>28.0</td>
<td>41.6</td>
<td>31.1</td>
</tr>
<tr>
<td>UFPE II</td>
<td>24.7</td>
<td>18.6</td>
<td>32.7</td>
<td>25.3</td>
</tr>
<tr>
<td>nogueira</td>
<td>7.1</td>
<td>5.5</td>
<td>7.2</td>
<td>6.6</td>
</tr>
<tr>
<td>titanz</td>
<td>12.6</td>
<td>10.5</td>
<td>7.6</td>
<td>10.2</td>
</tr>
<tr>
<td>unina</td>
<td>18.2</td>
<td>5.3</td>
<td>6.2</td>
<td>9.9</td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECGFP-5</td>
<td>6.4</td>
<td>3.0</td>
<td>3.9</td>
<td>4.4</td>
</tr>
<tr>
<td>ECGFP-7.5</td>
<td>5.2</td>
<td>2.3</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>ECGFP-10</td>
<td>5.0</td>
<td>2.2</td>
<td>3.0</td>
<td>3.4</td>
</tr>
<tr>
<td>ECGFP-30</td>
<td>3.4</td>
<td>1.3</td>
<td>2.0</td>
<td>2.2</td>
</tr>
</tbody>
</table>

algorithms participated in LivDet2015 competition. If we had an algorithm better than LCPD, we could even get better results. Note that although in this study we investigate fusion of ECG and a software-based liveness detection method, a hardware-based method can also be used in the same way because the proposed system performs the fusion on score level. In addition to 12 participants listed in Table 2.6, one complete fingerprint system is also submitted to LivDet2015 competition. Participants were provided by three spoof receipts and their submitted fingerprint system were tested on those spoof types as well as two unknown spoof materials. HTER of the submitted system was 8% as reported in [1] which is by far behind the proposed method.

In addition to the LiVDet2015 competition’s final report [1], recently organizers of LivDet series have presented a summary of LivDet competitions in [2]. Regarding the LivDet2015 competition, they have selected 3 of the 12 participants of the competition and reported rate of misclassified fake fingerprints (ferrfake) when rate of misclassified live fingerprints (ferrlive) is 1%. This represents the percent of spoof attacks that have been able to fool the system when only 1% of live attempts are mistakenly rejected. This results are reported from [2] in Table
Table 2.7: ferrfake (in percent) at ferrlive=1%. Results for 3 participants of LivDet2015 is reported from [2]. ferrfake of the MmLD (fusion of ECG and LCPD) with template updating for different ECG lengths in the 3rd test session is also presented. To be consistent with LivDet2015 competition, only LiquidEcoflex and RTV are considered as unknown materials.

<table>
<thead>
<tr>
<th>LivDet2015 Competition</th>
<th>Dig.Persona</th>
<th>GreenBit</th>
<th>Biometrika</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>unina</td>
<td>52.2</td>
<td>53.2</td>
<td>19.8</td>
<td>41.7</td>
</tr>
<tr>
<td>nogueira</td>
<td>25.3</td>
<td>23.4</td>
<td>20.4</td>
<td>23.0</td>
</tr>
<tr>
<td>anonym</td>
<td>84.0</td>
<td>75.9</td>
<td>57.1</td>
<td>72.3</td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECGFP-5</td>
<td>15.2</td>
<td>7.0</td>
<td>13.9</td>
<td>12.1</td>
</tr>
<tr>
<td>ECGFP-7.5</td>
<td>11.0</td>
<td>2.7</td>
<td>6.2</td>
<td>6.6</td>
</tr>
<tr>
<td>ECGFP-10</td>
<td>10.0</td>
<td>2.6</td>
<td>4.6</td>
<td>5.7</td>
</tr>
<tr>
<td>ECGFP-30</td>
<td>5.3</td>
<td>1.6</td>
<td>2.9</td>
<td>3.3</td>
</tr>
</tbody>
</table>

2.7. For comparison, ferrfake of the MmLD (fusion of ECG and LCPD with template updating) in the 3rd test session is also presented. The average ferrfake of these methods at ferrlive=1% is about 45%. This indicates the poor performance of the state-of-the-art liveness detection methods that is also pointed out in [2]. On the other side, the proposed method has been able to provide, for example, a ferrfake of 6.6% at ferrlive=1% incorporating only 7.5 seconds of ECG which is by far better that the state-of-the-art liveness detection methods.

2.5.3 Recognition Performance: Comparison with Bozorth

Figure 2.8 compares the recognition rate of the proposed system with Bozorth in the third test session. It can be seen that the proposed method performs significantly better than Bozorth. Note that the main focus of this study is liveness detection rather than recognition rate. We do not emphasis on fingerprint recognition performance and use Bozorth only as an average commercial fingerprint recognition system. If we had a more accurate and complicated fingerprint recognition algorithm, the results after fusion with ECG was even better. Therefore, we have considered a more challenging case. This is important because it demonstrates that two average low cost commercial fingerprint and ECG biometric systems with mediocre individual performances can be combined to provide significantly better performance. Note that our ECG signals are recorded using Vernier [81] which is a low cost nonprofessional commercially available hardware designed for educational use. Nevertheless, we have demonstrated promising results
under such a practical setup. One may get even better results by using a more complicated fingerprint recognition method or more sophisticated ECG acquisition hardware. But for the sake better generalization on real-world scenarios, we didn’t seek that direction.

2.6 Chapter Summary

ECG can be recorded from fingertips. Therefore, fingerprint is the natural choice to be fused with ECG. On the other side fingerprint is vulnerable to spoof attacks and ECG has inherent liveness detection. This chapter presented a unified approach for fusion of fingerprint and ECG that fills the gap between these two sides. To get the most out of ECG, the proposed system fuses ECG with a conventional fingerprint liveness detection method for a better liveness detection performance, and also fuses it with a fingerprint recognition method for a better recognition rate. Comprehensive experiments on LiVDet2015 database with 6 different spoofs, 3 different scanners, 6 state-of-the-art liveness detection algorithms and 3 different fusion methods have demonstrated that ECG significantly improves the performance in both tasks, i.e. liveness detection and biometric recognition. In addition, the proposed system automatically performs template updating, so that the performance of biometric system can be maintained in long term without manual re-training or re-enrollment. We also compared the proposed system with the results reported in LivDet2015 competition which is the latest competition in this field and demonstrated that it outperforms all participants of that competition.
Chapter 3

Feature Selection for Non-stationary Data

3.1 Introduction

In this chapter, we proposed a feature selection method for non-stationary signals such as ECG. The proposed method aims to build “templates” from samples of only one enrollment session such that the templates are robust against signal variations over different sessions. Realizing this goal requires extra information in addition to what can be extracted from the enrollment session during “enrollment” phase. The proposed approach, during a “development” phase, produces this information from an “auxiliary” dataset consists of recordings of some general subjects different from the biometric system users where recordings from multiple sessions are available for each member of the auxiliary dataset. This information is produced in the context of feature selection. The process of selecting a subset of informative features to be used for model construction is known as feature selection. Our feature selection criteria selects features that are more persistent across different sessions. The desired persistent features are selected from the auxiliary dataset. These features are then used as a prior knowledge in building the users templates.

The proposed feature selection criteria is formulated as a logistic regression optimization problem that can be efficiently solved through gradient descent. It makes no assumption re-
Chapter 3. Feature Selection for Non-stationary Data

Regarding the data distribution. The proposed method allows us to accommodate across session variations. None of the previous methods [23–29, 31–33, 35, 36, 38] have the advantage of handling across session variations and injecting a prior knowledge for building subjects templates that have only one enrollment session. We refer to the proposed algorithm as Multisession Feature Selection (MSFS).

An interesting characteristic of the ECG signals used in this research thesis is that they are collected from thumbs using dry electrodes (see Fig. 3.1), contrasting with most of the previous studies that are based on signals collected from chest using adhesive gel electrodes (e.g. PTB and MIT databases [16]). Recording ECG from fingertips is much more realistic for biometric application compared to recording from chest area in which the user has to undress for electrode placement. Adhesive gel electrodes have a very good signal to noise ratio because of their low electrode/skin impedance. In comparison, dry electrodes have lower signal quality due to unstable electrode/skin interface. Therefore, signals recorded from fingertips using dry electrodes are noisier and harder to deal with [83]. Moreover, unlike the previous ECG datasets [84], [6], [85], [37], [26], [86], [87], our dataset allows us to consider challenges related to ECG variation across different postures. In this study, in addition to analyzing ECG signals recorded at different times (scheduled at least one week apart), we consider across posture variations.

Furthermore, we demonstrate the effectiveness of the proposed method on the Transient Evoked Otoacoustic Emissions (TEOAE) database collected in BioSec lab at University of Toronto which includes recordings in 2 separate sessions scheduled at least one week apart. TEOAE is an acoustic response which is generated from within the inner ear after a low level transient click stimulus. It is related to inner ear structure and can be recorded using an earphone with built-in microphone. TEOAE has the following key characteristics: It can be recorded non-invasively. Speaker and microphone can be easily integrated into a headset or earphone connected to a smart phone to be used for access control or personalization of service (see Fig. 3.1). TEOAE presents in almost every living individual (99%+) [88]. So, it’s measurability is always guaranteed. Also, it might be useful in identification of newborn infants where performance of conventional methods based on face, fingerprint and iris degrade [89–91]. Clinically, TEOAE is simple, non-invasive yet effective in detecting hearing problems in newborn babies who are too young to cooperate in conventional hearing tests [92], and it has been
extensively used in infant screening programs around the world [93].

Note that all the previously existing TEOAE datasets (e.g. [94] and [95]) were recorded in the same session which is of little biometric value because such datasets do not allow to verify if the recognition algorithm does the job in a different session. In those datasets, both training and testing samples were recorded in the same session by either recording one signal and then dividing it into two parts (one for training and one for testing), or reinserting the earphone and capturing a second recording.

The remaining portion of this chapter is organized as follows: Section 3.2 briefly reviews recent works in area of TEOAE biometrics and feature selection. Note that prior works on ECG recognition has been previously reviewed in section 1.4. Details of the proposed feature selection algorithm are presented in Section 3.3. In section 3.4, an approach to speed up the proposed algorithm is presented. Experimental results are presented in section 3.5 and section 3.6 concludes the chapter.

3.2 Related Work

3.2.1 Transient evoked otoacoustic emission

The first study on feasibility of using TEOAE for human recognition was done by Swabey et al. [96] which uses power spectrum features and Euclidean distance for classification. In another work [94] by the same group, maximum likelihood estimation (MLE) is used to estimate probability density function (PDF) of inter-class and intra-class distance distributions. In [97] pair-wised distances in cepstrum-transform domain is used. A summary of above works is presented in [95]. The main drawback of these works is that due to difficulty of recording in
separate sessions, both testing and training recordings were collected in the same session or with only removing and refitting the earphone which limits its application for biometric purpose.

To address this limitation, the first multisession dataset of TEOAE signals was collected under a setup suitable for biometric evaluation [98] and bivariate empirical mode decomposition (BEMD) were employed for signal decomposition. Later, [99] used continuous wavelet transform (CWT) for signal representation. The main drawback of these methods is that they rely on expert’s knowledge to engineer features which were often chosen for their empirical performance, rather than a specific theoretical property related to across session variations.

3.2.2 Feature selection

Feature selections approaches can be divided into two categories: supervised and unsupervised. Supervised approaches use label information to guide the selection process whereas unsupervised approaches aims to describe structure of data in some feature space in the absence of label information [100,101]. In this study, we focus on the supervised approaches because in the biometric recognition application class labels are provided.

Supervised feature selection methods can be roughly categorized into filters, wrappers and embedded methods. Wrapper methods, such as sequential forward selection and sequential backward selection [102], select a feature subset (among all possible feature subsets) that gives the best performance [103]. A model for each candidate feature subset is trained and then tested and its performance is used to guide the feature selection process. Various searching techniques have been proposed in the literature. For example, in [104] particle swarm optimization is used as a search technique. However, wrappers are computationally very intensive specially if the chosen model is complex. The other drawback of such algorithms is the high risk of over-fitting because they are tuned to a specific model. While wrapper methods use a model, i.e. a specific classification algorithm, for evaluation of a feature subset, filter methods use a criterion function. The third category, embedded methods, embeds feature selection in classification [105,106]. However, such integration does not fit in our proposed approach because in the proposed approach feature selection and classification have to be performed on separate set of data. Data for feature selection have to be collected from multiple sessions but data for classifier training have to be collected from a single session to comply with “single session enrollment,
across-session test” scenario. More details on the proposed method will be provided in the next section.

On the other side, filter methods are relatively fast and do not suffer from aforementioned limitations. Different criterion functions have been proposed in the literature. Some feature selection approaches are based on mutual information and usually use some heuristics to handle the relevance-redundancy trade-off [107–111]. In [111] a minimum redundancy maximum relevance criterion is used which is based on mutual information. In [112], a greedy algorithm is presented that adds features to the current subset according to the selected and remaining features relevances with the labels.

