Cardiac Signals: Remote Measurement and Applications

Abhijit Sarkar

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

A. Lynn Abbott, Co-Chair
Zachary Doerzaph, Co-Chair
Devi Parikh
Daniel J. Stilwell
Jason Xuan

June 27, 2017
Blacksburg, Virginia

Keywords: Electrocardiogram, Blood volume pulse, Remote plethysmography, ECG biometrics, PPG biometrics, Skin detection, Driver monitoring,Face anti-spoofing.

Cardiac Signals: Remote Measurement and Applications

Abhijit Sarkar

(ABSTRACT)

The dissertation investigates the promises and challenges for application of cardiac signals in biometrics and affective computing, and noninvasive measurement of cardiac signals. We have mainly discussed two major cardiac signals: electrocardiogram (ECG), and photoplethysmogram (PPG).

ECG and PPG signals hold strong potential for biometric authentications and identifications. We have shown that by mapping each cardiac beat from time domain to an angular domain using a limit cycle, intra-class variability can be significantly minimized. This is in contrary to conventional time domain analysis. Our experiments with both ECG and PPG signal shows that the proposed method eliminates the effect of instantaneous heart rate on the shape morphology and improves authentication accuracy. For noninvasive measurement of PPG beats, we have developed a systematic algorithm to extract pulse rate from face video in diverse situations using video magnification. We have extracted signals from skin patches and then used frequency domain correlation to filter out non-cardiac signals. We have developed a novel entropy based method to automatically select skin patches from face. We report beat-to-beat accuracy of remote PPG (rPPG) in comparison to conventional average heart rate. The beat-to-beat accuracy is required for applications related to heart rate variability (HRV) and affective computing. The algorithm has been tested on two datasets, one with static illumination condition and the other with unrestricted ambient illumination condition.

Automatic skin detection is an intermediate step for rPPG. Existing methods always depend on color information to detect human skin. We have developed a novel standalone skin detection method to show that it is not necessary to have color cues for skin detection. We have used LBP lacunarity based micro-textures features and a region growing algorithm to find skin pixels in an image. Our experiment shows that the proposed method is applicable universally to any image including near infra-red images. This finding helps to extend the domain of many application including rPPG. To the best of our knowledge, this is first such method that is independent of color cues.

Cardiac Signals: Remote Measurement and Applications

Abhijit Sarkar

(GENERAL AUDIENCE ABSTRACT)

The heart is an integral part of the human body. With every beat, the heart continuously pumps oxygen-enriched blood to providing fuel to our cells and thus enabling life. The heartbeat is initiated by electrical signals generated in the heart muscles. This electrical activity, which are often governed by our autonomic nervous system, can be measured directly by electrocardiogram (ECG) using advanced and often obtrusive instrumentation. Photoplethysmogram (PPG), on the other hand, measures how the blood volume changes and can be readily measured with inexpensive instrumentation at certain locations (e.g. at the fingertip). The ECG and PPG are widely used cardiac signals in medical science for diagnosis and health monitoring. But, these signals hold greater potential than just its medical diagnostic applications. In this work, we have mainly investigated if these signals can be used to identify an individual. Every human heart differs by their size, shape, locations inside body, and internal structure. This motivated us to represent the signals using a mathematical model and use machine learning algorithm to identify individual persons. We have discussed how our method improves the identification accuracy and can be used with current biometric methods like fingerprint in our phone.

The measurement procedures of cardiac signals are often cumbersome and need instruments which may not be available outside medical facilities. Therefore, we have investigated alternative method of remote photoplethysmography (rPPG) that are relatively inexpensive and unobtrusive. In this dissertation, we have used face video of an individual to extract the heart rate information. The flow of blood causes small changes in the color of face skin. This is not visible to human eyes without digital magnification, but we have shown how knowledge of distinct behavior of human heart rate and use of advanced computer vision algorithms helped us to extract vital signals like heart rate with a significant accuracy.

In addition, to measure rPPG using face video, we integrated a method for automatic detection of skin from images and videos. Existing skin detection methods depended on color information which is not always available within available video sources. We have developed a novel standalone skin detection method to show that it is not necessary to have color cues

for skin detection. Our method relies on the context and the texture based appearance of skin. To the best of our knowledge, this is first such method that is independent of color cues.

In summary, the dissertation investigates the promises and challenges for application of cardiac signals in biometrics and nonobtrusive measurement of cardiac signals using face video.

dedicated in memory of dadu and dida ...

Acknowledgments

There are many people who have directly or indirectly contributed to this dissertation. Some of them I met here at Virginia Tech, the association with the rest continues from time even before.

First, I would like to thank my advisors, Dr. Lynn Abbott and Dr. Zachary Doerzaph. Your continuous help, co-operation and guidance have been the prime reason that I could complete this dissertation. Thank you for believing in my abilities. Both of you provided me with an open and amicable environment, where I could freely discuss my research ideas. You gave me the freedom to explore different research avenues and inspired me to try diverse problems. I thoroughly enjoyed those uncountable number of hours that I spent in your office discussing science and technology. And, thank you for patiently acknowledging all those last moment requests.

I would like to thank Dr. Devi Parikh for introducing me to computer vision through her courses "Advanced Topics in Computer Vision" and "Computer Vision Systems". I thank her for all those insightful discussions on research in general. I would also like to thank my other committee members Dr. Daniel Stilwell and Dr. Jason Xuan for their suggestions to improve this documents.

I would like to thank Dr. Krisnan Ramu for giving me the first research opportunity at Virginia Tech. I thank Dr. Scott Midkiff, the head of the department of ECE, (Virginia Tech, in 2012) for the help and valuable suggestions when I was going through the toughest time.

A special thank to Dr. Natalia Henao-Guerrero, from Virginia-Maryland college of veterinary medicine for helping me with my experimental setup. I thank Dr. Dhruv Batra from department of ECE for occasional discussions on research.

I have been associated with Virginia Tech Transportation Institute (VTTI) for a larger

portion of my time at Virginia Tech. I would like to thank VTTI for offering me graduate research assistantship over the years to support my graduate study. Association with VTTI has given me the opportunity to work in diverse projects and to work with some of the best minds in transportation research. In particular, I would like to thank Jeremy Sudweek, Miao Song, Mike Mollenhauer , Reginald Viray , Melissa Hulse, Leslie Harwood, Loren Stowe, Andy Petersen, Miguel Perez, Jon Hankey, Kayla Sykes, Rebekah Duke, Kelly McGowan, Ryan Johnson and Suzie Lee. I would like to thank National Surface Transport Safety Center for Excellence (NSTSCE) for partially supporting rPPG work in Chapter 4.

I would like to thank my colleagues from Delphi Inc. (Ron, Jim), Ramu Inc. (Nimal, Gray, Hong-sun), and Apple Inc. who gave me a glimpse of how industrial research is carried out.

Also, I like to thank all CESCA members, for giving me the platform to discuss research ideas. I thank all of my lab members at CESCA Computer Vision lab, especially to Ahmed who is a great friend and colleague.

During my stay at Blacksburg, I met some great people whose friendship I shall cherish for rest of my life: Gupta (I enjoyed every bit of our friendly banters and chat), Kriti Da (the perfect senior and mentor every junior would love to have), Abhra Da and Banga Da (the other seniors whom I always trusted for expert opinions), Pal (for those wonderful music sessions and discussions on music), Bikram (for being an awesome roommate and friend), Nath and Santa (the duo who holds a special place in my heart. Their love and mischievousness made my life in Blacksburg colorful and fun), folks from the Friday night gang (Suvo, Pal, Chatterjee, Basu, Surya, Roy, Wrik, Pandey, Pola, Krish, Mithun), the AID team members (Naresh, SRC, Vyas, Jaideep Da), and the 1309 loop members (Nath, Santa, Ranit, Lekha, Suchi, Leo, Purna, Donu, Sreeya, Srijan, Prasen Da, Arnab Pal, and GB).

There are some other people who are away from Blacksburg, but in one way or others have left their impact. Especially, I thank my childhood friend Anupam (we became friends even before we knew what friendship is), my high school friend Biswa (for all of those interesting discussions about academia), Nalini babu (my mathematics teacher and guru), and Don mama (Dr. Anirban Basu, for suggestions at some of the crucial moments of my life).

A special thank to my wife Shreya who has been my inseparable partner accompanying me in this most unpredictable road I have ever travelled. Thank you for tolerating all of my whims and idiosyncrasies patiently. Through her I got introduced to two more wonderful persons, Dr. Rabishankar Mitra (my father-in-law) and Dr. Sanghamitra Mitra (my mother-

in-law and my cooking guru).