There are some other approaches which are based on evolutionary algorithms. In [113] feature selection is realized by differential evolution to reduce the search space. Some other feature selection algorithms are based on a maximum margin criteria [114], [115], [116], [117], [118], [119]. These methods are sample-based where the “margin” is defined as the difference between distance to the nearest same class sample (near-hit) and the nearest sample from opposite classes (near-miss). Relief [116] selects features that are statistically relevant to the target values. The drawback of this method is that nearest miss and nearest hit samples are computed in the original space. This was addressed in Simba algorithm [115] through reevaluation of the margins in the induced space. However, its objective function is not convex and suffers from many local minima. Later, in [114] a local margin-based feature selection approach was presented in which uses a local learning approach to decompose a complex nonlinear problem into a set of locally linear problems within a large margin framework. Nevertheless, none of the aforementioned feature selection methods are designed to accommodate multisession data. Indeed, these are unable to benefit from the session label information provided in addition to the class labels in our problem. We previously presented a feature selection algorithm in section 2.3.3. However, that method treats features individually and does not consider the interaction between features. In this chapter, the idea of selecting features from a multi-session auxiliary dataset is realized in a more rigorous framework that does not suffer from aforementioned limitation.
3.3 Proposed feature selection method

Let $D = \{(x^{(i)}, y^{(i)}, s^{(i)})\}_{i=1}^M \subset \mathbb{R}^J \times \mathcal{Y} \times \mathcal{S}$ be an auxiliary dataset consists of $M$ samples belong to $N$ subjects and for each subject $S$ different sessions are available. $x^{(i)}$ is a $J$ dimensional feature vector, $\mathcal{Y} = \{Y_1, \ldots, Y_N\}$ is the set of all class labels and $y^{(i)} \in \mathcal{Y}$ is the class label of the $i$-th sample $x^{(i)}$. $S = \{S_1, \ldots, S_S\}$ is the set of all session labels and $s^{(i)} \in S$ is the session label of $x^{(i)}$. For example, $S$ may indicate recording sessions on $S$ different weeks, i.e., week 1, week 2, ..., week $S$.

Assume that $x^{(i)}$ is excluded from $D$ and we want to classify it while the corresponded subject, i.e. $y^{(i)}$, is enrolled using samples of only one session. However, there are $S$ possible options for enrollment session and we want $x^{(i)}$ to be correctly classified using either of them. Let adopt a classification rule such that given session $s \in S$ as enrollment, $x^{(i)}$ is correctly classified if and only if its corresponding margin $\ell^{(i)}_s$ is positive where $\ell^{(i)}_s$ is defined as:

$$\ell^{(i)}_s = 1^T d^{(i)}_s$$ (3.1)

where $d^{(i)}_s = d^{(i)}_{NM} - d^{(i)}_{NH}$ and $(\cdot)^T$ is transpose operator. $d^{(i)}_{NM}$ and $d^{(i)}_{NH}$ are absolute difference vectors determined as follows:

$$d^{(i)}_{NM} = |x^{(i)} - NM(x^{(i)})|,$$
$$d^{(i)}_{NH} = |x^{(i)} - NH_s(x^{(i)})|,$$ (3.2)

where $|.|$ is element-wise absolute operator, $NM(x^{(i)})$ is the nearest neighbor of $x^{(i)}$ with a different class label (nearest miss) and $NH_s(x^{(i)})$ is the nearest neighbor of $x^{(i)}$ from session $s$ with the same class label as $x^{(i)}$ (nearest hit). Note that equations (3.1) and (3.2) imply $l_1$ distance. While other options are possible (e.g. Euclidean distance by substituting $|.|$ with $(\cdot)^2$ in (3.2)), throughout this study we use $l_1$ distance. A basic definition of margin as the difference between distance to nearest miss and nearest hit was previously used in [116], [115] and [114], but it does not deal with the multiple session issue – Hence it is not appropriate for our problem.
A positive margin implies that the query sample \( x^{(i)} \) will be correctly classified. We realize the goal of having less misclassification error through weighting features by a nonnegative vector \( f \) and measuring margins in the corresponding weighted space specified by \( f \). A larger margin provides even more room for the query sample \( x^{(i)} \) to wander in the weighted space and still be correctly classified, i.e. a better generalization on unseen data. Therefore, we seek to maximize the margins in the “weighted” space. Equation (3.1) defines the margin in the original space. Similarly, in the weighted space, margin of sample \( x^{(i)} \) with respect to session \( s \) can be computed as follows:

\[
\ell_s^{(i)}(f) = f^T d_s^{(i)}. \tag{3.3}
\]

In fact \( f \) could be considered as a weighting vector that assigns higher weights to the relevant features where relevant features are those that results in larger margins. At the outset, \( f \) is unknown. Therefore, determining nearest miss and nearest hit in the weighted space defined by \( f \) is not possible. To overcome this issue, we use an iterative approach for computing \( f \), where at each iteration \( f \) is determined based on the distances in the weighted space defined at the previous iteration. Following [114], the problem of learning feature weights is formulated as a logistic regression problem and solved via gradient decent. However, determining \( \text{NM}(x^{(i)}) \) and \( \text{NH}_s(x^{(i)}) \) in the presence of irrelevant features, may not be accurate. To address this issue, margin is estimated as the expectation of \( \ell_s^{(i)}(f) \) over all possible candidates for \( \text{NM}(x^{(i)}) \) and \( \text{NH}_s(x^{(i)}) \) as follows:

\[
\bar{\ell}_s^{(i)}(f) = f^T \bar{d}_s^{(i)} \tag{3.4}
\]

where \( \bar{d}_s^{(i)} = \bar{d}_s^{(i)}_{NM} - \bar{d}_s^{(i)}_{NH_s} \); and \( \bar{d}_s^{(i)}_{NM} \) and \( \bar{d}_s^{(i)}_{NH_s} \) are:

\[
\bar{d}_s^{(i)}_{NM} = D_s^{(i)}_{NM} p^{(i)}_{NM}, \tag{3.5}
\]

\[
\bar{d}_s^{(i)}_{NH_s} = D_s^{(i)}_{NH_s} p^{(i)}_{NH_s}. \tag{3.6}
\]

\( D_s^{(i)}_{NM} \) and \( D_s^{(i)}_{NH_s} \) are matrices whose columns are absolute difference vectors with respect to \( x^{(i)} \):

\[
D_s^{(i)}_{NM} = \left( \left| x^{(i)} - x^{(M_t(1))} \right|, \ldots, \left| x^{(i)} - x^{(M_t(n))} \right| \right), \tag{3.7}
\]
\[ D_{NH_s}^{(i)} = \left( \left| x^{(i)} - x^{(H_s^i(1))} \right|, \ldots, \left| x^{(i)} - x^{(H_s^i(m))} \right| \right). \] (3.8)

\( \mathcal{M}^i \) and \( \mathcal{H}_s^i \) with cardinality of \( n \) and \( m \) denote set of all possible candidates for \( \text{NM}(x^{(i)}) \) and \( \text{NH}_s(x^{(i)}) \) respectively and are defined as:

\[ \mathcal{M}^i = \left\{ j \in \{1, \ldots, M\} \mid y^{(j)} \neq y^{(i)} \right\}, \] (3.9)

\[ \mathcal{H}_s^i = \left\{ j \in \{1, \ldots, M\} \mid y^{(j)} = y^{(i)}, s^{(j)} = s, j \neq i \right\}. \] (3.10)

\( \mathbf{p}_{NM}^{(i)} (\mathbf{p}_{NH_s}^{(i)}) \) in eq. (3.5) (eq. (3.6)) is an \( n \)-dimensional (\( m \)-dimensional) row vector indicates the probability of samples in \( \mathcal{M}^i (\mathcal{H}_s^i) \) being \( \text{NM}(x^{(i)}) (\text{NH}_s(x^{(i)})) \). Within the weighted space, samples situated closer to \( x^{(i)} \) are more probable to be the nearest sample. Therefore, it makes sense to define probabilities such that they are inversely related to distances. To this end, the probabilities can be determined as follows:

\[ \mathbf{p}_{NM}^{(i)} = \exp \left( \frac{-f^T D_{NM}^{(i)}}{\sigma} \right), \] (3.11)

\[ \mathbf{p}_{NH_s}^{(i)} = \exp \left( \frac{-f^T D_{NH_s}^{(i)}}{\sigma} \right), \] (3.12)

where \( \sigma \) is a user settable parameter. \( \mathbf{p}_{NM}^{(i)} \) and \( \mathbf{p}_{NH_s}^{(i)} \) are then normalized to sum to one to be the probabilities utilized in (3.5) and (3.6).

Having \( M \times S \) margins of the form \( \bar{\xi}_s^{(i)}(f) = f^T \bar{d}_s^{(i)}, s = 1, \ldots, S, i = 1, \ldots, M \), it is desired to maximize all margins. Considering a logistic regression formulation, the optimization problem can be expressed as follows:

\[ \max_f \sum_{i=1}^{M} \sum_{s=1}^{S} G(f^T \bar{d}_s^{(i)}), \text{ s.t. } f \geq 0, \] (3.13)

where \( G(\cdot) \) is a logistic function.

\[ G(b) = \log \left( \frac{1}{1 + \exp(-b)} \right) \] (3.14)
Input: $D = \{(x^{(i)}, y^{(i)}, s^{(i)})\}_{i=1}^{M}, \sigma, \lambda$

Output: $\{f\}$

1. Initialization: Set $f = (1, \ldots, 1)^T$;
2. repeat
3. $f_{\text{prev.}} = f$;
4. for $i \leftarrow 1$ to $M$ do
5. Compute $D^{(i)}_{NM}$ as in (3.7);
6. Compute $p^{(i)}_{NM}$ using $f_{\text{prev.}}$ as in (3.11);
7. Compute $\bar{d}^{(i)}_{NM}$ as in (3.5);
8. for $s \leftarrow 1$ to $S$ do
9. Compute $D^{(i)}_{NH_s}$ as in (3.8);
10. Compute $p^{(i)}_{NH_s}$ using $f_{\text{prev.}}$ as in (3.12);
11. Compute $\bar{d}^{(i)}_{NH_s}$ as in (3.6);
12. $\bar{d}^{(i)}_s = \bar{d}^{(i)}_{NM} - \bar{d}^{(i)}_{NH_s}$;
end
13. Compute $a$ through solving (3.17);
14. Set $f_j = a_j^2, 1 \leq j \leq J$
15. end
16. until $\|f - f_{\text{prev.}}\|_2 < \varepsilon$;

Algorithm 1: Pseudo code of the proposed MSFS feature selection algorithm.

$G$ is a strictly increasing function, therefore maximizing $G\left(f^T \bar{d}^{(i)}_s\right)$ indeed implies maximizing $f^T \bar{d}^{(i)}_s$. However, $G$ is useful because it can take an input that can vary from negative to positive infinity whereas the output always ranges between 0 and 1. Then (3.13) can be simplified as follows:

$$\min_f \sum_{i=1}^M \sum_{s=1}^S \log \left(1 + \exp \left(-f^T \bar{d}^{(i)}_s\right)\right), \quad \text{s.t. } f \geq 0, \quad (3.15)$$

A regularization term may also be added that has been shown to be useful in many applications to control over-fitting [120]. Hence, the problem can be formulated as follows:

$$\min_f \sum_{i=1}^M \sum_{s=1}^S \log \left(1 + \exp \left(-f^T \bar{d}^{(i)}_s\right)\right) + \lambda \|f\|_1, \quad \text{s.t. } f \geq 0, \quad (3.16)$$

where $\lambda$ is a user settable parameter. Equation (3.16) is a constrained convex optimization problem with respect to $f$. We then replace the vector $f$ with a new vector $a$ such that $f_j = a_j^2, 1 \leq j \leq J$ and reformulate the problem in an unconstrained form as follows:

$$\min_a \sum_{i=1}^M \sum_{s=1}^S \log \left(1 + \exp \left(-\sum_{j=1}^J a_j^2 \bar{d}^{(i)}_s(j)\right)\right) + \lambda \|a\|_2^2 \quad (3.17)$$
where \( \bar{d}_s^{(i)}(j) \) is the j-th element of vector \( \bar{d}_s^{(i)} \). Therefore (3.17) can be solved via gradient descent with step size of \( \tau \) and updated as follows:

\[
a \leftarrow a - \tau \Delta
\]

\[
\Delta = a \odot \left( \lambda 1 - \sum_{i=1}^{M} \sum_{s=1}^{S} \frac{\exp \left( - \sum_{j} a_j^2 \bar{d}_s^{(i)}(j) \right)}{1 + \exp \left( - \sum_{j} a_j^2 \bar{d}_s^{(i)}(j) \right)} \bar{d}_s^{(i)} \right)
\]

(3.18)

where \( \odot \) is Hadamard operator. We utilized fminbnd function in MATLAB to determine \( \tau \) using the line search algorithm. Features weight vector is initialized to 1, so that all features have the same weight at the beginning and then will be updated using (3.18) until a stopping criteria is satisfied. In this study, the algorithm stops when the difference between the weights in two successive iterations is less than a threshold \( \varepsilon \) which is set to 0.01. Pseudo code of the proposed MSFS feature selection method is presented in Algorithm 1.

The proposed method considers the “single session enrollment, across-session test” scenario and indeed, our objective function is directly related to the classification error under this scenario. This is an interesting characteristic of the proposed method that makes it a perfect choice for human recognition using medical biometrics such as ECG and TEOAE which are known to be non-stationary across different sessions compare to many other conventional biometrics. Therefore it is superior to all previously existing feature selection algorithms in the literature.

### 3.4 Speeding up using Random Projection

The most time consuming part of the proposed algorithm is computing distances to determine the nearest miss and nearest hit for every sample. In practice, about 2% of the CPU time is spent on solving (3.15) using gradient descent and the rest of CPU time is spend mainly on computing NM and NH\(_s\) which requires computing distance between \( x^{(i)} \) and all samples in \( M^i \) and \( H_s^i, s = 1, ..., S \) which is problematic if both \( M \) (number of samples) and \( J \) (dimensionality) have high values. However, majority of samples have very small probability of being the nearest neighbor because they are too far from \( x^{(i)} \). To speed up the process of determining NM and NH\(_s\), we use Random Projection to project the data onto a lower \( K \)-dimensional space and determine probabilities of being the nearest neighbor in the lower dimensional space and pick
a subset of top samples, i.e. high probable samples. Only selected samples are then used to
determine nearest neighbor in the $J$-dimensional space scaled by $f$ which results in a faster
processing time.