Thank you Baba (my father, Mr. Amalesh Sarkar, who helped me to find my love and passion for mathematics and inspired me to think independently) and Ma (my mother Smt. Rupa Sarkar, whose strict discipline helped me in my initial days as a student) for your unconditional support, love and blessings. Thanks to my little brother Tutun for taking up all of the responsibilities on behalf of me. And finally, I would like to thank dadu and dida (my grandparents, Sri Arun Kumar Biswas and Smt. Renuka Biswas) who has been the biggest inspiration in my life.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

This dissertation explores the scope of remote measurement of physiological signals and related applications. Physiological signals have been a major tool for medical diagnosis and prognosis. Several medical applications have often benefited from these measurements. Electrocardiography (ECG), electroencephalography (EEG), blood pressure (BP), photoplethysmography (PPG), respiration rate, body and skin temperature, and skin conductance are some of the vital signals conventionally used in medical science. Recent interests in wearable health monitoring systems (WHMS) and body area networks (BAN) have propelled the scope and research interest for measurement and applications of these signals outside of standard medical centers, clinics and facilities [10, 11]. This dissertation mainly focuses on measurements and related applications of cardiac signals. Electrocardiogram and pulse oximetry are the most common methods to study cardiac activity of a person. Apart from traditional ECG and PPG based measurement of cardiac signals, a number of non-contact methods have been investigated over the last decades. Radar-Doppler vibrometry [12], bio-impedance [13], laser Doppler vibrometry [14], and thermal imaging [15] are some of the methods worth mentioning. Kranjec et al. have given a summary of recent advances in emerging technology in heart rate measurement [16]. Advances in computer vision, imaging techniques and availability of high quality cameras have also inspired researchers for video based measurements and analysis of vital signals. Although all of them have their own limitations, they have shown promise for comfortable measurement of cardiac signals in the future. In this dissertation, we discuss video based heart rate measurement as a possible

non-contact method, and we provide experimental evaluation. Also, we explore novel applications for the two most commonly used cardiac signals, ECG and PPG. These cardiac signals primarily reflect the behavior of human heart and related cardiovascular activities. ECG is the measurement of the electrical activities in the heart muscles. BVP, on the other hand, measures the change in the volume of blood flowing through the body with every cardiac beat. PPG is the standard method of measuring blood volume pulse (BVP), often using a fingertip based pulse oximeter. In the course of this document, we shall mainly discuss the following:

- How the shape of these signals convey the identity of an individual.
- How non-invasive measurements of these signals are feasible in natural conditions.
- Scope of cardiac signals in daily life.

Over the last decade, we have witnessed an increasing use of portable devices embedded with multiple sensors. Advances in materials science, embedded systems, and sensory mechanisms have significantly reduced sizes of the devices while enhancing their performance. Recent breakthroughs in artificial intelligence (AI) and high power embedded computing have also motivated scientific communities and industries to develop state of the art portable devices that are accessible to the public. For example, a mobile phone today is an ensemble of multiple sensors and AI algorithms. These applications are targeted for wide variety of applications and a diverse audience.

For cardiac signal measurements, a number of portable noninvasive sensors and devices are available. A survey of contact and non-contact sensor types for ECG is given in [17]. A popular wearable device is the Fitbit, which can estimate heart rate along with other physiological measurements [18]. Apple [19] also has an interest in ECG-based sensing. The AliveCor [20] heart monitor allows an individual to obtain an ECG-like signal by touching 2 electrodes that are coupled with a smart-phone. The vendor claims the ability to recognize cases of atrial fibrillation (AF) from the sensed ECG signals. The Nymi band [21] is worn on the wrist, and provides electrodes that allow sensing of an ECG signal. It has been used for biometric authentication, e.g., to make credit card payments verified by the user's unique heartbeat. Body Area Networks (BAN) have also been used for heart monitoring.' The emphasis in [22] is low-power contact sensors within a BAN for acquiring PPG signals related to sensing of heart rate. Another PPG sensor is reported by [2], in the form of an earring. Piezoelectric sensors, which are sensitive to pressure, have also been proposed for

sensing blood volume pulses [1]. In summary, researchers and industry are excited by the potential to sense and utilize signals that can be obtained from the cardiovascular system. Following these promises and growing interests, we explore the scope and motivation of our work with the following applications in mind.

1.1.1 Cardiac Biometrics for Personal Security

With advances in mobile technology, we are transferring enormous amounts of data over the internet. These data often include personal and sensitive information that may lead to identity theft or other critical problems. With growing interest in mobile usage and the internet of things (IoT), the chances of attack are also increasing. Currently, more than 19 billion devices including mobile phones and home appliances, are connected through the internet, and this number is expected to increase to 30 billion by the end of 2020 [23]. Therefore we need to develop new modalities that can improve personal security. An attack is possible at multiple levels of operation including hardware level, data encryption level or user interaction level. In this work, we mainly concentrate on security in the user authentication level using biometrics.

Biometrics is a specialized field of research which focuses on identifying individuals from their physical attributes. Common biometric modalities include recognition using features from the fingerprint, retina, face, shape of hand or a combination of these. Recently fingerprint detections are widely used in mobile devices as biometric identification (BioID) and authentication. As an alternative to conventional password based access control, BioIDs are becoming increasingly popular as a more convenient and more robust identification method. One of the major drawbacks of any image based biometric system is that these systems rely on one shot evaluation. A copy of the fingerprint or retina image or face image (photo or video) of a subject may fool the system. Recent reports [24, 25] show that fingerprint spoofing in mobile phones may not need complex tools, but may be fooled by *play clay*. Researchers have cited similar constraints and risks in recent literature (e.g. [26, 27]). One of the basic reasons is that these systems often lack the proof of liveness of the subject during authentication. While liveness detection is an active field of research in the biometric community, we investigate new biometric modalities using physiological signals to address these limitations.

Physiological signals reflect the intrinsic characteristics of a person. ECG and PPG signals capture the structural and operational behavior of the cardiovascular system. The

inter-individual variability of the acquired signal is the result of the physiological differences between the hearts. This includes the differences in the Purkinje system, in the orientation of heart muscles, their tensile characteristics, the electric conductivity of different parts of the heart, and its size and position in the body [28]. These differences make cardiac signals from each individual unique and this fact motivates researcher to use these signals as biometric signature. Da Silva et al. [29] have stated that ECG modality satisfies all factors to be an accepted modality defined by Jain et al. [30]:

> "... it can be found in virtually all living humans (Universality), its authentication capabilities for circumscribed groups of individuals has been shown (Uniqueness), it can be easily acquired using suitable devices (Measurability), it has been shown to perform accurately for subsets of the population (Performance), the off-the-person approach has made it acceptable (Acceptability), and it's not easily spoofed as it depends on an internal body organ, the heart (Circumvention) particularly."

Furthermore, the temporal nature of cardiac signals proves intrinsic liveness of the subject. This temporal nature holds the capacity for continuous authentication, as compared to single validation for image based methods. Most researchers have used the differences in the shape and morphology of the ECG and PPG signals as the biometric trait. Odinaka et al. [31] have given a comprehensive analysis with comparison of ECG biometrics methods.

1.1.2 Emotion Recognition

Researchers have used analysis of facial, gestural and vocal responses to recognize human emotions. In certain cases these responses show clear indication of the emotion of a person. At the same time, humans have been trained to control these external expressions. Therefore, human emotions exhibit complex patterns. These patterns are often challenging for another human or machine to understand based on their external behavior only. But, in most cases, the human is not trained to control the intrinsic behavior of an emotion. Over the last decade, researchers from the affective computing have shown that different emotions can be measured from skin conductance, EEG and heart rate variability (HRV) [32, 33]. In the current work, we shall discuss HRV which reflects the activity and balance in the autonomic nervous system (ANS). HRV refers to the variations of instantaneous heart rate (HR) over time. The balance in the sympathetic and parasympathetic nervous system dictates the

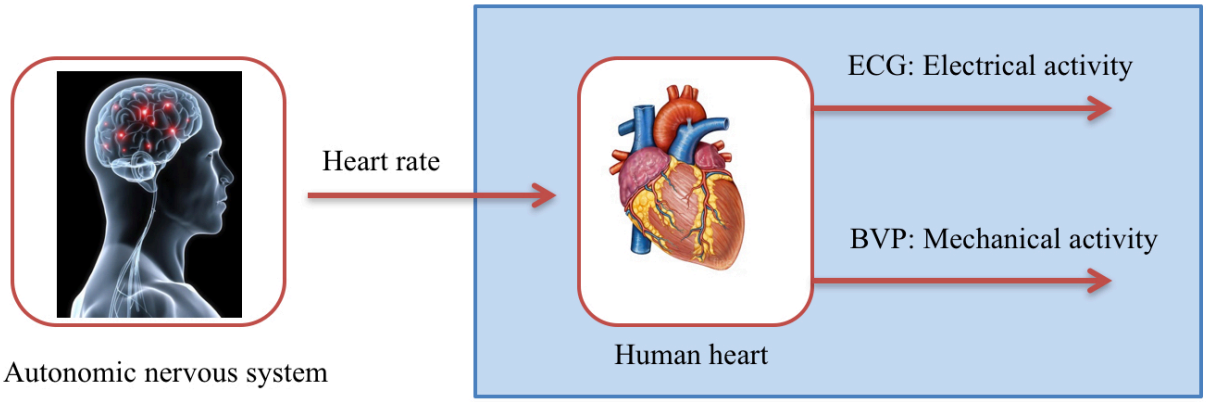


Figure 1.1: Relation of autonomic nervous system and heart rate. Heart rate and its variability is a function of the present condition of the ANS. We can measure heart rate using ECG or PPG. Hence by these measurements we can assess the psychophysiological conditions reflected in the ANS.

duration of every heartbeat. Hence, the rhythm of the heart is controlled by the state of the nervous system. Again, the state of ANS is governed by the psychophysiological condition of a person. Conditions like fatigue, drowsiness, stress or any human emotion are reflected in the balance of the ANS. Therefore, knowledge of HRV can help us understand the psychophysiological condition of a person. Acharya et al. [34] have provided a comprehensive review of heart rate variability, including its analysis procedure in time and frequency domain and its relation to ANS under different states of ANS. To measure the HRV, we need to first record the heart rate information. ECG and PPG are the most common methods for this measurement. Figure 1.1 depicts the schematic where the heart receives a command signal from the ANS. We can measure the activity of the heart (mainly heart rate) by recording ECG or PPG. In other words, we can measure the behavior of the ANS by measuring ECG or PPG signals from the heart.