The idea of random projection is based on the Johnson-Lindenstrauss lemma [121] which
states that if points in a vector space are projected onto a randomly selected subspace of suitably
high dimension, then the distances between points are approximately preserved. Computation
of random projection is simple and involves only forming a random $J \times K$ matrix $R$ ($K \ll J$)
which its elements are often generated from Gaussian distribution. Later, Achlioptas [122]
showed that $R$ can be generated by a much simpler distribution such that its elements can be
drawn from $\{\sqrt{3}, 0, -\sqrt{3}\}$ with probability of $\frac{1}{6}$, $\frac{2}{3}$ and $\frac{1}{6}$ respectively.

Samples in the weighted space are projected to a $K$-dimensional space:

$$\hat{X}_f = R^T X_f,$$

where $X_f = (x_f^{(1)}, \ldots, x_f^{(M)})$ is data matrix in the weighted space, i.e. $x_f^{(j)} = f \otimes x^{(j)}$,
$j = 1, \ldots, M$. Computing matrices of absolute differences are much faster in the $K$-dimensional
space because $K \ll J$:

$$\hat{D}^{(i)}_{NM} = \left( |\hat{x}^{(i)}_f - \hat{x}_f^{(M^i(1))}|, \ldots, |\hat{x}^{(i)}_f - \hat{x}_f^{(M^i(n))}| \right)$$

$$\hat{D}^{(i)}_{NH_s} = \left( |\hat{x}^{(i)}_f - \hat{x}_f^{(H^i_s(1))}|, \ldots, |\hat{x}^{(i)}_f - \hat{x}_f^{(H^i_s(m))}| \right)$$

where $\hat{x}_f^{(j)}$ is projection of $x_f^{(j)}$ using $R$. Then probabilities of being the nearest neighbor can
be determined as follows:

$$\hat{p}^{(i)}_{NM} = \exp \left( -\frac{1^T \hat{D}^{(i)}_{NM}}{\hat{\sigma}} \right)$$

$$\hat{p}^{(i)}_{NH_s} = \exp \left( -\frac{1^T \hat{D}^{(i)}_{NH_s}}{\hat{\sigma}} \right)$$

where $\hat{\sigma}$ has the same role as $\sigma$ and is defined as $\sigma \times \left( \frac{K}{J} \right)$ to account for change in dimensionality.$\hat{p}^{(i)}_{NM}$ and $\hat{p}^{(i)}_{NH_s}$ are then normalized to sum to one. We form shortened lists of candidates for
$NM(x^{(i)})$ and $NH_s(x^{(i)})$ denoted by $\hat{M}^i$ and $\hat{H}_s^i$ where $\hat{M}^i \subset M^i$ and $\hat{H}_s^i \subset H_s^i$. To this end, we
sort members of $\hat{M}_i$ and $\hat{H}_s$ in descending order according to their corresponding probabilities $p_{NM}^{(i)}$ and $p_{NH_s}^{(i)}$. Then we pick the top $t$ members of $M^i (H^i_s)$ to form $\hat{M}_i (\hat{H}_s^i)$. $t$ is defined such that the summation of their corresponding probabilities are not less than a user-settable parameter $\gamma$. In all our experiments, $\gamma$ is fixed at the value of 0.95. We use $\hat{M}_i (\hat{H}_s^i)$ instead of $M^i (H^i_s)$ in (3.7) (3.8) to determine $D^{(i)}_{NM}$ ($D^{(i)}_{NH_s}$) which indeed has less number of columns and is faster to compute. Then $p_{NM}^{(i)}$ and $p_{NH_s}^{(i)}$ are computed using (3.11) and (3.12) based on the resulting $D^{(i)}_{NM}$ and $D^{(i)}_{NH_s}$. Other steps are the same as the MSFS. Pseudo code of the proposed algorithm based on random projection (hereafter MSFS-RP) is presented in Algorithm 2.

The time spent on projection in (3.19) is very small and can be ignored because in each iteration $\hat{X}(f)$ is computed only once. The time spent on computing $D^{(i)}_{NM}$, $D^{(i)}_{NH_s}$, $\hat{D}^{(i)}_{NM}$ and $\hat{D}^{(i)}_{NH_s}$ are dominant because they must be computed for every sample. Assume that the total number of samples in $\hat{M}_i$ and $\hat{H}_s$ is $L$. In MSFS algorithm, computing $D^{(i)}_{NM}$ and $D^{(i)}_{NH_s}$ is of the order $O(MJ)$. In MSFS-RP algorithm, $\hat{D}^{(i)}_{NM}$ and $\hat{D}^{(i)}_{NH_s}$ are computed in the lower dimensional space which is of the order $O(MK)$. In addition, $D^{(i)}_{NM}$ and $D^{(i)}_{NH_s}$ are to be computed on $\hat{M}_i$ and $\hat{H}_s^i$ which is of the order $O(LJ)$. Therefore, the complexity is of the order $O(MK + LJ)$. As will be discussed in section 3.5.9, $MK \ll LJ$ and this complexity for MSFS-RP is approximately $O(LJ)$ Therefore, the CPU time of MSFS-RP is approximately $\frac{M}{N}$ of the CPU time of MSFS.

### 3.5 Experimental results

In this section, we perform several experiments on synthetic and real-world datasets to demonstrate the effectiveness of the proposed method. The first two experiments evaluate the effectiveness of the proposed method to cope with variations of ECG and TEOAE across different sessions where our feature selection algorithm aims to select a subset of features which are more steady across sessions. In the third experiment, we use the proposed feature selection algorithm to select a subset of features which are more steady across different body postures. We consider the case that subjects are enrolled in sitting (standing) posture but tested in standing (sitting) posture. The experiments are designed with respect to two objectives:
Input: \( D = \{ (x^{(i)}, y^{(i)}, s^{(i)}) \}_{i=1}^{M}, \sigma, \lambda, K, \gamma \)

Output: \{f\}

1. Initialization: Set \( f = (1, \ldots, 1)^T \);
2. repeat
   3. \( f_{\text{prev.}} = f \);
   4. Compute \( \hat{X}_{f} \) using \( f_{\text{prev.}} \) as in (3.19);
   5. for \( i \leftarrow 1 \) to \( M \) do
      6. Compute \( \hat{D}^{(i)}_{NM} \) as in (3.20);
      7. Compute \( \hat{p}^{(i)}_{NM} \) as in (3.22);
      8. Form \( \hat{M}^{i} \subset M \) according to \( \hat{p}^{(i)}_{NM} \);
      9. Compute \( \hat{D}^{(i)}_{NM} \) as in (3.7) using \( \hat{M}^{i} \);
     10. Compute \( \hat{p}^{(i)}_{NM} \) using \( f_{\text{prev.}} \) as in (3.11);
     11. Compute \( \hat{d}^{(i)}_{NM} \) as in (3.5);
     12. for \( s \leftarrow 1 \) to \( S \) do
          13. Compute \( \hat{D}^{(i)}_{NHs} \) as in (3.21);
          14. Compute \( \hat{p}^{(i)}_{NHs} \) as in (3.23);
          15. Form \( \hat{H}^{i}_{s} \subset H^{i} \) using \( \hat{p}^{(i)}_{NHs} \);
          16. Compute \( \hat{D}^{(i)}_{NHs} \) as in (3.8) using \( \hat{H}^{i}_{s} \);
          17. Compute \( \hat{p}^{(i)}_{NHs} \) using \( f_{\text{prev.}} \) as in (3.12);
          18. Compute \( \hat{d}^{(i)}_{NHs} \) as in (3.6);
          19. Compute \( \bar{d}^{(i)}_{s} = \hat{d}^{(i)}_{NM} - \hat{d}^{(i)}_{NHs} \);
     20. end
   21. end
   22. Compute \( a \) through solving (3.17);
   23. Set \( f_{j} = a_{j}^{2}, 1 \leq j \leq J \);
 24. until \( \| f - f_{\text{prev.}} \|_{2} < \varepsilon \);


- Comparing performance of the proposed method against state-of-the-art methods in the area of ECG biometric recognition (AC/LDA [25], Odinaka [31] and Wang [38]) and TEOAE biometric recognition (CWT/LDA [99], BEMD [98] and PDF Estimation [94]).

- Comparing performance of the proposed method against existing feature selection algorithms in the literature. To this end, performance of the proposed method is compared against seven state-of-the-art feature selection algorithms including locally linear feature selection (LLFS) [114], iterative search margin based algorithm (Simba) [115], differential evolution feature selection (DEFS) [113], joint mutual information (JMI) [107], interaction capping (ICAP) [108], Markov random field (MRF) [123] and conditional infomax feature extraction (CIFE) [109].
3.5.1 Experimental Setup

Synthetic Dataset– Illustration of Multisession Concept

We use a synthetic dataset to illustrate the concept of feature selection for multisession data and the misclassification that may happen when enrollment and testing are from different sessions (as described in the beginning of section 3.3). This dataset consists of samples of 3 subjects and for each subject 2 sessions are available where each session includes 50 samples. Illustration of this dataset in the space defined by its four relevant features \( \{f_1, f_2\} \) and \( \{f_3, f_4\} \) is shown in Fig. 3.2 where two different sessions are shown in red and blue. In addition to the four relevant features, to have a more challenging problem, dataset is contaminated with 5000 irrelevant features drawn randomly from a Gaussian distribution with zero mean and unit variance. One may say that \( \{f_1, f_2\} \) are better than \( \{f_3, f_4\} \) because in the space defined by \( f_1 \) and \( f_2 \) samples of different sessions overlap but in the case of \( \{f_3, f_4\} \) such an overlap does not exist. Moreover, if we take the first session (blue) as enrollment and the second session (red) for testing, it can be seen that in the case of using \( \{f_1, f_2\} \), almost all samples will be correctly classified. But, using \( \{f_3, f_4\} \) results in misclassification of many samples. Misclassified samples are marked by a circle around them. We will later use this “toy” example to demonstrate the effectiveness of the proposed feature selection algorithm to select a subset of features which are more steady across sessions (\( f_1 \) and \( f_2 \) in this case).

Real-world Datasets

Performance of the proposed method is also demonstrated using three experiments on real-world datasets. In the first experiment, we use ECG database collected in our lab (BioSec) in University of Toronto [80]. There are 82 subjects that have 2 or more ECG recordings in sitting posture. 46 out of 82 subjects have exactly 5 sessions. Follow-up sessions are collected over a six months period. Unlike many previous ECG databases, signals were collected from fingertips as shown in Fig. 3.1. We use aforementioned 46 subjects as auxiliary dataset and the remaining 36 subjects with 2-4 recording sessions (on average 4.28 sessions) are used for enrollment and testing. Hence, there is no overlap between these two sets.

The purpose of having an auxiliary dataset is to select a subset of more persistent features.
Chapter 3. Feature Selection for Non-stationary Data

Figure 3.2: Visualization of across-session training and testing on the synthetic dataset consists of four features, i.e., $f_1$ to $f_4$ with 5000 additional irrelevant features. Blue and red samples were used for training and testing respectively. Classification result using $f_1$ and $f_2$ is shown on the left and result for $f_3$ and $f_4$ is shown on the right. Classification boundaries are shown with different background colors.

Therefore, it is important that the auxiliary dataset have as many sessions as possible. According to the above decomposition scheme, 5 sessions per subject in auxiliary dataset is available. It also divides the available 82 subjects into two approximately equal sets of 46 and 36 subjects, so that the size of both sets are in the range of the size of multisession datasets reported in literature [84], [6], [85], [37], [26], [86], [87].

In the second experiment, we use TEOAE dataset collected in our lab (BioSec) in University of Toronto [80]. Unlike previous TEOAE datasets collected primarily for diagnostic purpose (e.g. [94] and [95]), our dataset is collected solely for biometric purpose. Therefore, optimal test conditions like a clinical setup in a sound proof room is avoided and signals were collected in a regular office where there were people talking and entering or leaving the office. Signals from 54 subjects are collected for two sessions scheduled at least one week apart. Following [124], 30 subjects are randomly chosen as auxiliary dataset and the remaining subjects are used for enrollment and testing.

In the third experiment, we use the ECG database collected in our lab (BioSec) in University of Toronto [80]. There are 79 subjects that have recordings in both sitting and standing postures. Subjects in the auxiliary dataset are the same as those in the first experiment. All the remaining
subjects (33 subjects) are used for enrollment and testing.

**Performance measure**

Considering the “single session enrollment, across-session test” scenario described earlier, we pick a session for enrollment and use the rest of the available sessions for testing. That is, testing session can be any session in the past or future. For simplicity assume that there are P subjects and for each subject S sessions are available. For each of S sessions, we train P SVM classifiers in one-vs-all configuration. That gives a total of $P \times S$ templates. To simulate the positive trials (genuine attempts), each template is tested against $S - 1$ opposite sessions (same class), hence the term “across-session testing”. This gives a total of $P \times S \times (S - 1)$ positive trials. To simulate the negative trials (intruder attempts), each template is tested against $P - 1$ other subjects (opposite class) in $S - 1$ opposite sessions. That is $(P - 1) \times (S - 1)$ intruder attempts against each template. This gives a total of $P \times S \times (P - 1) \times (S - 1)$ negative trials. Note that we refer to the number of subjects by $P$ as oppose to $N$ to highlight that EER is computed using a separate dataset that does not overlap the auxiliary dataset. Auxiliary dataset is dedicated to feature selection task.