1.1.3 Driver Monitoring

One of the major motivations of our work has been noninvasive assessment of psychophysiological condition of drivers and knowing their role in the safety critical events like crash or near crash. Factors such as the cognitive workload, stress level of a driver, inattentions including drowsiness, fatigue, driving under influence (DUI), as well as on and off road dis-

tractions are example of many contributing factors leading to a crash. A report from the National Highway Traffic Safety Administration (NHTSA) states that 35092 people lost their lives and more than 2.44 million people were injured by road accidents in the year 2015 [35]. More than 6.3 million non-fatal crashes have been reported including injuries and property damage in the same year. These statistics have led to a naturalistic driving study (NDS), which is an effective way to understand the role of drivers' performance and behavior in traffic safety. For example, the Strategic Highway Research Program 2 Naturalistic Driving Study (SHRP2) data contains numerous variables to record the interaction of a driver with the vehicle and traffic by monitoring data such as speed, acceleration, following distance, maneuver selection, and braking patterns to name a few. The main shortcoming of the study is that it cannot directly record and monitor a driver's physiological state. Rather, data reductionists need to watch videos of the driver (also available with the SHRP2 data) and make subjective assessments of the drivers' state (e.g., fatigued, drunk driving, etc.). With the enormous amount of video data, it is infeasible to manually watch and analyze all the videos. Hence the research community would benefit from automated system which will be able to read such situations. Objective measures of physiological variables, like blood pressure, heart rate (HR), and respiration rate may be better indicators of a driver's true psychophysiological state. However, the major constraint for measuring physiological variables is that the available methods are intrusive in nature and interfere with the natural behavior of a driver. Most of the measurement procedures available today need devices with wires and heavy instruments which often interfere with the natural action. These types of instrumentation and measurement procedures are not possible to implement in natural scenarios, especially for NDS. Therefore in order to investigate the psychophysiological condition of a driver, one needs to implement strategies that do not impede natural driving.

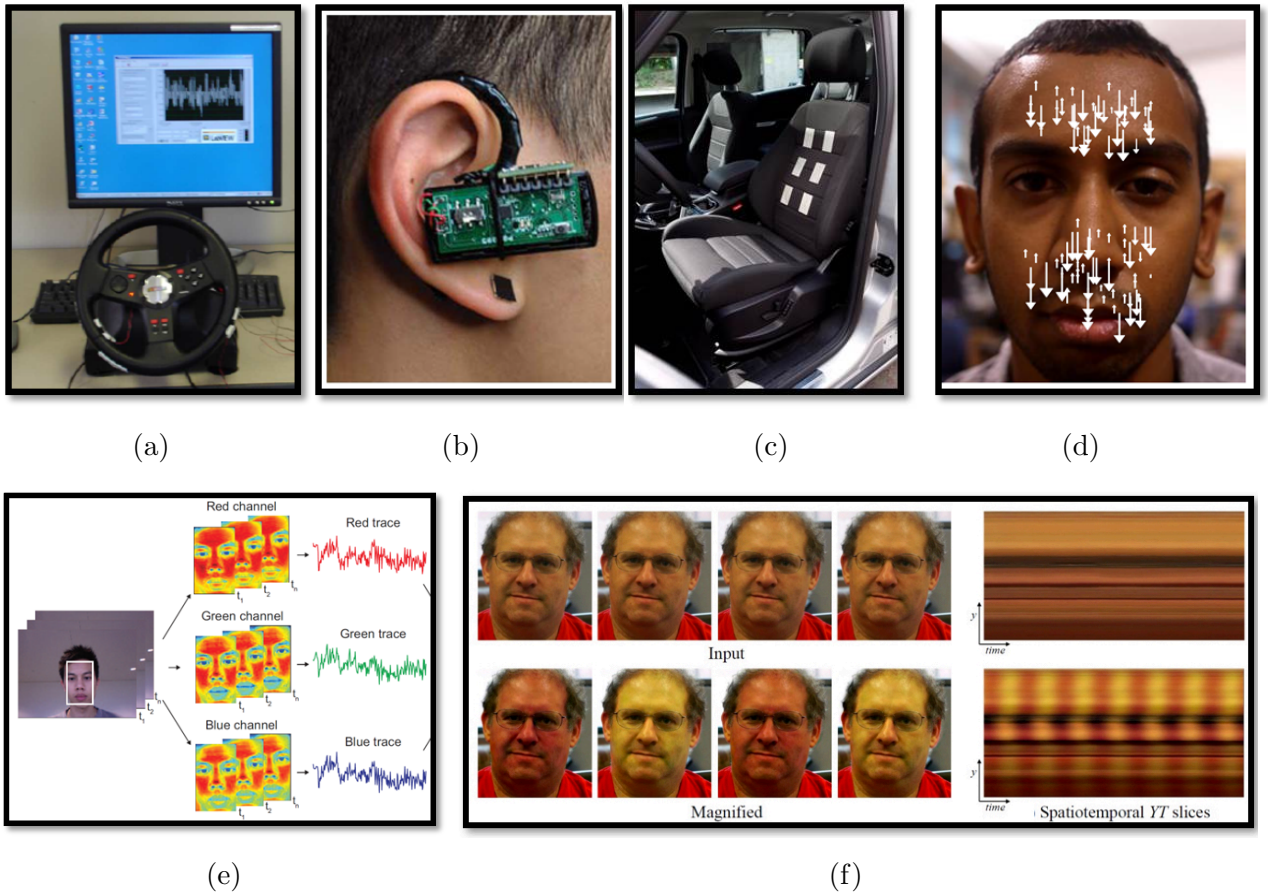
1.1.4 Noninvasive Measurement of Plethysmography Signals

Although biosignals are important for understanding the physiological and emotional condition of a person, extensive use of these signals is restricted by the rigorous measurement procedures. Expensive medical instruments and their complexity often restrict their uses outside medical clinics. For example, heart rate is one of the major parameters extracted from ECG signals. Standard ECG machines require a person to wear adhesive gel and electrodes. This is often uncomfortable for the subject and hinders their natural movement. It is tiring and uncomfortable for patients who need continuous monitoring in the hospital. Even wearing a pulse oximeter for long periods may cause discomfort. Additionally, as mentioned

in the earlier section, applications like driver monitoring and health monitoring shall also benefit with remote and noninvasive measurement of HR.

Over the last decade several other methods and measuring instruments have been engineered. Figure 1.2 shows some of those works reported in literature. Lin et al. [1] have used an embedded pulse rate sensor in the steering wheel for driver monitoring. Poh et al. [2] presented an ear wearable plethysmography sensor for a motion tolerant pulse measurement. Ford has demonstrated a HR monitor embedded in a car seat [36]. With advances in computer vision and camera technology, several researchers have used video data to extract heart rate information. Guha et al. [37] have reported a novel method to track the motion of the head. With every cardiovascular pulse, the thrust of blood causes a Newtonian reaction in the veins around neck and head. This creates a small pulsating movement in the head. This is not generally visible, but by carefully tracking points in the head and face, this motion can be measured. The main merit of this method is that it does not depend on the ambient illumination, but the accuracy of results depends largely on the natural motion of the head. There have been two other video monitoring strategies, in which BVP can be measured directly from a stream of video data. Poh et al. [3] have used a blind source separation (BSS) method to extract this signal. Although this method shows interesting results, there lies an ambiguity in the choice of correct channel in the required independent component analysis (ICA, a class of BSS). Secondly, this method is designed for color video, whereas most SHRP2 face videos are grayscale only. Poh et al. have used three different color channel coding schemes based on face skin and applied an ICA to extract the pulse signal. McDuff et al. [38] have shown in their following work that this method gives best results while using a five color-band camera and 12 bit image depth. It also requires very high resolution image with large number of face pixels. The high-quality video is very different from any available real life scenario and large scale affordable naturalistic implementation. Wu et al. [4] have used a method of Eulerian video magnification (VidMag) to extract a BVP signal from a video sequence and magnify it to make it visible to normal eyes.

We find the VidMag method to be more suitable for naturalistic conditions, for several reasons. This method traces the changes in pixel intensities. Hence, it can be used for the grayscale video data of SHRP2 as well as future studies which use color video. We have tested a video in RGB and in a transformed grayscale version for comparison. In both cases, the estimated blood volume changes were identical. Secondly, VidMag uses a Gaussian pyramid which reduce spatial noise. Naturalistic videos often contain noise either from the camera sensor or from the surroundings. In this report we have used VidMag to



(a) (b) (c) (d)

(e) (f)

Figure 1.2: Recent work in remote PPG: (a) Pulse rate sensor embedded in steering wheel [1]. (b) Photo-plethysmography sensor in the ear [2]. (c) Pulse measurement sensor installed on the seat by Ford. (d) Pulse rate estimation from subtle head movement. (e) Pulse rate using blind source separation [3]. (f) Video magnification for pulse rate measurement [4].

.

evaluate the performance of the remote plethysmography. The next section briefly describes the algorithm.

1.1.5 Face Anti-Spoofing

Face biometric is a widely used modality for access control. Face recognition techniques are used in different devices to grant access to authorized personnel. The domain of application spans from online banking services to access of restricted premises. Unfortunately, the increasing reliability of biometric methods has instigated imposters to gain access by means of spoofing. It has been shown that use of a simple photograph of a client (one who has access) [39] can fool a face biometric system if proper anti-spoofing countermeasures are not included in the algorithm. With wide availability of images of individuals in social media and access of advanced imaging and video recording devices to the public, other modes of spoofing attack have emerged. These include the playing of a recorded video of an individual as input to the system. A more advanced method could be the use of a 3D mask. All of these possible spoof attacks have instigated several researchers to study anti-spoofing measures for face biometrics similar to liveness detection for fingerprint and iris.

1.2 Contributions in this Dissertation

We have used techniques from computer vision, machine learning, and signal processing to study the scope of cardiac signal for diverse and novel applications:

- In Chapter 2 and Chapter 3, we have proposed a novel method that uses a dynamical model of the heart to represent the morphology of the cardiac signals (ECG and PPG) as a function of the parameters of the model. We have used these parameters as the feature vectors and shown that they preserve the cardiac signature of an individual. We have also shown that the proposed method is robust to external emotional stimuli.
- In Chapter 4 we have experimentally evaluated the feasibility of remote heart rate measurement in naturalistic conditions. With the help of the video magnification algorithm [4], we have analyzed face videos of different individuals with synchronous ground truth heart rate measurements, and we have proposed a post processing method based on frequency analysis of the temporal signals. Our method has helped to improve

the estimation of the instantaneous heart rate measurement and propose a metric to select skin patches suitable for rPPG. Also we have discussed the limitations of such methods in different scenarios.

- In Chapter 5 we have introduced a novel skin detection algorithm that does not require the use of color cues. Skin detection is an integral part of remote plethysmography (rPPG). We have proposed a universal skin detection algorithm using local texture features. Historically, all of the skin detection methods have used color cues. To the best of our knowledge, this is the first instance where skin detection is performed without using any color information. The proposed method is adaptive to the applied image and can be extended for different spectra such as near infrared (NIR).

- In Chapter 6, we have discussed a micro-texture based face anti-spoofing counter measure. Borrowing the texture features from the work of [40], we have experimentally demonstrated the robustness of texture features in capturing the differences between real and spoof images and replay videos. We have also discussed whether texture based methods are robust in case of partial occlusion.

Following our work in this document we have proposed possible directions and open research questions for each of the topics in Chapter 7 .

1.3 List of Publications

Most of the work reported in this dissertation has appeared in the following lists of publications:

- A. Sarkar, A. L. Abbott, and Z. Doerzaph, 2014, September. Assessment of psychophysiological characteristics using heart rate from naturalistic face video data. In International Joint Conference on Biometrics (IJCB) (pp. 1-6).

- A. Sarkar, A. L. Abbott, and Z. Doerzaph, 2015, September. ECG biometric authentication using a dynamical model. In Biometrics Theory, Applications and Systems (BTAS), IEEE 7th International Conference on (pp. 1-6).

- A. Sarkar, A. L. Abbott, and Z. Doerzaph, and K. Sykes, 2016, January. Evaluation of video magnification for nonintrusive heart rate measurement. In Control, Measurement and Instrumentation (CMI), IEEE First International Conference on (pp. 494-498).

- A. Sarkar, A. L. Abbott, and Z. Doerzaph, 2016, September. Biometric authentication using photoplethysmography signals. In Biometrics Theory, Applications and Systems (BTAS), IEEE 8th International Conference on (pp. 1-7).
- A. Sarkar, A. L. Abbott, and Z. Doerzaph, 2017, March. Universal Skin Detection Without Color Information. In Applications of Computer Vision (WACV), IEEE Winter Conference on (pp. 20-28).
- A. Sarkar, Z. Doerzaph, and A. L. Abbott, Video Magnification to Detect Heart Rate for Drivers, Research report, National Surface Transport Safety Center for Excellence (NSTSCE), (To be published).

Chapter 2

Cardiac Biometrics Using ECG Signals

2.1 Introduction

Biometric authentication and identification systems have assumed increasing importance in recent years. Because of this importance, there is a continuing need to explore and develop new biometrics capabilities. Each of the different sensing modalities, including traditional iris-based and fingerprint-based approaches, exhibits particular strengths and weaknesses [41, 42]. The main focus of interest in this work is authentication based on electrocardiogram (ECG) signals (Figure 2.1) and photoplysmogram (PPG) signals. Cardiac biometrics represents an emerging modality that has strong potential to complement or supplement existing biometrics approaches. Authentication systems that depend on signals from an individual ' s heart are attractive because of universality (the signals are present in all living individuals), and because the signals are difficult to falsify (as they derive from the autonomic nervous system, ANS). The geometry and physiology of the heart differ from one individual to another [43]. The tension in the heart muscles, shape, size, and position of the heart, conductivity of heart muscle are different for different person. And because of these variations it may be possible to distinguish different individuals based on observed electrical activity.

An ECG represents electrical activity that causes contraction and relaxation of cardiac muscle fibers [44]. This electrical activity is measured by ECG recording devices. Tradition-

ally, a basic ECG is obtained using electrodes that are placed on the surface of the body, measuring the potential difference between 2 electrodes. A typical single ECG pulse is shown in Figure 2.1a. A possible objection to the use of ECG as a biometric modality is that ECG sensing is relatively invasive, as compared to more common modalities such as fingerprint scanning or keystroke monitoring. Although less invasive techniques are possible [17], most ECG sensing is performed using electrodes that are placed at several locations on the body.

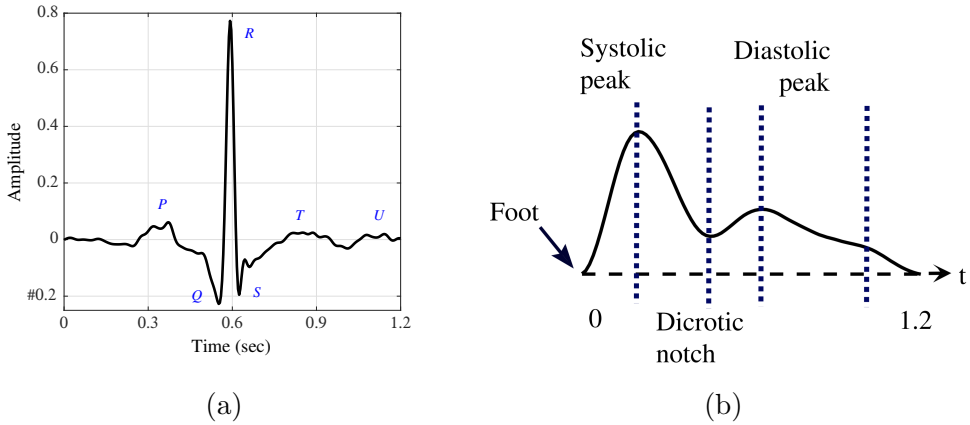
Photoplethysmography (PPG) is an alternative to ECG for capturing signals related to cardiac activity. The resulting PPG waveform is related to volumetric changes that occur underneath the skin as the heart pumps blood throughout the body by aforementioned contraction and expansion. An example of a single PPG beat is shown in Fig. 2.1b. A typical PPG sensor uses an illumination source to transmit light through the skin, and captures a time-varying waveform using a photodetector [45]. Example devices that employ PPG are pulse oximeters, which are commonly placed on a fingertip or in an ear. Blood volume pulses (BVP) are sensed in this way, and are often used to estimate heart rate. An advantage of PPG-based biometric analysis is that the sensors typically are less invasive than those for ECG.

In this document, we explore the potential of both the signals in two chapters. This chapter mainly focuses on the ECG based biometric method. Chapter 3 discusses the scope and experimental analysis of PPG based biometric method. In both cases we have used a dynamical model to model the heart. All the signals tested for this work are from publicly available databases([46, 5]), and are obtained using common techniques.

2.2 Previous Work in ECG Biometrics

Research in ECG-based biometrics can broadly be divided into two major categories: those depending on the detection of fiducial points, and those based more broadly on appearance of the signal. In fiducial based methods, the emphasis is on detection of particular distinctive parts of the ECG signal. By convention, the labels are assigned as shown in Figure 2.1. Biel et al. [47], Israel et al. [48], Kyoso and Uchiyama [49], Irvine and Israel [50], and Singh and Singh [51] have used all or a subset of these six fiducial points for ECG biometrics. These algorithms rely on careful localizations of fiducial points. Hence the performance of these methods largely depends on the accuracy of the fiducial point detection.