We use linear SVM classifier because it has been successfully employed for ECG biometric recognition in recent works [50,125,126]. For a fair comparison, auxiliary datasets, enrollment and testing sets are common for all algorithms. Algorithms were implemented in MATLAB and executed on a desktop with an Intel core i7-3770 CPU and 16GB RAM. MATLAB’s default settings for linear SVM classifier is used. More specifically, the penalty parameter is set to its default value i.e. 1. The number of selected features ranges from 1 to 1000 with a step size of 50. For each algorithm the appropriate number of selected features were determined using a bootstrapping algorithm [114] on the auxiliary dataset. For this purpose, each algorithm is run 10 times where at each run two subsets of equal size were randomly selected from the auxiliary dataset. One subset is used for feature selection and the other one is used for validation. The appropriate value for the number of features is the one that gives the minimum error on average over 10 runs. We use it for the final feature selection over entire auxiliary dataset.
Preprocessing

For ECG signals, following [13] and [25], a fourth order band-pass Butterworth filter between 0.5-40 Hz was employed for preprocessing. All signals are visually inspected before filtering. Each recording is segmented to its heartbeats using the method in [127] and [128] and aligned with respect to R peaks. For each session, Euclidean distance between segments and median is computed and segments whose distance are bigger than a threshold are considered as outlier and removed. Normalization is done by converting each feature to its z-score, i.e., we subtract the mean and divide by standard deviation so that each feature has zero mean and unit variance. Regarding the TEOAE signals, considering that Vivosonic Integrity System [129] used for signal recording has a built-in noise cancellation, no further preprocessing were needed. We only converted features to their z-score as described above.

Features’ Description

There has been a variety of representations suggested in the literature to represent medical signals. For example, STFT has been used in [31] for ECG recognition. Similarly, CWT has been used in [35], Autocorrelation has been used in [25] and some statistical quantities from different frequency bands has been used in [130]. To have a comprehensive feature pool and consider the potential advantages of each of these features, we form a pool containing all features and select a subset of features using a feature selection algorithm.

Features of the feature pool are as follows: Continuous wavelet transform (CWT) with Daubechies 5 as mother wavelet is computed on 1 second window centered at R peak location. STFT with Hamming window of length 16 with step size of 13 is also computed on 1 second window centered at R peak location. Autocorrelation is computed on windows of length 6 seconds with 80 lags. We also considered six frequency bands: 8-13Hz, 13-18Hz, 18-25Hz, 25-30Hz, 30-35Hz, 35-50Hz and for each band, mean of power, standard deviation of power, maximum amplitude, standard deviation of amplitude, kurtosis and skewness were computed. Maximum, standard deviation, kurtosis and skewness are also computed from the signal itself. The signal amplitude, i.e. a one second window centered around R peak, was also considered. Aforementioned feature vectors were concatenated to form the final feature vector of the length...
For TEOAE signals, we formed a similar feature pool with the addition of cepstrum features suggested in [97] for TEOAE which results in a feature vector of the length 7522. We only considered recordings from left ear which has been shown to be more challenging (compare to right ear) for biometric recognition [99] due to the ear asymmetries [131].

### 3.5.2 Experiments on across-session ECG recognition

In this section, performance of the proposed feature selection method is demonstrated on ECG signals when enrollment and testing are from different sessions. We use feature selection to find a subset of more steady features across different sessions. Length of each session is 2-5 minutes but we randomly selected 16 heartbeats to represent each session. This is a more challenging problem compare to the previous study in [57] which uses the whole session. While a variable ECG length strategy like what we presented in chapter 2 is possible, for simplicity in this chapter we set the number of heartbeats to 16. The interested reader is referred to [31] and [51] for further information on effect of number of heartbeats. Each experiment is repeated 5 times and the average EER and standard deviations are reported. We compare the performance of the proposed method against three existing methods for ECG biometric recognition: AC/LDA [25], Odinaka [31] and Wang [38]. Table 3.1 shows EER and standard deviation for these comparison methods. It can be seen that the proposed MSFS method has significantly better performance than the other methods. We have carefully implemented these methods because they are not publicly available. In AC/LDA method, the number of autocorrelation lags is set to 30 and window length is set to 6 seconds as suggested in [25]. The Wang method has two parameters, dictionary size and length of segments. These are set to 2000 and 64 respectively as suggested in [38].

Table 3.1: Across-session ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>14.4(1.1)</td>
<td>18.4(0.7)</td>
<td>20.3(1)</td>
<td>6.9(0.3)</td>
</tr>
</tbody>
</table>
3.5.3 Experiments on across-session TEOAE recognition

In this section, generalization of the proposed method on other medical biometrics is demonstrated by applying it on TEOAE biometric. Similar to across-session ECG experiment, feature selection is used to find a subset of more steady features across different sessions. Duration of each session varies among the subjects. The average length of recordings is about 39 seconds. On average, there are 122 samples for the left ear and 113 samples for right ear per session. Length of each sample is 17.2ms. Note that the actual raw samples of the length 17.2ms are not available as output. Samples undergo an internally implemented noise cancellation process which implies averaging over blocks of raw samples to produce output samples. Stimulus (click) interval is set to 21.12ms and as a result of internal noise cancellation process, the amount of time required to produce one sample at the output is about 338ms. Fig. 3.3 shows the first and last 10 samples for one subject. Note that responses are less noisier as the time proceeds.
Table 3.2: Across-session TEOAE experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CWT/LDA [99]</td>
<td>25.2(1.8)</td>
<td></td>
</tr>
<tr>
<td>BEMD [98]</td>
<td>22.6(1)</td>
<td></td>
</tr>
<tr>
<td>PDF Est. [94]</td>
<td>28.4(1.7)</td>
<td></td>
</tr>
<tr>
<td>MSFS</td>
<td>3.3(1.2)</td>
<td></td>
</tr>
</tbody>
</table>

due to the aforementioned noise cancellation. Unlike [99] that has used the last 10 samples of each session (requires the entire session), we randomly pick 10 samples from the first quarter of each session which in average reduce the length of each session to about 10 seconds. Each experiment is repeated 5 times and the average EER and standard deviations are reported.

We compare the performance of the proposed method against three state-of-the-art methods in the area of TEOAE biometric recognition including CWT/LDA [99], BEMD [98] and PDF Estimation [94]. Results are presented in Table 3.2 which demonstrate significant improvement compare to other methods. We have carefully implemented these methods because they are not publicly available. In CWT/LDA method, Daubechies 5 is used as mother wavelet and CWT scale is set to 7 as suggested in [99]. In BEMD method, following [99] the second most refined level of intrinsic mode function (IMF) is used because it provides the best results.

3.5.4 Experiments on across-posture ECG recognition

In this section, we demonstrate the generalization of the proposed method by applying it to signals collected in different postures. In the previous experiments, feature selection was used to cope with long-term variation of ECG and TEOAE over different sessions. However, such variations can also happen in short-term due to different body postures. Here, we use feature selection to cope with variation of ECG due to different body postures. We only consider sitting and standing postures because this is the most challenging case as indicated in the previous study in [57]. Similar to section 3.5.2, 16 heartbeats were randomly selected to represent each session. The proposed method is compared against three state-of-the-art methods and results are presented in Table 3.3. These results demonstrate the effectiveness of the proposed method in handling ECG variation due to different postures.
Table 3.3: Across-posture ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC/LDA [25]</td>
<td>12.2(1.4)</td>
<td></td>
</tr>
<tr>
<td>Odinaka [31]</td>
<td>13.7(1.1)</td>
<td></td>
</tr>
<tr>
<td>Wang [38]</td>
<td>18.3(1.6)</td>
<td></td>
</tr>
<tr>
<td>MSFS</td>
<td>3.7(0.9)</td>
<td></td>
</tr>
</tbody>
</table>

### 3.5.5 Comparison with prior art in feature selection

The results presented in Table 3.1, 3.2 and 3.3 shows significant improvement in the “single session enrollment, across session test” scenario. This demonstrates that the idea of utilizing an auxiliary multisession dataset is indeed effective in dealing with nonstationarity across sessions. Since this idea is realized through a novel features selection algorithm, the reader may be interested to compare MSFS algorithm with existing feature selection algorithms in the literature. To address this issue, performance of the proposed feature selection algorithm is also compared with seven feature selection algorithms in the literature and the results for all three experiments are presented in Table 3.4. The best result at each experiment is in bold. The last row is the average EER computed over all three experiments (i.e., row 1 to 3). This indicates that the proposed algorithm performs significantly better than the other comparison algorithms. For reference, EER for SVM without feature selection is also presented in Table 3.4. The code for our comparison feature selection algorithms are all available on the respective author’s websites and their parameters are set to their default values. To have a fair comparison, parameters of the proposed method are also set to default values and are discussed in Section 3.5.6.

In addition, receiver operating characteristic (ROC) curves for comparison algorithms which exhibited the best performance as well as the proposed MSFS method are shown in Fig. 3.4. The number of selected features for these methods (i.e. LLFS, Simba, MRF and MSFS) in across-session ECG experiment are respectively 50, 50, 200 and 50. For across sessions TEOAE experiment, these are 250, 350, 250 and 250 and for across posture ECG experiment, these are 150, 200, 300 and 100 respectively.
Table 3.4: Comparison of different feature selection algorithms: EER (in percent) and standard deviation (in percent) are reported for three experiments. Standard deviations are presented in parentheses. First and second rows are corresponded to across-session ECG and TEOAE experiments, respectively. Third row is corresponded to across-posture ECG experiment. Last row shows the average EER over the three experiments. Results without feature selection (No FS) is also appended.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG (Across-session)</td>
<td>10.6(0.8)</td>
<td>8.8(0.6)</td>
<td>10.1(0.5)</td>
<td>13.2(0.6)</td>
<td>12.6(0.2)</td>
<td>11.5(1)</td>
<td>12.4(0.2)</td>
<td>13.5(1)</td>
<td>6.9(0.3)</td>
</tr>
<tr>
<td>TEOAE (Across-session)</td>
<td>6.6(1.7)</td>
<td>6(1.7)</td>
<td>7.5(1.9)</td>
<td>8.9(1.4)</td>
<td>7.8(1.2)</td>
<td>6.8(1.9)</td>
<td>7.9(1.6)</td>
<td>8.8(1.3)</td>
<td>3.3(1.2)</td>
</tr>
<tr>
<td>ECG (Across-Posture)</td>
<td>7.9(1.3)</td>
<td>9(1.2)</td>
<td>8.3(1.1)</td>
<td>8.3(0.9)</td>
<td>7.9(0.8)</td>
<td>6.5(1.1)</td>
<td>7.3(1.2)</td>
<td>9.9(1.1)</td>
<td>3.7(0.9)</td>
</tr>
<tr>
<td>Average</td>
<td>8.37</td>
<td>7.93</td>
<td>8.63</td>
<td>10.13</td>
<td>9.43</td>
<td>8.27</td>
<td>9.20</td>
<td>10.73</td>
<td>4.63</td>
</tr>
</tbody>
</table>
3.5.6 Effect of parameters $\sigma$ and $\lambda$

The proposed feature selection algorithm has two parameters: $\sigma$ and $\lambda$. Parameter $\sigma$ controls the local behavior of the proposed feature selection algorithm. A large value of $\sigma$ increases the effect of far samples in estimating the nearest neighbor. The second parameter, $\lambda$, controls robustness against over-fitting. Fig. 3.5 and 3.6 show the feature weights computed by the proposed feature selection algorithm on the synthetic dataset for different values of $\sigma$ and $\lambda$. It can be seen that the feature weights do not change significantly over a relatively wide range of values for $\sigma$ and $\lambda$. The more stable features $f_1$ and $f_2$ indeed have greater weights compare to $f_3$ and $f_4$ that indeed suffer from across session variation (see Fig. 3.2). These results also assert that the proposed algorithm assigns very small weights to irrelevant features $\{f_5, \ldots, f_{5004}\}$. Generally these parameters can be determined using cross validation. However,
in our experiments on real world datasets we set these parameters to 2 and 6 respectively – i.e., default values and they are fixed during all experiments.

3.5.7 Margin improvement over successive iterations

In this section, we provide a demonstration of the property that the proposed method promotes positive margins. Fig. 3.7 shows the percentage of positive margins (for different values of $\sigma$ and $\lambda$) on the synthetic dataset. It can be seen that as the algorithm proceeds, the number of positive margins in the weighted space is increased. It demonstrates that the classification error discussed in section 3.3, related to the “single session enrollment, across session test” scenario, is reduced through iteratively updating the weights which is an important characteristic of the proposed method.
3.5.8 Effect of the number of sessions used in feature selection

The proposed method performs feature selection on an auxiliary dataset that consists of data from multiple sessions. In this section, we investigate the effect of changing the number of sessions incorporated in feature selection process. The auxiliary dataset in across session ECG experiment has 5 sessions. Table 3.5 shows the EER of the proposed method when feature selection is performed on varying number of sessions. It can be seen that increasing the number of sessions improves the performance. This can be attributed to better stability of selected features across sessions when more sessions are involved in feature selection process.

Table 3.5: Effect of changing the number of sessions used in feature selection.

<table>
<thead>
<tr>
<th>Number of sessions</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>EER (%)</td>
<td>8.7</td>
<td>8.1</td>
<td>7.3</td>
<td>6.9</td>
</tr>
</tbody>
</table>

3.5.9 CPU time and speeding up using Random Projection

In this section, CPU time of the proposed algorithm is investigated. Fig. 3.8 shows the CPU time for different number of samples and dimensions on the synthetic dataset. This demonstrates that the proposed algorithm can be applied to high dimensional data with thousands of samples. More specifically, complexity of the proposed feature selection algorithm is linear in number of features and quadratic in number of samples. These results show that the proposed feature selection algorithm can be applied to large datasets where both number of samples and features are large.
Chapter 3. Feature Selection for Non-stationary Data

Figure 3.8: The CPU time (in seconds) taken by the proposed algorithm to perform feature selection on the synthetic dataset where the parameters $\sigma$ and $\lambda$ are set to 1. (a) CPU time vs number of features. The figure shows linear complexity of the proposed method with respect to feature dimensionality. Number of samples is set to 300. (b) CPU time vs number of samples. The figure shows the proposed algorithm has quadratic time with respect to number of samples. Dimensionality is set to 5000.