An appearance based model, on the other hand, does not rely on exact localization of

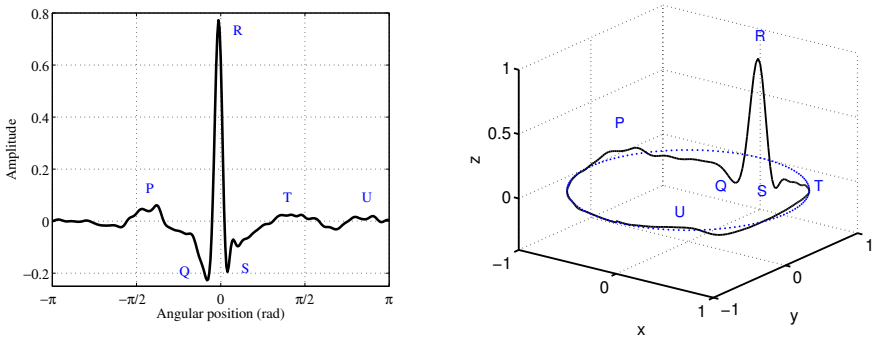


(a) (b)

Figure 2.1: Examples of ECG and PPG signals. (a) A typical electrocardiogram (ECG) beat, with fiducial points indicated as $P, ..., U$. (b) One beat of an example photoplethysmography (PPG) signal. The primary fiducial points are indicated: foot, systolic peak, dicrotic notch, and diastolic peak.

characteristic points. In general, these algorithms align several ECG beats by their R peaks and apply different time domain and frequency domain operations to extract feature vectors. Wubbeler et al. [52] have used a simple distance based classifier on the temporal derivative of the ECG signal. Temporal autocorrelation (AC) [53, 54], discrete cosine transform (DCT)[54], wavelets [55], Eigen decomposition [56], and short term Fourier transform [57] are some of the notable strategies that have been applied for feature selection. Odinaka et al. [31] have given a comparative review for various feature selection strategies and classification methods used in ECG biometrics.

In both cases, the biometric attributes are dictated by the morphology of the ECG signal. To the best of our knowledge, no prior work utilizes an analytical model to study this morphology in the context of biometric identification. We have used a dynamical system to model the ECG signal as a function of time and instantaneous heart rate. The model parameters can be used with a discriminant classifier to identify and authenticate human subjects. We have introduced a new alignment strategy for the ECG signals by using relative angular position from the R peak. This is in contrast to conventional time-based alignment.



(a) (b)

Figure 2.2: A typical ECG signal and its phase-wrapped version in plane. The axes denote the and function value for each angular position in the limit cycle. Values on the axis are the ECG signal.

2.3 Dynamical Model: Pulse Beat Morphology

Figure 2.2 shows one single sinus cycle of the ECG signal. One such cycle or beat can be broadly divided into two phases, known as depolarization and repolarization of the heart muscle fibers. The P wave is caused by the atrial depolarization, whereas the QRS waves are caused by ventricular depolarization. The T and U, on the other hand, are the result of ventricular repolarization. The balance between the sympathetic and parasympathetic nervous system determines the variability of the heart rate. Therefore if we assume the heart as an electromechanical system in a black box, the nervous system governs the actuation signals, determining the heart rate. We replace the black box by the dynamical model proposed by McSharry et al. [44]. We consider the parameters of the dynamical model to be the characteristic features of a person. If the system can reliably estimate these parameters and capture their changing behavior, biometric authentication is possible.

The nonlinear dynamical model proposed by [44] has three coupled differential equations to simulate the ECG signal

$$\dot{x} = \beta x - \omega y \tag{2.1a}$$

$$\dot{y} = \beta y + \omega x \tag{2.1b}$$

$$\dot{z} = -\sum_{i \in \mathcal{F}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0) \tag{2.1c}$$

Intuitively, $\theta = \text{atan2}(y, x)$ represents the angular position along the circular limit cycle in

the (x, y) plane, and z represents the amplitude of the ECG signal. Then $\omega = \dot{\theta}$ is the angular velocity, with $\beta = 1 - \sqrt{x^2 + y^2}$. The third dimension when plotted against time gives the quasi-periodic ECG signal. Depending on the relative angular position in the limit cycle, the dynamics of z decides its distance from the attractor in the $x - y$ plane. [24] represents each of the $PQRSTU$ complexes as Gaussian functions with amplitude a_i, mean θ_i, and standard deviation b_i, here, $i \in \mathcal{F} = \{P, Q, R, S, T, U\}$. The value z_0 is the contribution from the baseline low frequency signal governed by the respiratory sinus arrhythmia (RSA). A typical ECG signal and its phase-wrapped trace in the 3D model space are shown in Figure 2.2b. It is often argued that the change in heart rate has influences on the amplitude of the R-peak. In the current scenario we have normalized the amplitude of a single ECG cycle for simplicity and generalization.

Significantly, if the baseline term $(z - z_0)$ is ignored for beat-to-beat analysis, the solution of the third equation (3.1c) reduces to a sum of Gaussians, one per fiducial point in the set $\mathcal{F} = \{P, Q, R, S, T, U\}$, and $\Delta\theta_i = (\theta - \theta_i) \mod 2\pi$. The parameters a_i, b_i, and θ_i uniquely specify the Gaussian functions, and are determined in the implementation through non-linear least squares optimization. The system registers beats from ECG signals, and normalizes the magnitudes of the R peaks.

2.4 Preprocessing

The ECG signal is often contaminated with high frequency and low frequency noise. The low frequency noise is mainly incorporated into the baseline electrical potential (< 1 Hz). The high frequency component of ECG signal is affected by the power line frequency (60 Hz) and measurement noise from the signal lead (> 40 Hz). Using a standard filtering strategy from [54], we have employed a notch filter of 60 Hz to eliminate the power-line frequency, and a 4^{th} order band-pass filter with passband of 1 Hz to 40 Hz.

2.5 ECG Alignment by Angular Position

After the raw ECG signal has been filtered for noise, we have used a peak detection algorithm to identify the R peaks. According to the dynamical model, each ECG beat traces the unit limit cycle once. Therefore, starting from one R peak, the signal traverses 2π radians in the $x - y$ plane to reach the next R peak. In all previous research the alignment has been

performed in the time domain by aligning only the R peaks from each single beat. Alignment for the rest of the signal is assumed to be automatic. In contrast, we have aligned each ECG signal by its relative angular position (θ) from the R peak. The angular positions 0 to 2π (alternatively, $-\pi$ to π) are quantized into 360 divisions. This method aligns all the data points of an ECG signal to a fixed framework, and reduces the average variance in the signal by 23%. Finally, we normalize the peak to peak amplitude for each beat to unit height (Q-R in Figure 2.2).

2.6 Parameter Estimation of ECG Complex

The dynamic equations in (2.1) can be transformed into polar coordinates [58]:

$$\dot{r} = r\left(1 - r\right) \tag{2.2a}$$

$$\dot{\theta} = \omega \tag{2.2b}$$

$$\dot{z} = \sum_{i \in \mathcal{F}} -a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - \left(z - z_0\right) \tag{2.2c}$$

This transformation makes the first two equations obvious and the dynamics of r redundant to current analysis. Also, for the problem of interest, we align an ECG signal by individual beats. Hence, we can ignore the baseline term $(z - z_0)$. This leaves the dynamics of z as only a simple derivative of a sum of Gaussians, and it is possible to get an analytical solution of z from

$$z = \sum_{i \in \mathcal{F}} \alpha_i \exp\left(-\frac{\left(\theta - \theta_i\right)^2}{2b_i^2}\right) \tag{2.3}$$

where $\alpha_i = f\left(a_i, b_i, \omega\right)$ is the amplitude of the Gaussian function. Figure 2.3 shows how different Gaussian function at different fiducial points constitute the beat morphology of the ECG signal.

For a given ECG cycle with known angular position, $\theta_q = \omega t_q$, z can be assumed as a direct function of θ and it is possible to estimate the model parameters $\boldsymbol{\alpha} = \{\alpha_i\}, \boldsymbol{b} = \{b_i\}, \boldsymbol{\Theta} = \{\theta_i\}, i \in \mathcal{F}$ by formulating an optimization problem that minimizes the squared

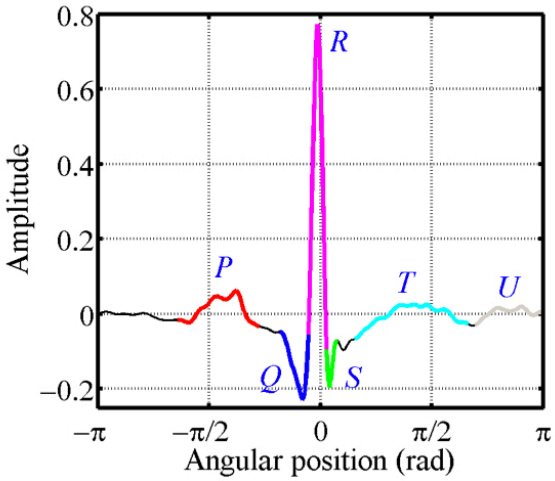


Figure 2.3: The figure schematically shows how each of the Gaussian signals covers part of the ECG morphology.

error between each observation $s(\theta_q)$ and model output $z(\theta_q)$. θ_q is the any quantized angular position between 0 and 360.

$$\{\boldsymbol{\alpha}, \boldsymbol{b}, \boldsymbol{\Theta}\} = \underset{\alpha_i,\ b_i,\ \theta_i}{\operatorname{argmin}} \sum_{\theta_q=1}^{360} \|s\left(\theta_q\right) - z(\theta_q)\|_2^2 \tag{2.4}$$

We have used the Levenberg-Marquardt algorithm for a non-linear least squares optimization, as implemented in Matlab. The initial estimates of the parameters for the normalized ECG signal are shown in Table 2.1.

Before computing the parameters we prepare an ECG template. Each template comprises n consecutive heart beats. The ECG signal for n beats are aligned by their relative angular position from the R peak. Figure 2.4 shows one such template for $n = 30$. Clearly, the template has a variation for each of the quantized angular positions. To include the variation in the signal, we perform the parameter estimation in two steps. First we calculate the mean signal as shown in Figure 2.4, and estimate parameters for the mean signal through the minimization problem in (2.4). Figure 2.5 shows the mean signal of a template and its reconstructed version after the parameter estimation. The template division and alignment are shown in Figure 2.6.