Regarding the real-world datasets, CPU time of the proposed feature selection algorithms is presented in Table 3.6. EER for both MSFS and MSFS-RP algorithms are also provided for comparison. As may be seen, MSFS-RP speeds up on average about 4 times compare to MSFS algorithm while its effect on EER can be neglected. Here, dimensionality is reduced from $\sim7500$ to 500 (i.e., $K = 500$) by random projection. We realized that on average only 21% of samples were selected by the criteria described in section 3.4 to estimate the nearest miss and nearest hit for each sample. This leads to less complexity in computing $d_s(i)$ and absolute difference matrices in (3.7) and (3.8) (see section 3.4).

Table 3.6: CPU time (in seconds) of the MSFS and MSFS-RP algorithms. EER (in percent) is also presented in parentheses.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>MSFS-RP</th>
<th>MSFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG (Across-session)</td>
<td>$M = 3675$, $J = 7278$</td>
<td>790(7.6)</td>
</tr>
<tr>
<td>TEOAE (Across-session)</td>
<td>$M = 600$, $J = 7522$</td>
<td>65(3.4)</td>
</tr>
<tr>
<td>ECG (Across-Posture)</td>
<td>$M = 1472$, $J = 7278$</td>
<td>151(3.9)</td>
</tr>
</tbody>
</table>

We also report the CPU time for our comparison feature selection algorithms in Table 3.7.
Table 3.7: CPU time (in seconds) for our comparison feature selection algorithms. The symbols (*) and (†) denote algorithms are respectively implemented in MATLAB and C. The total time is reported in the last row. For convenience, CPU time of the proposed MSFS-RP method is repeated in the last column.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>LLFS*</th>
<th>Simba*</th>
<th>DEFS*</th>
<th>JMI†</th>
<th>ICAP†</th>
<th>MRF†</th>
<th>CIFE†</th>
<th>MSFS-RP*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG (Across-session)</td>
<td>4108</td>
<td>1556</td>
<td>1256</td>
<td>400</td>
<td>769</td>
<td>110</td>
<td>754</td>
<td>790</td>
</tr>
<tr>
<td>TEOAE (Across-session)</td>
<td>126</td>
<td>47</td>
<td>304</td>
<td>69</td>
<td>147</td>
<td>3</td>
<td>139</td>
<td>65</td>
</tr>
<tr>
<td>ECG (Across-Posture)</td>
<td>682</td>
<td>260</td>
<td>694</td>
<td>188</td>
<td>350</td>
<td>18</td>
<td>331</td>
<td>151</td>
</tr>
<tr>
<td>Total time</td>
<td>4916</td>
<td>1863</td>
<td>2254</td>
<td>657</td>
<td>1266</td>
<td>131</td>
<td>1224</td>
<td>1006</td>
</tr>
</tbody>
</table>

The proposed method is implemented in MATLAB that is known to be much slower than many other programming languages especially C. Codes for LLFS, Simba and DEFS are written in MATLAB and JMI, ICAP, MRF and CIFE are written in C. It can be seen that the proposed MSFS-RP algorithm on average is faster than all MATLAB-implemented methods LLFS, Simba and DEFS as well as the two C implemented methods ICAP and CIFE. Although JMI algorithm is implemented in C, it still runs slower than MSFS-RP in 2 out of 3 experiments. The only method that is faster than the proposed method is MRF. This can mainly be because of its faster programming language (i.e. C) since the complexity of our method and MRF are the same – i.e. quadratic in number of samples and linear in number of features [123]. Note that this is the development phase time which is performed off-line. On the other hand, enrollment and test phases involve only training and testing using SVM. This is much faster than the process in development phase, since no feature selection is required. In our experiments, the test phase is performed in a fraction of a second.

Compare to MSFS algorithm, MSFS-RP reduces the run-time by incorporating only a portion of samples in computing \( \tilde{d}^{(i)}_{NM} \) and \( \tilde{d}^{(i)}_{NH_s} \) (see equation (3.5) and (3.6)). Figure 3.9 shows the error in computing \( \tilde{d}^{(i)}_{NM} \) on TEOAE dataset. Error is defined as \( \frac{\| \tilde{d}_{ORG} - \tilde{d}_{RP} \|_2}{\| \tilde{d}_{ORG} \|_2} \) where \( \tilde{d}_{ORG} \) corresponds to MSFS algorithm and \( \tilde{d}_{RP} \) corresponds to MSFS-RP algorithm. As may be seen, the error is very small – i.e., less that 5%. This can explain why the MSFS-RP method has similar performance compare to the MSFS method because MSFS-RP and MSFS mainly differs in the way that \( \tilde{d}^{(i)}_{NM} \) and \( \tilde{d}^{(i)}_{NH_s} \) are computed and in fact other steps of both algorithms are the same.

We also compare the MSFS and MSFS-RP algorithms in terms of the number of features
Figure 3.9: Approximation error in computing $\bar{d}_{NM}^{(i)}$ in (3.5) for TEOAE dataset. Error is defined as $\frac{||\bar{d}_{ORG}-\bar{d}_{RP}||_2}{||\bar{d}_{ORG}||_2}$ where $\bar{d}_{ORG}$ corresponds to MSFS method and $\bar{d}_{RP}$ corresponds to MSFS-RP method.

they have in common. Fig. 3.10 shows the percentage of common features between two methods for different values of $\gamma \in \{75\%, 85\%, 95\%, 99\%\}$ in TEOAE dataset. $\gamma$ controls the number of selected samples. Smaller values of $\gamma$ lead to selecting a smaller number of samples. Note that at around 300 selected features (where the proposed algorithm usually gives its best performance on this dataset), MSFS-RP with $\gamma = 0.95$ or 0.99 gives a feature subset that has almost all features in common with MSFS (about 98%). However, further decreasing $\gamma$ significantly decreases the percentage of common features. We set $\gamma$ to 0.95 (i.e., its default value) without tuning and it is fixed during all experiments.

While MSFS-RP algorithm benefits from less complexity due to dealing with smaller subsets of samples, determining these subsets adds some extra time to the algorithm compared to MSFS algorithm. This implies that the dimensionality of low-dimensional space, $K$, should

Figure 3.10: Percentage of common features shared between MSFS and MSFS-RP versus the number of selected features in TEOAE dataset for $\gamma \in \{75\%, 85\%, 95\%, 99\%\}$. 
Chapter 3. Feature Selection for Non-stationary Data

Figure 3.11: Average error in the distance between members of a pair of samples versus $K$. Error is averaged over 30 runs where in each run errors of 1000 pairs are computed. It can be seen that the error significantly increases when $K$ is smaller than approximately 500.

be small compared to the original dimensionality $L$. We randomly pick 1000 pairs of samples and compute the $l_1$ distance between members of each pair before and after being projected into a $K$-dimensional spaces. After normalizing by $\left(\frac{K}{L}\right)$, the root-mean-square error between the distances before and after projection is computed. We run this procedure 30 times and compute the average error over all runs. Fig. 3.11 shows the average error for different values of $K$. As may be seen, $K$ can be reduced while its effect on error is unnoticeable. But if we keep reducing $K$ beyond approximately 500, error will significantly increase. Therefore, we set $K$ to 500 (i.e., its default value) without tuning and it is fixed during all experiments.

3.6 Chapter Summary

We propose a feature selection method that selects features which are more persistent across sessions. A multisession auxiliary dataset of some general subjects is used for feature selection. Selected features are then used for building templates of biometric system users. The proposed approach overcomes the weakness of previously existing feature selection algorithms that ignore the fact that the input data belong to different sessions. Moreover, we embedded an across session criteria in the feature selection process that perfectly fits the aforementioned single session enrollment, across session testing scenario. The proposed feature selection algorithm is computationally efficient, i.e. linear in number of features and quadratic in number of samples and we have demonstrated that it can be applied to large datasets with thousands of features and samples. In addition, we have done experiments on TEOAE signals. The proposed
method has been compared with six state-of-the-art methods for ECG and TEOAE biometric recognition. In addition, it has been compared with seven existing feature selection algorithms in the literature. Experimental results demonstrate the effectiveness of the proposed method to cope with variation of such signals due to time lapse and body posture.
Chapter 4

Cross View Matching for Biometric Recognition

4.1 Introduction

In cross view matching problem, the system aims to measure the similarity between two samples that are in different views and decide whether the two samples match? In this context, view can refer to facial pose in cross pose face recognition, imaging spectrum in near infrared versus visible light face recognition or heart rate in ECG recognition where heart rate of gallery and probe are different. This problem may also arise in handwriting word spotting where one view is image of a word and the other view is string of that word or text-image retrieval where one view is an image and the other view is the text surrounding the image.

We focus on a challenging scenario where testing have to be done on some new classes that have not seen during training. This is more challenging than the conventional classification problems where testing is done on some samples that although are new but belong to known classes that have been seen during training. For example this may arise in a biometric system where training is done on an auxiliary dataset consists of some generic subjects but it has to be used to recognize subjects that have not been seen during training –i.e. using the system for new subjects without retraining the system. Similarly, in the context of word spotting or text-image retrieval, it might be desired to spot respectively a new name or image that has not
been seen during the training. In this chapter, we focus on cross view matching for biometric recognition.

During training, we form a rich pool of different features extracted from original signals/images. Each view can have its own feature set that can be different from that of other views in terms of type of features involved and length of the feature vector. Therefore, direct comparison doesn’t help. Instead, we represent a new subject by its (dis)similarities to people that we have seen before –i.e. subjects in an auxiliary dataset. The auxiliary dataset consists of some generic subjects for which samples of all views are available.

Ideally, such representation is independent of the view in which the (dis)similarities are computed. We realize this via a feature weighting approach that emphasis on most relevant features. Feature weights are learned on the auxiliary dataset. More precisely, we find a set of weights for each view such that cross matching error is minimized and class separability within each view is maximized. The cross matching error is defined as the error between view-independent representation of samples belongs to the same subject but are in different views. Class separability is defined as a local margin around samples in each view.

Testing involves computing the distance between a pair of samples from different views. Once weights are learned on the auxiliary dataset during a training phase, the view-independent
representation of each test sample is computed in its corresponding weighted space. Dissimilarity of two test samples in different views is obtained via computing the distance between the view-independent representation of the samples. Figure 4.1 illustrates the proposed approach where two samples that have different length due different heart rates are compared via their (dis)similarities to some generic subjects in an auxiliary set in the weighted space. In addition to cross heart rate ECG recognition, we demonstrate the generalization of the proposed method on two other applications: near infrared versus visible light face recognition and, cross pose face recognition.

The remaining portion of this chapter is organized as follows: Section 4.2 briefly reviews recent works in the area of cross-view matching. Details of the proposed method is presented in section 4.3. Experimental results are presented in section 4.4 and section 4.5 concludes the chapter.

4.2 Related Work

The main steam of current approaches transform samples of different views into a common space where comparison is done in the common space. Generally, they can be divided in unsupervised and supervised approaches.

Unsupervised approaches such as Canonical Correlation Analysis (CCA) [132] and its kernel version, Kernel CCA (KCCA) [132] projects the two views onto common space such that the correlation between the projected views is maximized. [133] is a deep extensions of CCA based on deep neural networks and autoencoders. While CCA finds the directions of maximum correlation, Partial Least Squares (PLS) [134] which is based on Partial Least Squares regression finds projection matrices that maximum the pairwise covariance. [135] is a multiview subspace learning for clustering where one view is constrained by the spectral embedding from the other view.

On the other side, supervised methods such as Correlation Discriminant Analysis [136] and discriminant CCA [137,138] use class labels to learn a discriminant common space. Multiview Fisher Discriminant Analysis [139,140] learns classifiers in different views and a regularization term that penalizes the disagreement between views. However, it is confined to binary classifi-
cation problems. In Common Discriminant Feature Extraction (CDFE) [141] scatter matrices are defined on two different views, and local consistency of manifold is imposed to regularize the dimension reduction. Coupled Spectral Regression (CSR) [142] finds the most effective low-dimensional embeddings for the original sample data and learns the projection between the low-dimensional embedding and the original data through least squares regression. In [143] an iterative algorithm based on sparsity is used to learn two coupled linear regression models that projects data from different modalities into a common subspace. In [144], views are projected to a common subspace learned in an unsupervised way and then classification is done with softly assigned metrics which are locally optimized. In Generalized Multiview Analysis (GMA) [145] a supervised extension of CCA is presented that uses both pair-wised and label information of views to learn a common subspace. This method is extended in Multiveiw Discriminant Analysis (MvDA) [146] such that it also considers inter-view discriminant information while maximizing the between-class and minimizing the within-class variations. In [147] a discriminative shared Gaussian process latent variable model is used to discover correlations between different views. Multiview Uncorrelated Discriminant Analysis (MULDA) [148] extracts uncorrelated features in each view and computes transformations of each view such that the extracted features contain minimum redundancy. A modified version that method called MULDA-m also proposed by replacing CCA with DCCA in their method. They also proposed MLDA-m which can be seen as extension of the method in [145] via replacing CCA with DCCA.

While prior works are subspace learning methods that learn projection matrices for each view, we approach the cross-view matching problem from a feature weighting perspective which can be seen as an especial case of subspace learning methods where the projection matrices are forced to be diagonal. Therefore, it has a lower degree of freedom. This constraint may explain the better generalization of the proposed method on data from unseen classes which is an important characteristic of our experimental protocol.

4.3 Proposed method

Let $\mathcal{D} = \{x^{(i)}, y^{(i)}, z^{(i)}\}_{i=1}^{M} \subset \mathbb{R}^{J_x} \times \mathbb{R}^{J_y} \times Z$ be an auxiliary dataset consisting of $M$ pairs of samples in $\mathcal{X}$ and $\mathcal{Y}$ views. $x^{(i)}$ and $y^{(i)}$ are respectively $J_x$ and $J_y$ dimensional feature vectors
with class label \( z^{(i)} \in \mathcal{Z} \), where \( \mathcal{Z} = \{Z_1, \ldots, Z_N\} \) is the set of all class labels. Considering the i-th pair i.e. \( \mathbf{x}^{(i)} \) and \( \mathbf{y}^{(i)} \) with class label \( z^{(i)} = Z_j \), we represent them by their distances to the centres of other classes in the corresponding view as follows:

\[
\phi(\mathbf{x}^{(i)}) = \left( \begin{array}{c} d^{(i,1)}_x, \ldots, d^{(i,j-1)}_x, d^{(i,j+1)}_x, \ldots, d^{(i,Z_N)}_x \end{array} \right) \quad (4.1)
\]

\[
\phi(\mathbf{y}^{(i)}) = \left( \begin{array}{c} d^{(i,1)}_y, \ldots, d^{(i,j-1)}_y, d^{(i,j+1)}_y, \ldots, d^{(i,Z_N)}_y \end{array} \right) \quad (4.2)
\]

where \( \phi(\mathbf{x}^{(i)}) \) and \( \phi(\mathbf{y}^{(i)}) \) are \( N - 1 \) dimensional vectors. Considering \( l_1 \) distance, \( d^{(i,z)}_x \) and \( d^{(i,z)}_y \), \( z \in \{1, \ldots, N\} \), \( z \neq j \) can be computed as follows:

\[
d^{(i,z)}_x = 1^T d^{(i,z)}_x \\
d^{(i,z)}_y = 1^T d^{(i,z)}_y \quad (4.3)
\]

where

\[
d^{(i,z)}_x = |\mathbf{x}^{(i)} - \mathbf{r}^{(z)}_x| \\
d^{(i,z)}_y = |\mathbf{y}^{(i)} - \mathbf{r}^{(z)}_y| \quad (4.4)
\]

and \( \mathbf{r}^{(z)}_x \) and \( \mathbf{r}^{(z)}_y \) are the centre of \( z \)-th class in \( \mathcal{X} \) and \( \mathcal{Y} \) views respectively. \((\cdot)^T\) is transpose operator and \(|\cdot|\) is element-wise absolute operator. While other options are possible (e.g. Euclidean distance by substituting \(|\cdot|\) with \((\cdot)^2\) in (4.4)), throughout this study we use \( l_1 \) distance.

Since \( \mathbf{x}^{(i)} \) and \( \mathbf{y}^{(i)} \) belong to the same class, we want their representations \( \phi(\mathbf{x}^{(i)}) \) and \( \phi(\mathbf{y}^{(i)}) \) be similar. We realize this goal by “weighting” the feature space such that the cross view matching error \( \|\phi(\mathbf{x}^{(i)}) - \phi(\mathbf{y}^{(i)})\|_2 \) be minimized. Equation (4.3) defines the distances in the original space. Similarly, we can compute distances in the weighted space as follows:

\[
d^{(i,z)}_x = \mathbf{f}^T_x d^{(i,z)}_x \\
d^{(i,z)}_y = \mathbf{f}^T_y d^{(i,z)}_y \quad (4.5)
\]

where \( \mathbf{f}_x \) and \( \mathbf{f}_y \) are nonnegative weight vectors. Hence, the problem of minimizing the cross-view error in the weighted spaces can be stated as follows:

\[
\min_{\mathbf{f}_x, \mathbf{f}_y} \sum_{i=1}^M \|\mathbf{f}^T_x \mathbf{D}_x^{(i)} - \mathbf{f}^T_x \mathbf{D}_x^{(i)}\|_2^2 \quad \text{s.t.} \quad \mathbf{f}_x, \mathbf{f}_y \geq 0 \quad (4.6)
\]

\( \mathbf{D}_x^{(i)} \) and \( \mathbf{D}_y^{(i)} \) are difference matrices where their columns are the difference vectors with respect
to the i-th sample and are computed as follows:

\[
D_x^{(i)} = \left( d_x^{(i,1)}, \ldots, d_x^{(i,j-1)}, d_x^{(i,j+1)}, \ldots, d_x^{(i,Z_N)} \right) \quad (4.7)
\]

\[
D_y^{(i)} = \left( d_y^{(i,1)}, \ldots, d_y^{(i,j-1)}, d_y^{(i,j+1)}, \ldots, d_y^{(i,Z_N)} \right) \quad (4.8)
\]

Note that \(D_x^{(i)}\) and \(D_y^{(i)}\) have \(N - 1\) columns, i.e. class \(z^{(i)} = Z_j\) is excluded. In addition, we want to have classes be separated. To this end, we maximize the margin around each sample. For the sake of brevity, in the following, only equations for \(X\) view are described. Driving the equations for \(Y\) view is straightforward. The margin for the i-th sample in \(X\) is defined as follows:

\[
\ell_x^{(i)} = f_x^T d_x^{(i)} \quad (4.9)
\]

where \(d_x^{(i)} = d_{NM_x}^{(i)} - d_{NH_x}^{(i)}\) and \(d_{NM_x}^{(i)}\) and \(d_{NH_x}^{(i)}\) are absolute difference vectors determined as follows:

\[
d_{NM_x}^{(i)} = \left| x^{(i)} - NM(x^{(i)}) \right|, \quad d_{NH_x}^{(i)} = \left| x^{(i)} - NH(x^{(i)}) \right| \quad (4.10)
\]

where \(NM(x^{(i)})\) is the nearest neighbor of \(x^{(i)}\) with a different class label (nearest miss) and \(NH(x^{(i)})\) is the nearest neighbor of \(x^{(i)}\) with the same class label as \(x^{(i)}\) (nearest hit). Intuitively, a positive margin allows the sample to wander in the sample space and still be correctly classified by a nearest neighbor classifier in a leave-one-out fashion, i.e. a better generalization on unseen data. As a part of a multi-objective optimization problem, we aim to maximize the margin in the “weighted” space of each view. This can also be seen as minimizing the leave-one-out classification error on each view.

At the problem outset, \(f_x\) and \(f_y\) are unknown. Therefore, determining nearest miss and nearest hit in the weighted space defined by \(f_x\) and \(f_y\) is a challenging issue. To overcome this issue, we use an iterative approach for computing \(f_x\) and \(f_y\), where at each iteration \(f_x\) and \(f_y\) are determined based on the distances in the weighted space defined at the previous iteration. However, determining \(NM_x(x^{(i)})\) and \(NH_x(x^{(i)})\) in the original space, may not be accurate. To address this issue, margin is estimated as the expectation of \(\ell_x^{(i)}(f)\) over all possible candidates.
for $\text{NM}_x(\mathbf{x}^{(i)})$ and $\text{NH}_x(\mathbf{x}^{(i)})$ as follows:

$$
\ell_x^{(i)}(f_x) = f_x^T \bar{\mathbf{d}}_x^{(i)} \quad (4.11)
$$

where

$$
\bar{d}_x^{(i)} = \bar{d}_x^{(i)}_{\text{NM}_x} - \bar{d}_x^{(i)}_{\text{NH}_x} \quad (4.12)
$$

and

$$
\bar{d}_x^{(i)} = \mathbf{D}^{(i)}_{\text{NM}_x} \mathbf{p}^{(i)\text{T}}_{\text{NM}_x} \quad \bar{d}_x^{(i)} = \mathbf{D}^{(i)}_{\text{NH}_x} \mathbf{p}^{(i)\text{T}}_{\text{NH}_x} \quad (4.13)
$$

$\mathbf{D}^{(i)}_{\text{NM}_x}$ and $\mathbf{D}^{(i)}_{\text{NH}_x}$ are matrices whose columns are absolute difference vectors with respect to $\mathbf{x}^{(i)}$:

$$
\mathbf{D}^{(i)}_{\text{NM}_x} = \left( \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{M}^i(1))} \right|, \ldots, \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{M}^i(n))} \right| \right) \quad (4.14)
$$

$$
\mathbf{D}^{(i)}_{\text{NH}_x} = \left( \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{H}^i(1))} \right|, \ldots, \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{H}^i(m))} \right| \right) \quad (4.15)
$$

$\mathcal{M}^i$ and $\mathcal{H}^i$ with cardinality of $n$ and $m$ denote set of all possible candidates for $\text{NM}(\mathbf{x}^{(i)})$ and $\text{NH}(\mathbf{x}^{(i)})$ respectively and are defined as:

$$
\mathcal{M}^i = \left\{ j \in \{1, \ldots, M\} \mid z^{(j)} \neq z^{(i)} \right\} \quad (4.16)
$$

$$
\mathcal{H}^i = \left\{ j \in \{1, \ldots, M\} \mid z^{(j)} = z^{(i)}, j \neq i \right\} \quad (4.17)
$$

$\mathbf{p}^{(i)}_{\text{NM}_x} \ (\mathbf{p}^{(i)}_{\text{NH}_x})$ in eq. (4.13) is an n-dimensional (m-dimensional) row vector indicates the probability of samples in $\mathcal{M}^i \ (\mathcal{H}^i)$ being $\text{NM}_x(\mathbf{x}^{(i)}) \ (\text{NH}_x(\mathbf{x}^{(i)}))$. Within the weighted space, samples situated closer to $\mathbf{x}^{(i)}$ are more probable to be the nearest sample. So the probabilities are determined as follows:

$$
\mathbf{p}^{(i)}_{\text{NM}_x} = \exp \left( -\mathbf{f}_x^T \mathbf{D}^{(i)}_{\text{NM}_x} / \sigma \right) \quad (4.18)
$$

$$
\mathbf{p}^{(i)}_{\text{NH}_x} = \exp \left( -\mathbf{f}_x^T \mathbf{D}^{(i)}_{\text{NH}_x} / \sigma \right) \quad (4.19)
$$

where $\sigma$ is a user settable parameter. $\mathbf{p}^{(i)}_{\text{NM}_x}$ and $\mathbf{p}^{(i)}_{\text{NH}_x}$ are then normalized to sum to one to be the probabilities utilized in (4.13).
Having $2 \times M$ margins of the form $\ell_x^{(i)}(f_x) = f_x^T \bar{d}_x^{(i)}$ and $\ell_y^{(i)}(f_y) = f_y^T \bar{d}_y^{(i)}$, it is desired to maximize all margins. Considering a logistic regression formulation, the optimization problem can be expressed as follows:

$$\max_{f_x,f_y} \sum_{i=1}^{M} G(f_x^T \bar{d}_x^{(i)}) + \sum_{i=1}^{M} G(f_y^T \bar{d}_y^{(i)}), \quad \text{s.t.} \quad f_x, f_y \geq 0,$$

(4.20)

where $G(\cdot)$ is a logistic function.

$$G(b) = \log \left( \frac{1}{1 + \exp(-b)} \right)$$

(4.21)

$G$ is a strictly increasing function, therefore maximizing $G(f_x^T \bar{d}_x^{(i)})$ indeed implies maximizing $f_x^T \bar{d}_x^{(i)}$. However, $G$ is useful because it can take an input that can vary from negative to positive infinity whereas the output always ranges between 0 and 1. Then (4.20) can be simplified as follows:

$$\min_{f_x,f_y} \sum_{i=1}^{M} \log \left( 1 + \exp \left( -f_x^T \bar{d}_x^{(i)} \right) \right) + \log \left( 1 + \exp \left( -f_y^T \bar{d}_y^{(i)} \right) \right), \quad \text{s.t.} \quad f_x, f_y \geq 0,$$

(4.22)

We define $f = (f_x^T, f_y^T)^T$ and rewrite (4.22) in terms of $f$ as follows:

$$\min_{f} \sum_{i=1}^{2M} \log \left( 1 + \exp \left( -f^T r^{(i)} \right) \right), \quad \text{s.t.} \quad f \geq 0$$

(4.23)

where $r^{(i)}, i = 1, \ldots, 2M$ are columns of the following matrix:

$$R = \begin{pmatrix} \bar{d}_x^{(1)} & \cdots & \bar{d}_x^{(M)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \bar{d}_y^{(1)} & \cdots & \bar{d}_y^{(M)} \end{pmatrix}$$

(4.24)

and $0$ is a zero matrix of appropriate size. We also rewrite (4.6) in terms of $f$. As shown in Appendix A, the cross view matching error in (4.6) can be expressed as a quadratic form in terms of $f$ as follows:

$$\sum_{i=1}^{M} \| f_x^T D_x^{(i)} - f_y^T D_y^{(i)} \|_2^2 = \sum_{i=1}^{M} f^T D^{(i)} f = f^T D f$$