In the second part, we use this initial estimation and formulate another minimization

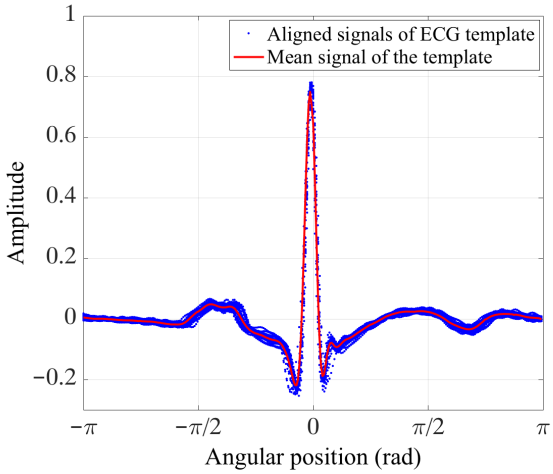


Figure 2.4: All the ECG signals are shown for the thirty beats in a template and their statistical mean for each angular position.

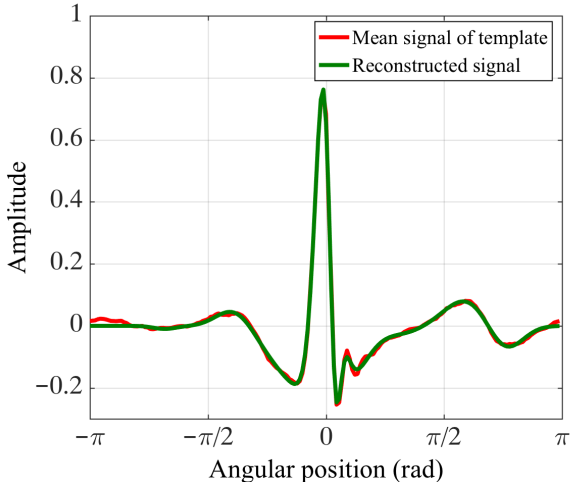


Figure 2.5: The mean signal of the template and its reconstructed version after parameter estimation.

Table 2.1: Initial estimate of the parameters for ECG fiducial points

Index	*P*	*Q*	*R*	*S*	*T*	*U*
α_i	0.12	-0.25	1.2	-0.25	0.12	0.12
b_i	$\pi/16$	$\pi/16$	$\pi/16$	$\pi/16$	$\pi/16$	$\pi/16$
θ_i	$-\pi/2$	$-\pi/16$	0	$\pi/16$	$\pi/3$	$2\pi/3$

problem to include the variation between each beat in a given template by

$$\left\{\boldsymbol{\alpha}', \boldsymbol{b}', \boldsymbol{\Theta}'\right\} = \underset{\alpha_i,\ b_i,\ \theta_i}{\operatorname{argmin}} \sum_{p=1}^{n} \sum_{\theta_q} \left\| s_p\left(\theta_q\right) - z(\theta_q) \right\|_2^2 \tag{2.5}$$

$s_p\left(\theta_q\right)$ denotes the signal value of p^{th} signal in the template at angular position θ_q. The second step helps the system to adapt to the variation in heart rate. This can also be thought of an approach to minimize the modeling noise. It is to be noted that a large value n of will lead towards a conservative algorithm, and requiring a large number of samples to authenticate.

2.7 Quadratic Discriminant Classification

We have tested with three classifiers: linear discriminant classifier, k-nearest neighbor and quadratic discriminant classifier (QDC). After performing 10-fold cross-validation for all the three classifiers, we observed the best result for QDC. Formally, for any given set of observation pairs $(x_i,\ y_i)$, where $x_i \in \mathbb{R}^n$ is the feature vector for the i^{th} example and $y_i \in \{1,\ \ldots, K\}$ is the i^{th} class label, we assume that the feature vectors have a multivariate Gaussian distribution. The QDC trains to find the distinct mean and variance for each of the classes in the n dimension feature space. For us, $n = 18$, $K = 47$. For a uniform prior distribution of the classes, we test our observation to minimize the misclassification cost:

$$\hat{y} = \underset{y=1,\ldots,K}{\operatorname{argmin}} \sum_{k=1}^{K} P\left(k \mid x\right) C\left(y \mid k\right) \tag{2.6}$$

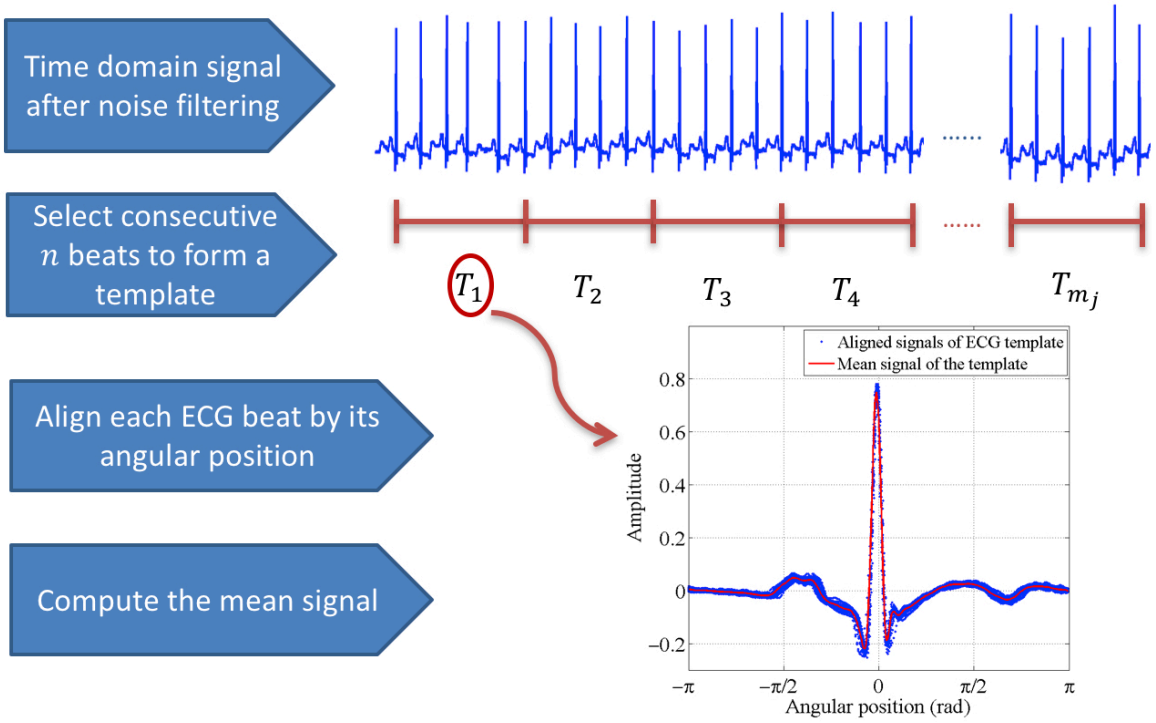


Figure 2.6: A sequence of ECG signal is first divided into a number of beats. Those beats constitute a template. We first align them in angular domain and then find out the mean signal.

where $\hat{y}$ is the predicted class, the posterior probability $P(k \mid x)$ is defined by the prior probability (uniform in this case) and the likelihood given by a multivariate Gaussian density function:

$$P(x \mid k) = \frac{1}{(2\pi |\Sigma_k|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right) \tag{2.7}$$

and the cost function $C(y \mid k)$ by an indicator function,

$$C(y|k) = \begin{cases} 1 & \text{if} \quad y_k = k \\ 0 & \text{if} \quad y_k \neq k \end{cases} \tag{2.8}$$

The terms μ_k and Σ_k are the learnt multivariate mean vectors and multivariate co-variance for class k.

2.8 Experiments and Results

2.8.1 Dataset

We have used the MIT-BIH Arrhythmia database [5] to test the authentication process. This dataset is freely available from physionet.org [59]. It comprises records from 47 participants (25 men and 22 women with age between 23 and 89 years), each with an average of 30 minutes using a single lead ECG signal, digitized at 360 Hz. All the records are filtered using the methods described in section 2.4. We have divided each signal into templates of n beats and estimated the parameters that best match the template. The value of n is selected from:

$$n \in \{1, 5, 10, \ 20, 30\} \tag{2.9}$$

We test the algorithm for different numbers of training and testing beats to record its accuracy for identification and authentication. First, for any given subject j, we divide the full ECG signal into m_j templates, $\mathcal{T}_n^j(k)$, , each comprising n beats and $k = 1..m_j$. After preprocessing and alignment, we estimate the 18 parameters from each of the template using (2.3), (2.4), and (2.5). So, for a given subject we have m_j observations and the feature vector

for the j^{th} subject is $\mathcal{F}_n^j \in \mathbb{R}^{(m_j \times 18)}$. Next, we take an equal number of samples, N_{train}, from each subject to train the classifier and an equal number of test samples, N_{test}, all randomly chosen from each subject. The training and testing sets are disjoint with no overlap between them.

$$N_{train} + N_{test} < \min_j (m_j) \tag{2.10}$$

For each $\{N_{train},\ N_{test}\}$ we run the test for 45 times and take an average over all the results. The number 45 is chosen as an arbitrary number close to the number of subjects, 47.

Authentication Results. The authentication procedure takes N_{train} templates from each subject and classifies each template individually. The final class is chosen by voting from each of the templates. The probe templates are assigned to the class that observes the largest number of votes. We check if individual classes are correctly identified for the given template set from each class. The authentication accuracy is given by:

$$\eta_{Auth} = \frac{Correctly\ identifed\ subjects}{Total\ number\ of\ subjects} \tag{2.11}$$

Figure 2.7 shows results for mean authentication accuracy values using different template sizes. Parts (a) – (e) show accuracy for different numbers of training samples (varying from 25 to 1000) and testing sample (varying from 1 to 500 beats). Although we have used templates of different sizes (n), we have represented all results in number of heart beats ($number\ of\ template \times template\ size$) for illustration. It has been noticed that with the increase of number of training samples, the accuracy also increases, but eventually saturates. We observed that, templates of size $n = 5$ outperforms all other templates for most of the training and testing combination. Figure 2.5(f) shows a performance comparison for a fixed training size of 625 beats. With N_{Test} more than 300, we can guaranty accuracy of $(97 \pm 2.3)\,\%$, where 97% is the mean accuracy and 2.3% is the maximum deviation. For $N_{Test} = 50$, we observed an accuracy of $(90 \pm 2.1)\,\%$.

Identification Results. The initial part of the identification test is performed in the same way as the authentication process. We vary the number of training samples and testing samples for each subject and run the classifier. From the posterior probability matrix of the

Table 2.2: Comparison of authentication accuracy with other studies. F/NF indicates if the method is fiducial based (F) or non-fiducial based (NF).

Study	***Sample size***	***F/NF***	η_{Auth}
Wang et al. [54]	13	F	100%
Wang et al. [54]	13	NF	98%
Odinaka et al. [57]	269	NF	99%
Israel et al. [17]	29	F	97-98%
Irvine et al. [56]	104	F	91%
Ours	47	F	97% $\pm$ 2.3%

classification, we compute the confusion matrix for each rank of recognition. This shows the number of testing samples correctly classified with an accuracy measurement:

$$\eta_{Iden} = \frac{Correctly\ identifed\ testing\ samples}{Total\ number\ of\ test\ samples} \tag{2.12}$$

It was observed that for a fixed number of training samples, identification accuracy does not change much with the change in the number of testing samples. This invariance demonstrates the stability of the algorithm. Figure 2.9a shows one such case for 200 training templates of size $n = 5$ (1000 beats) taken from each subject. The identification accuracy does not change appreciably up to rank 15 when we change our test template population from 5 beats to 500 beats. On the other hand, when we vary the size of the training set but fix the number of N_{Test}, we observe some increase in rank accuracy. When we used a fixed $N_{test} = 5$ beats and vary the training set size from 250 – 1000 beats using a 5 beat template, we can achieve rank 1 accuracy of 65% and rank 15 accuracy of 95% (Figure 2.9b).

2.8.2 Conclusions

The authentication accuracy has been often referred as subject recognition rate or identification performance, whereas the identification accuracy is referred as window matching accuracy (same as our template) or heartbeat recognition rate in different literatures. Table

Table 2.3: A Comparison on feature space dimension.

Methods	*Initial feature space (Final)*
Irvine et al.[50]	10 - 250
Odinaka et al.[57]	2048
Wang et al. [54]	300
Raj et al. [53]	300 (10)
Ours	18

2.2 shows a comparison with the state of the art algorithm for authentication accuracy for within session ECG records, and our algorithm produces comparable results. We have used an 18 dimensional feature space for the QDC. This is comparatively lower than most of the recent works. Table 2.3 compares the dimensionality of the feature space used in literature.

We have described a new approach for recognizing individuals using ECG signals. Although the potential for ECG-based biometric identification has been discussed for roughly 15 years, our algorithm is the first (to our knowledge) to employ an analytical approach to the problem. We have described a novel biometric algorithm in which a dynamical system has been used to model ECG signals. We have introduced an angular position based ECG alignment strategy in contrast to the time based alignments used previously. Our method can be considered to be closer to the fiducial based methods in which the feature parameters of the six fiducial points are learned using a sum-of-Gaussians representation, resulting in an 18-component feature vector that is used for classification.

Using a classifier based on quadratic discriminant analysis (QDA), and using a heartbeat template, our system achieved an accuracy level for human subject authentication of 97%. This method is comparable with state of the art as shown in Table 2.2. The analytical model [24, 36] has proved effective in noise elimination using an extended Kalman filter or unscented Kalman filter. Future work will explore the possibility of better noise elimination, as well as performance of our technique when using ECG signals from different sessions, and consideration of heart rate variability.

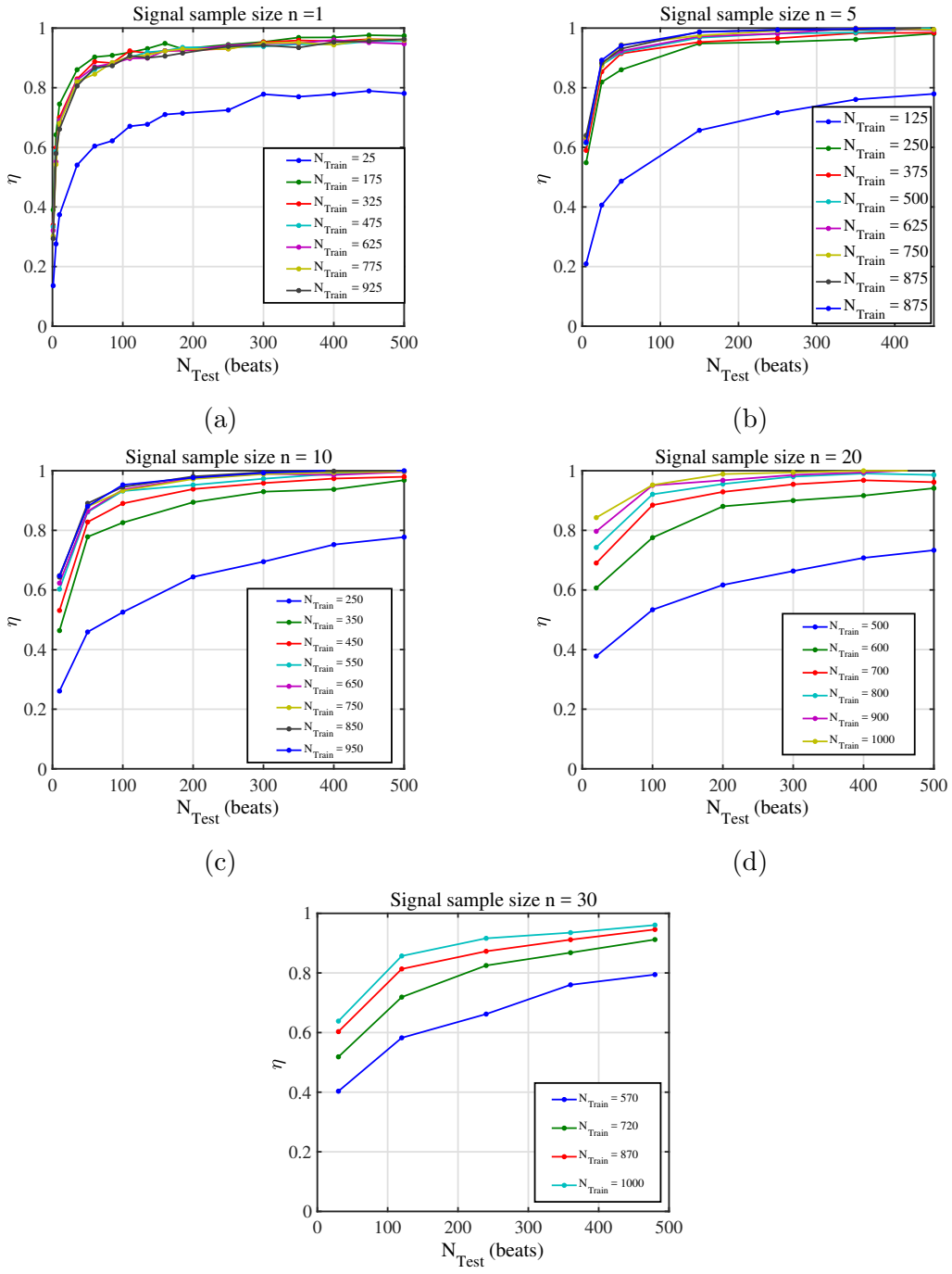


Figure 2.7: Authentication accuracy η shown with different numbers of heartbeats in training and testing. Feature vectors are extracted from a batch of (a) $n = 1$, (b) $n = 5$, (c) $n = 10$, (d) $n = 20$, (e) $n = 30$, heartbeats respectively.