(4.25)
\[
\text{Input: } D = \{x^{(i)}, y^{(i)}, z^{(i)}\}_{i=1}^{M}, \sigma, \lambda \\
\text{Output: } \{f\} \\
\text{1 Initialization: Set } f = (1, \ldots, 1)^{T}; \\
\text{2 repeat} \\
\text{3 } f_{\text{prev.}} = f; \\
\text{4 for } i \leftarrow 1 \text{ to } M \text{ do} \\
\text{5 Compute } D^{(i)}_x \text{ and } D^{(i)}_y \text{ as in (4.7) and (4.8); Compute } D^{(i)} \text{ as in (4.26);} \\
\text{6 Compute } D^{(i)}_{NMx} \text{ and } D^{(i)}_{NMy} \text{ as in (4.14);} \\
\text{7 Compute } \tilde{d}^{(i)}_{NMx} \text{ and } \tilde{d}^{(i)}_{NMy} \text{ using } f_{\text{prev.}} \text{ as in (4.18);} \\
\text{8 Compute } \tilde{d}^{(i)}_{NHx} \text{ and } \tilde{d}^{(i)}_{NHy} \text{ as in (4.12);} \\
\text{9 Compute } \tilde{d}^{(i)}_{NHx} \text{ and } \tilde{d}^{(i)}_{NHy} \text{ as in (4.13);} \\
\text{10 Compute } \tilde{d}^{(i)}_x \text{ and } \tilde{d}^{(i)}_y \text{ as in (4.12);} \\
\text{11 Compute } \tilde{d}^{(i)}_x \text{ and } \tilde{d}^{(i)}_y \text{ as in (4.12);} \\
\text{12 Compute } \tilde{d}^{(i)}_x \text{ and } \tilde{d}^{(i)}_y \text{ as in (4.12);} \\
\text{13 Compute } \tilde{d}^{(i)}_x \text{ and } \tilde{d}^{(i)}_y \text{ as in (4.12);} \\
\text{14 end} \\
\text{15 Compute } R \text{ as in (4.24);} \\
\text{16 Compute } D = \sum_{i=1}^{M} D^{(i)}; \\
\text{17 Compute } a \text{ through solving (3.17);} \\
\text{18 Set } f_j = a^2_j, \text{ } 1 \leq j \leq J \text{ until } \|f - f_{\text{prev.}}\|_2 < \varepsilon; \\
\]

**Algorithm 3:** Pseudo code of the proposed algorithm.

\[
D^{(i)} = \begin{pmatrix} D^{(i)}_x D^{(i)}_x^T & -D^{(i)}_x D^{(i)}_y^T \\ -D^{(i)}_y D^{(i)}_x^T & D^{(i)}_y D^{(i)}_y^T \end{pmatrix}
\tag{4.26}
\]

and \( D = \sum_{i=1}^{M} D^{(i)} \). Considering both objectives, i.e. minimizing the cross-view matching error and maximizing the sample margins, the problem can be formulated as follows:

\[
\min_f \frac{1}{2M} \sum_{i=1}^{2M} \log \left( 1 + \exp \left( -f^T r^{(i)} \right) \right) + \lambda f^T D f, \quad \text{s.t. } f \geq 0
\tag{4.27}
\]

where \( \lambda \) is a user settable parameter. Equation (4.27) is a constrained convex optimization problem with respect to \( f \). We then replace the vector \( f \) with a new vector \( a \) such that \( f_j = a^2_j, \text{ } 1 \leq j \leq J \) where \( J = J_x + J_y \) and reformulate the problem in an unconstrained form as follows:

\[
\min_a \frac{1}{2M} \sum_{i=1}^{2M} \log \left( 1 + \exp \left( -\sum_{j=1}^{J} a^2_j r^{(i)}(j) \right) \right) + \lambda \sum_{i,j=1}^{J} a^2_i d_{ij} a^2_j
\tag{4.28}
\]
$r^{(i)}(j)$ is the $j$-th element of vector $r^{(i)}$ and $d_{ij}$ denotes elements of matrix $D$. (4.28) can be solved via gradient decent with step size of $\tau$ and updated as follows:

$$a \leftarrow a - \tau \Delta$$

$$\Delta = a \otimes \left( \lambda \sum_j a_j^2 d_{ij} - \sum_{i=1}^{2M} \frac{\exp \left( - \sum_j a_j^2 r^{(i)}(j) \right)}{1 + \exp \left( - \sum_j a_j^2 r^{(i)}(j) \right)} r^{(i)} \right)$$

(4.29)

where $\otimes$ is Hadamard operator. We utilized fminbnd function in MATLAB to determine $\tau$ using the line search algorithm. $f$ is initialized to 1, so that all weights are the same at the beginning and then will be updated using (4.29) until a stopping criteria is satisfied. In this study, the algorithm stops when the difference between the weights in two successive iterations is less than a threshold $\varepsilon$ which is set to 0.01. Pseudo code of the proposed method is presented in Algorithm 3.

4.4 Experimental results

We evaluate performance of the proposed method on three different applications. ECG recognition in different heart rates, near infrared versus visible light face recognition and cross pose face recognition. Performance of the proposed method is compared against state-of-the-art method including MULDA, MLDA-m, MULDA-m presented in [148], GMA [145] and MvDA [146].

The proposed method have two parameters: $\sigma$ and $\lambda$. $\sigma$ controls the local behavior of the proposed algorithm. A large value of $\sigma$ increase the effect of far samples in determining the nearest sample. If $\sigma$ goes to infinity, all samples will have equal probability to be the the nearest. Therefore the nearest sample would be the average of all other samples i.e no localization. On the other side, if $\sigma$ goes to zero, we experimentally realized that the algorithm will not converge. The second parameter $\lambda$ control the trade off between maximum margin criterion and cross matching error. A large value of $\lambda$ emphasizes on reducing cross matching error while a small value of $\lambda$ emphasizes on maximizing margins i.e. better discrimination. Generally these parameters can be determined using cross validation. However, in our experiments $\sigma$ and $\lambda$ are set to 3 and $10^{-3}$ respectively –i.e. default values and they are fixed during all experiments. For
all comparison methods input dimension in each view is reduced using Principal Component Analysis (PCA) such that 95% of energy is preserved. The output dimension of the comparison methods is also tuned with step size of 50 and the best results is reported. In the GMA method, as suggested in [145] and [146], $\mu$ and $\gamma$ are set to 1 and the trace ratio respectively, and $\lambda$ is tuned in [1 100]. For MULDA, MULDA-m and MLDA-m methods, as suggested in [148], $\gamma$ is tuned among [1,5,10,15,20]. The codes for all comparison methods have been provided by the respective authors.

4.4.1 Experiments of ECG recognition

In this experiment, we use the ECG database collected in BioSec lab in University of Toronto [80]. There are 82 subjects that have recordings in up to 5 sessions in rest and after exercise conditions. We take the first 40 subjects as auxiliary dataset and the rest goes for enrollment and testing. ECG signals were recorded using Vernier EKG sensor and Go!Link interface [81] with 12 bits resolution and sampling rate of 200 Hz using three dry AgCl electrodes from fingertips. We consider the heart rate between 66 to 100 beats per minutes which is equivalent to RR interval of 120 to 180 samples and this interval is uniformly divided into 15 non-overlapping bins. We pick the middle bin i.e. bin 8 for enrollment and the remaining 14 bins goes for testing. Therefore we compute 14 pairs of feature weights on the auxiliary set. There is no overlap between the auxiliary set and the test set. In the auxiliary dataset, we randomly selects 15 samples to represent each bin. In the test dataset, we randomly pick 20 samples for enrollment and 5 samples for testing.

Each bin includes a range for RR intervals. However, it is desired that length of samples within a bin be the same. Therefore, we pick the center of a interval to represent the length of that bin. We consider a window centred at R peak where the length of the window is twice the the length of its corresponding bin and compute Continues Wavelet Transform (CWT) with Daubechies 5 as mother wavelet. We also consider six frequency bands: 8-13Hz, 13-18Hz, 18-25Hz, 25-30Hz, 30-35Hz, 35-50Hz and for each band, mean of power, standard deviation of power, maximum amplitude, standard deviation of amplitude, kurtosis and skewness were computed. Maximum, standard deviation, kurtosis and skewness are also computed from the signal itself. The signal amplitude, i.e. a window centred around R peak, is also considered.
Table 4.1: ECG recognition across different heart rates. EER (in percent) and standard deviation (in percent) are reported for different methods. Standard deviations are presented in parentheses.

<table>
<thead>
<tr>
<th>Heart Rate</th>
<th>MULDA</th>
<th>MLDA-m</th>
<th>MULDA-m</th>
<th>GMA</th>
<th>MvDA</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>67.4</td>
<td>13.0(1.3)</td>
<td>13.0(1.5)</td>
<td>13.3(1.1)</td>
<td>14.3(0.0)</td>
<td>16.6(2.4)</td>
<td>12.7(1.1)</td>
</tr>
<tr>
<td>69.0</td>
<td>16.0(0.9)</td>
<td>16.6(0.1)</td>
<td>16.5(0.4)</td>
<td>16.7(0.1)</td>
<td>22.9(1.5)</td>
<td>12.5(1.6)</td>
</tr>
<tr>
<td>70.6</td>
<td>14.5(0.8)</td>
<td>13.2(1.4)</td>
<td>13.3(1.4)</td>
<td>12.1(1.9)</td>
<td>15.8(0.8)</td>
<td>10.7(1.1)</td>
</tr>
<tr>
<td>72.3</td>
<td>13.0(1.7)</td>
<td>13.8(1.9)</td>
<td>13.9(1.7)</td>
<td>11.6(1.4)</td>
<td>13.6(2.2)</td>
<td>10.0(1.0)</td>
</tr>
<tr>
<td>74.1</td>
<td>8.2(0.7)</td>
<td>8.8(1.5)</td>
<td>8.8(1.5)</td>
<td>8.6(0.6)</td>
<td>11.9(2.4)</td>
<td>10.2(1.6)</td>
</tr>
<tr>
<td>76.0</td>
<td>11.9(1.8)</td>
<td>12.8(1.9)</td>
<td>13.0(1.8)</td>
<td>11.0(0.5)</td>
<td>12.7(2.0)</td>
<td>11.2(1.4)</td>
</tr>
<tr>
<td>77.9</td>
<td>12.8(1.9)</td>
<td>13.4(1.4)</td>
<td>13.5(1.4)</td>
<td>13.6(1.3)</td>
<td>15.6(1.2)</td>
<td>8.5(1.4)</td>
</tr>
<tr>
<td>82.2</td>
<td>16.5(0.4)</td>
<td>11.8(1.1)</td>
<td>11.6(0.9)</td>
<td>13.0(0.7)</td>
<td>15.3(0.9)</td>
<td>8.1(1.4)</td>
</tr>
<tr>
<td>84.5</td>
<td>13.9(2.2)</td>
<td>12.4(2.5)</td>
<td>12.4(2.5)</td>
<td>9.9(0.6)</td>
<td>12.1(0.4)</td>
<td>11.6(1.5)</td>
</tr>
<tr>
<td>87.0</td>
<td>16.1(1.7)</td>
<td>17.8(3.8)</td>
<td>17.2(3.8)</td>
<td>11.7(1.1)</td>
<td>14.2(1.8)</td>
<td>10.2(1.7)</td>
</tr>
<tr>
<td>89.6</td>
<td>8.8(0.6)</td>
<td>10.3(1.1)</td>
<td>10.5(1.1)</td>
<td>11.4(1.0)</td>
<td>12.6(1.1)</td>
<td>11.1(0.6)</td>
</tr>
<tr>
<td>92.3</td>
<td>17.3(1.9)</td>
<td>15.8(2.6)</td>
<td>15.0(2.6)</td>
<td>12.1(0.5)</td>
<td>18.5(5.6)</td>
<td>12.5(1.8)</td>
</tr>
<tr>
<td>95.3</td>
<td>25.1(2.4)</td>
<td>21.3(3.0)</td>
<td>21.4(3.0)</td>
<td>20.3(2.2)</td>
<td>20.2(1.7)</td>
<td>11.5(2.3)</td>
</tr>
<tr>
<td>98.4</td>
<td>24.6(1.7)</td>
<td>24.4(2.6)</td>
<td>25.0(2.6)</td>
<td>16.2(0.7)</td>
<td>12.9(1.0)</td>
<td>10.6(1.7)</td>
</tr>
</tbody>
</table>

| Average    | 15.1  | 14.7  | 14.7  | 13.0  | 15.4  | **10.8** |

Autocorrelation is computed and the number of lags is 80% of the corresponding bin length. Aforementioned feature vectors were concatenated to form the final feature vector. The length of the final feature vector of the first bin i.e. the slowest heart rate is 6252 and for the last bin i.e. the fastest heart rate is 4304. Features are normalized such that they have zero mean and unit variance. To account for the randomness in selecting samples in each bin, experiments were repeated 5 times and the average EER is reported in Table 4.1. It can be seen that the proposed method significantly performs better than the comparison methods.

### 4.4.2 Experiments of near infrared versus visible light face recognition

In this experiment, we use the NIR-VIS database in [149] consists of 202 subjects. We randomly select 4 samples per subject. The auxiliary dataset consists of the first 100 subjects and the rest goes for enrollment and testing. Images are aligned by an affine transform by eyes and center of mouth such that the distance between eyes is 100 pixels. Illumination is normalized by the method suggested in [150]. SURF features [151] are computed on 64 landmarks detected by the method in [152]. SURF features are also computed from 36 points uniformly distributed in a square area on inner face area. Also holistic representation is obtained by cropping and
Table 4.2: Near infrared versus visible light face recognition. Recognition rate (in percent) for different methods are compared.

<table>
<thead>
<tr>
<th>MULDA</th>
<th>MLDA-m</th>
<th>MULDA-m</th>
<th>GMA</th>
<th>MvDA</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>79.26</td>
<td>69.33</td>
<td>69.33</td>
<td>61.14</td>
<td>66.05</td>
<td>84.6</td>
</tr>
</tbody>
</table>

resizing the images to 41x36 pixels. These features are concatenated to from a final feature vector of the length 8132. Features are normalized such that they have zero mean and unit variance. Recognition rate of the proposed method and 5 comparison methods are reported in Table 4.2. It can be seen that the proposed method significantly performs better than the comparison methods. Note that since recognition rate is the most widely used performance measure in face recognition, we use it in our face experiments.

4.4.3 Experiments of cross pose face recognition

We use MultPie face database [153] consists of 337 subjects for which face images in 15 different poses, 20 illuminations, 6 expressions and 4 sessions are provided. We use the first 100 subjects for auxiliary dataset and the rest goes for enrollment and testing. We consider images from 4 different sessions and 5 illuminations i.e. 1,4,7,12,17. Enrollment is done on frontal lighting i.e. illumination 7 and testing is done on the aforementioned 5 lightings. Similar to NIR-VIS experiment, face images are aligned according to eyes and center of mouth by an affine transform and illumination is normalized by the method suggested in [150]. SURF features are computed on 100 points uniformly distributed on face area. Holistic features are obtained by cropping and resizing the images to 45x56 pixels. These features are concatenated to from a final feature vector of the length 8920. Features are normalized such that they have zero mean and unit variance. Recognition rate of the proposed method and 5 comparison methods are reported in Table 4.3. It can be seen that the proposed method significantly performs better than the comparison methods.

4.5 Chapter Summary

In this chapter, we proposed a novel method for cross view matching. Each view has its own feature set. Different types of features have been concatenated to form a rich pool of features for
Table 4.3: Cross pose face recognition. Recognition rate (in percent) for different methods are compared. *ah* refers to the two additional cameras (08-1 and 19-1) located above the subject, simulating a typical surveillance camera view.

<table>
<thead>
<tr>
<th>Angle</th>
<th>MULDA</th>
<th>MLDA-m</th>
<th>MULDA-m</th>
<th>GMA</th>
<th>MvDA</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>-90</td>
<td>10.1</td>
<td>9.4</td>
<td>9.5</td>
<td>15.2</td>
<td>22.2</td>
<td>26.5</td>
</tr>
<tr>
<td>-75</td>
<td>13.1</td>
<td>12.5</td>
<td>12.6</td>
<td>20.3</td>
<td>34.3</td>
<td>37.1</td>
</tr>
<tr>
<td>-60</td>
<td>16.0</td>
<td>15.3</td>
<td>15.3</td>
<td>28.3</td>
<td>45.7</td>
<td>58.6</td>
</tr>
<tr>
<td>-45</td>
<td>32.2</td>
<td>29.0</td>
<td>29.0</td>
<td>38.2</td>
<td>62.0</td>
<td>78.6</td>
</tr>
<tr>
<td>-30</td>
<td>53.1</td>
<td>47.5</td>
<td>47.5</td>
<td>55.6</td>
<td>80.2</td>
<td>81.8</td>
</tr>
<tr>
<td>-15</td>
<td>59.1</td>
<td>55.3</td>
<td>55.2</td>
<td>66.2</td>
<td>87.6</td>
<td>93.6</td>
</tr>
<tr>
<td>15</td>
<td>64.5</td>
<td>59.9</td>
<td>60.0</td>
<td>70.4</td>
<td>85.0</td>
<td>93.4</td>
</tr>
<tr>
<td>30</td>
<td>38.0</td>
<td>37.0</td>
<td>36.9</td>
<td>47.1</td>
<td>74.2</td>
<td>85.3</td>
</tr>
<tr>
<td>45</td>
<td>27.0</td>
<td>24.6</td>
<td>24.5</td>
<td>35.3</td>
<td>60.1</td>
<td>76.1</td>
</tr>
<tr>
<td>60</td>
<td>19.5</td>
<td>18.3</td>
<td>18.2</td>
<td>25.8</td>
<td>44.7</td>
<td>62.7</td>
</tr>
<tr>
<td>75</td>
<td>14.1</td>
<td>12.3</td>
<td>12.4</td>
<td>17.7</td>
<td>32.2</td>
<td>42.0</td>
</tr>
<tr>
<td>90</td>
<td>10.0</td>
<td>9.3</td>
<td>9.0</td>
<td>16.2</td>
<td>24.5</td>
<td>33.8</td>
</tr>
<tr>
<td><em>ah</em></td>
<td>25.6</td>
<td>22.7</td>
<td>22.8</td>
<td>31.8</td>
<td>47.4</td>
<td>64.6</td>
</tr>
</tbody>
</table>

| Average | 28.9 | 26.7 | 26.6 | 35.7 | 53.4 | **63.8** |

Each view. The proposed method finds a set of feature weights for each view such that the cross-view matching error is minimized and the local margin around each sample is maximized. The proposed method is a matcher that computes the similarity of two samples situated in different views. Once the feature weights are computed over an auxiliary set of some generic subjects, it can be readily applied to new subjects without requiring a training phase. Therefore, it is not confined to a predefined set of enrolled subjects.
Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis presented new ideas on human recognition based on ECG. Compare to conventional biometrics ECG is easy to collect i.e. from fingertips and more importantly, it is robust to spoof attacks due to its inherent liveness requirement and difficulty of stealing or fabrication.

Considering that both ECG and fingerprint can be recorded from fingertips, in chapter 2, we proposed an interesting application of ECG for fingerprint spoof detection. On one side, fingerprint is vulnerable to spoof attacks. On the other side, ECG comes with its inherent liveness detection. We proposed a unified approach for fusion of ECG and fingerprint that fills the gap between these two sides. The fusion process has two folds: fusion of ECG and fingerprint for spoof detection and fusion of ECG and fingerprint for identity verification. Unlike the conventional fusion of ECG and fingerprint that aims to improve the identity verification rate, the proposed method aims to improve spoof detection rate. In addition, the proposed system automatically performs template updating, so that the performance of biometric system can be maintained in long term without manual re-training or re-enrollment. We also proposed a simple yet effective criterion to automatically decide when to stop recording ECG. We showed that not all subjects need to have the same length of ECG recording. The recording time can be shorter for Subjects that their ECG is highly repetitive. Comprehensive experiments on LiVDet2015 database with 6 different spoofs, 3 different scanners and 3 different fusion methods were done. The proposed system achieved a liveness detection EER of 4.4% incorporating only 5 seconds
of ECG. By extending the recording time to 30 seconds, liveness detection EER is reduced to 3% which is about 4 times better than the best of six comparison methods. This is also about 2 times better than the best results achieved by participants of LivDet2015 competition.

Beside its advantages, ECG comes also with its challenges. Non-stationary nature of ECG makes it hard to deal with especially when only one session is available for enrollment and testing is done in a another session that could be a few weeks apart. In chapter 3, we proposed a feature selection method that selects features which are more persistent across sessions. A multisession auxiliary dataset of some generic subjects were used for feature selection. Selected features are then used for building templates of biometric system users. The proposed approach overcomes the weakness of previously existing feature selection algorithms that ignore the fact that the input data belong to different sessions. Moreover, we embedded an across session criteria in the feature selection process that perfectly fits the aforementioned single session enrollment, across session testing scenario. The proposed feature selection algorithm is computationally efficient, i.e. linear in number of features and quadratic in number of samples and we have demonstrated that it can be applied to large datasets with thousands of features and samples. In addition, we examined the effectiveness of the proposed feature selection method for ECG recognition where enrollment and testing are in different body postures. Also, we evaluated the proposed method for biometric recognition using TEOAE signals. Performance of the proposed method was compared against seven state-of-the-art feature selection algorithms as well as another six approaches in the area of ECG and TEOAE biometric recognition. On average, the proposed feature selection algorithm achieved an EER of 4.63% which is about 2 times better than the best comparison method.

Further in chapter 4, we proposed a cross-view matching method to cope with the effect of heart rate. In this context, “view” refers to different heart rates. We used signals recorded after exercise as well as rest condition. In the proposed approach a sample is represented by its similarity to subjects of the auxiliary set in the respective weighted space. Ideally, such representation should be view-independent, i.e. same-labelled samples situated in different views should have similar representations. To this end, the proposed method computes a set feature weights for each view such that the cross-view matching error is minimized and also the margin around each sample is maximized. We have shown the generalization of the proposed
method on other applications including visible light versus near infrared face recognition and cross-pose face recognition. The proposed algorithm was compared against 5 state-of-the-art cross-view matching methods. On average, the proposed algorithm improved the EER in cross heart rate experiment by 2.2%. In the near infrared versus visible light face recognition experiment, it improved the recognition rate by 5.3% compared to the best comparison method. Finally, in the cross-pose face recognition experiment it performed on average about 10% better than the best comparison method.

Overall, in this thesis a novel application for ECG was introduced where ECG was combined with fingerprint for spoof detection. In addition, this thesis attempted to provide algorithms to address some challenges in ECG biometric recognition, though they may also be used on other medical biometrics as well. The results established in this thesis provides useful insights to design and deploy biometric systems based on medical signals in practical scenarios.

5.2 Future Work

The following possible directions can be considered for further improvements and extensions of the current work:

- In the current work we have used different features such as continuous wavelet transform and autocorrelation for signals and SURF features for images. However, other types of features can also be investigated to further enrich the feature pool.

- The cross-view matching method presented in chapter 4 works on pairs of samples situated in different views. In the current work, samples are paired randomly. As a future direction, a more rigorous strategy can be developed for pairing ECG samples such that among all possible pairing options, those that result in a smaller cross-view matching error are considered.

- ECG signal is the most widely explored medical biometric and in this thesis we focused on fusion of ECG and fingerprint. However, other modalities such as Photoplethysmogram (PPG) that can also be recorded from fingertips can used for fusion.
• In the unified approach for fusion of ECG and fingerprint, fusion was performed in Score level. As a future direction, fusion at feature level or other advanced fusion methods can also be explored. Although, this may need to extend the current database to include more subjects.
Appendix A

Cross-view matching error: Deriving Equation (4.25)

\[ ||f_x^T D^{(i)}_x - f_y^T D^{(i)}_y||_2^2 = f_x^T D^{(i)}_x D^{(i)T}_x f_x - f_x^T D^{(i)}_x D^{(i)}_y f_y \]

\[ - f_y^T D^{(i)}_y D^{(i)T}_x f_x + f_y^T D^{(i)}_y D^{(i)}_y f_y \]

\[ = \begin{pmatrix} f_x^T & f_y^T \end{pmatrix} \begin{pmatrix} D^{(i)}_x \\ 0 \end{pmatrix} \begin{pmatrix} D^{(i)T}_x & 0 \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix} \]

\[ - \begin{pmatrix} f_x^T & f_y^T \end{pmatrix} \begin{pmatrix} D^{(i)}_x \\ 0 \end{pmatrix} \begin{pmatrix} 0 & D^{(i)T}_y \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix} \]

\[ - \begin{pmatrix} f_x^T & f_y^T \end{pmatrix} \begin{pmatrix} 0 \\ D^{(i)}_y \end{pmatrix} \begin{pmatrix} D^{(i)T}_x & 0 \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix} \]

\[ + \begin{pmatrix} f_x^T & f_y^T \end{pmatrix} \begin{pmatrix} 0 \\ D^{(i)}_y \end{pmatrix} \begin{pmatrix} 0 & D^{(i)T}_y \end{pmatrix} \begin{pmatrix} f_x \\ f_y \end{pmatrix} \]

\[ = f^T D^{(i)}_x f - f^T D^{(i)}_y f - f^T D^{(i)}_y f + f^T D^{(i)}_y f \]

\[ = f^T D^{(i)} f \quad \text{(A.1)} \]
where

\[
\begin{align*}
    f &= \begin{pmatrix} f_x^T & f_y^T \end{pmatrix}^T, \\
    D_{xx}^{(i)} &= \begin{pmatrix} D_{x}^{(i)} & 0 \\ 0 & 0 \end{pmatrix}, \\
    D_{xy}^{(i)} &= \begin{pmatrix} 0 & D_{y}^{(i)} \\ 0 & 0 \end{pmatrix}, \\
    D_{yx}^{(i)} &= \begin{pmatrix} 0 & 0 \\ D_{y}^{(i)} & D_{x}^{(i)} \end{pmatrix}, \\
    D_{yy}^{(i)} &= \begin{pmatrix} 0 & 0 \\ 0 & D_{y}^{(i)} \end{pmatrix},
\end{align*}
\]

\[
    D^{(i)} = D_{xx}^{(i)} - D_{xy}^{(i)} - D_{yx}^{(i)} + D_{yy}^{(i)}
\]

\[
    = \begin{pmatrix} D_{x}^{(i)} D_{x}^{(i)T} & -D_{x}^{(i)} D_{y}^{(i)T} \\ -D_{y}^{(i)} D_{x}^{(i)T} & D_{y}^{(i)} D_{y}^{(i)T} \end{pmatrix}.
\]
Bibliography


[34] N. Kumar and A. G. Andreou, “Heteroscedastic discriminant analysis and reduced rank
hmmss for improved speech recognition,” *Speech communication*, vol. 26, no. 4, pp. 283–


[37] F. Agrafioti and D. Hatzinakos, “Signal validation for cardiac biometrics,” in *Acoustics
Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE,

[38] J. Wang, M. She, S. Nahavandi, and A. Kouzani, “Human identification from ecg signals
via sparse representation of local segments,” *Signal Processing Letters, IEEE*, vol. 20,

[39] G. L. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli,
and S. Schuckers, “First international fingerprint liveness detection competitionlivdet

2011fingerprint liveness detection competition 2011,” in *Biometrics (ICB), 2012 5th IAPR

[41] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli, and S. Schuckers,
“Livdet 2013 fingerprint liveness detection competition 2013,” in *Biometrics (ICB), 2013

[42] C. Sousedik and C. Busch, “Presentation attack detection methods for fingerprint recogni-

electrocardiogram (ecg).” *IEEE transactions on bio-medical engineering*, 2016.