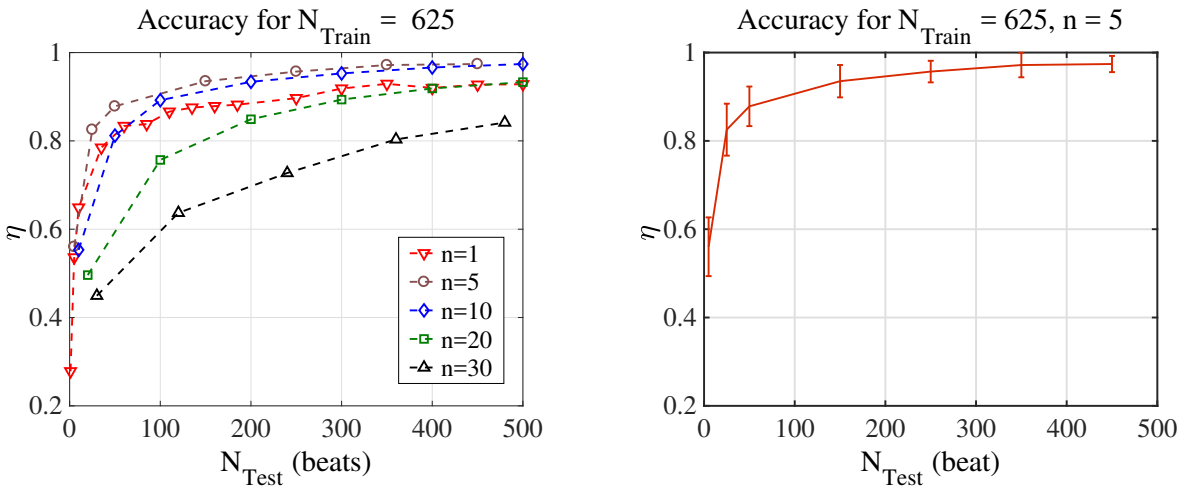


Figure 2.8: A comparison of all the five templates when training is performed on 625 heartbeats per subject. For these tests, a template comprising a 5-beat ECG signal provides the best authentication accuracy

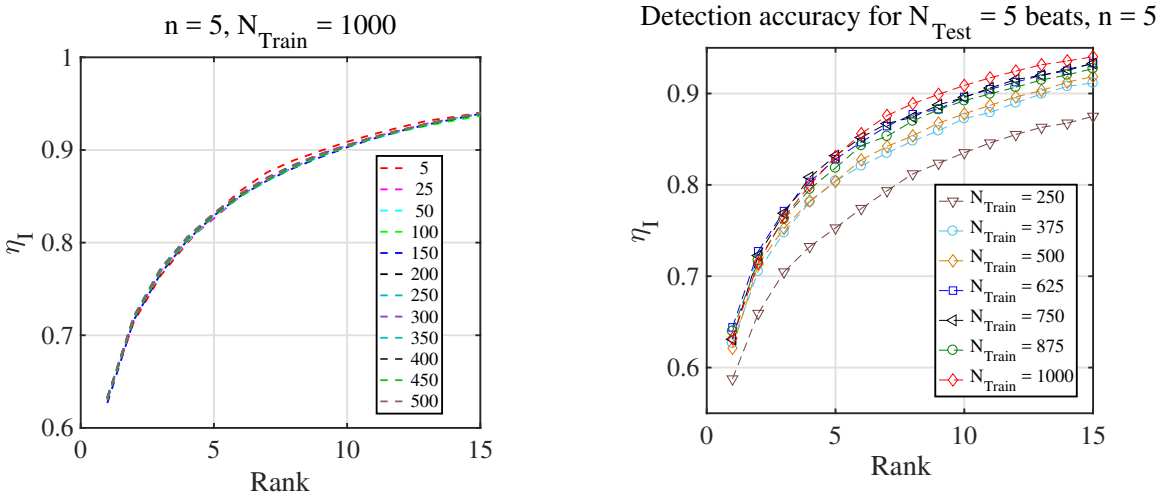


Figure 2.9: (a) Identification accuracy with fixed training sample size, $N_{Train} = 1000$ and variation in N_{Test} from 5 to 500 templates, each with 5 beats. η_I does not change with variation in test sample size. (b) Identification accuracy for a fixed test size of 5 beats and varying training size from 250 to 1000 beats.

Chapter 3

Biometrics Using PPG Signals

3.1 Introduction

We have seen in the previous chapter that ECG signal proves to be an effective biometric modality. Due to the physiological differences of the cardiovascular system from one person to the next [43], other signals representing the behavior of the cardiovascular system hold strong potential for biometric recognition and authentication. This part of the work is concerned with another cardiac sensing modality known as photoplethysmography (PPG). PPG refers to an illumination-based sensing of volumetric changes of blood in the microvascular bed of tissues with every sinus cycle. Measurements can be obtained from the fingertip, toe or ear using commercially available pulse oximeters. Unlike ECG, which measures the electrical activity of the heart, PPG more closely represents the mechanical functioning of the cardiovascular system. A comprehensive analysis of PPG measurement procedures and their possible variations has been presented by Allen [45].

The premise of our work is that it should be possible to distinguish different individuals based on observed PPG signals because of the physiological variations in the cardiovascular system. Due to these variations the shapes morphology of the recorded PPG from different person vary. As an example, Figure 3.1 shows examples of PPG signals from three subjects. Visually, these three sets of signals are distinguishable by their shapes. Following the dynamical model presented in the previous chapter, we have approximated each PPG signal as a sum of Gaussians, and we have used the parameters in a discriminant analysis framework to identify individuals. We have tested our algorithm using the publicly avail-

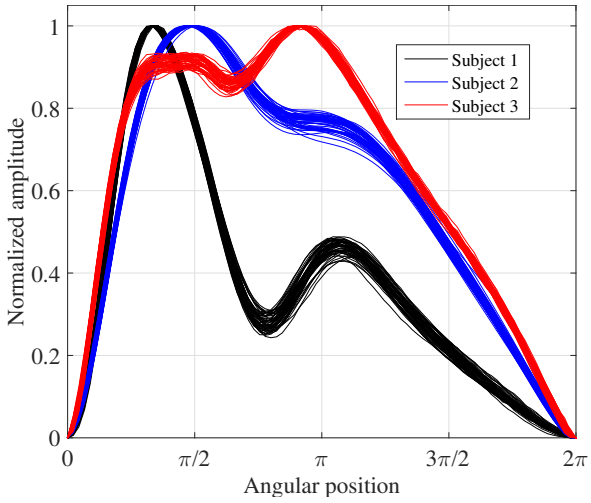


Figure 3.1: Photoplethysmography signals from three different subjects, obtained using a fingertip pulse oximeter. Different individuals can exhibit significant differences in PPG signal morphology, and these differences can be exploited for biometric authentication.

able DEAP dataset [46], which provides PPG signals from individuals in different emotional states under controlled experimental conditions.

3.2 Photoplethysmography Signals

PPG refers to the measurement of change of volume using an optical sensor, generally used for blood volume pulse (BVP) measurement. These signals are often obtained using a pulse oximeter attached to the skin. Although different parts of the body including fingertip, toe tip and ears are accessible sources for PPG, it has been shown that fingertip and ear PPG signals are more prominent and reliable than the toe tip measurements.

As the heart propels blood through the body, fluid dynamics cause small expansions and contractions of the vasculature, which in turn give rise to a pulsating PPG signal. Figure 3.2 shows an example of the signal morphology for a single beat. It comprises two main phases: the anacrotic phase and the catacrotic phase. The anacrotic phase signifies the rising of the systolic pressure and the rising edge of the signal. The catacrotic phase, on the other hand, reflects the diastole and the wave reflection from the periphery. As indicated in the figure, the diastolic peak and the dicrotic notch appear in this phase. The foot of the pulse beat shows the lowest point in the diastole and starting point of the systole. Elgendi [60] has

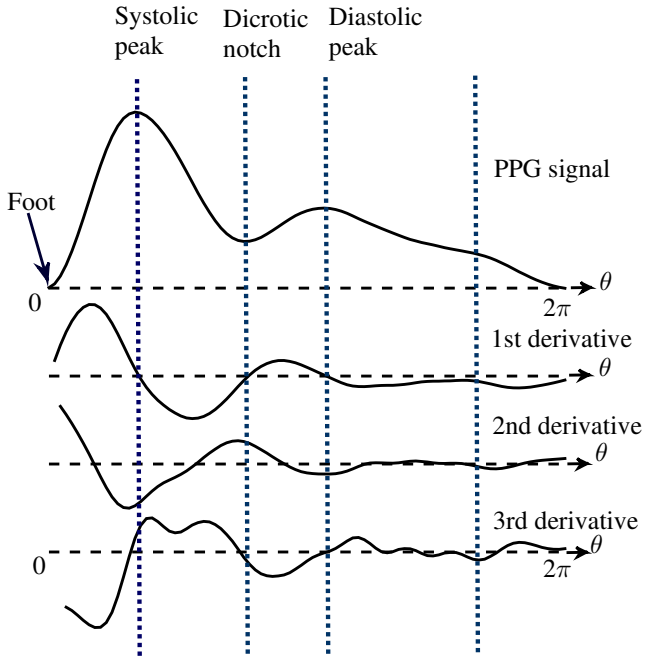


Figure 3.2: Morphology of the PPG signal. The primary fiducial points of the PPG signal are indicated: the foot, the systolic peak, the dicrotic notch, and the diastolic peak. The 1st, 2nd and 3rd temporal derivatives are also shown. Derivatives have been used for time-domain analysis by other researchers. We have used them for initial parameter estimation.

given a comprehensive analysis of the morphology of fingertip PPG and explained reasons for their diversity and variation.

In most of the previous work, researchers have used different time-domain features as biometric descriptors. For example, Gu et al. [61] have used upward and downward slopes, and peak time as features. Yao et al. [62] have used zero crossings of the first and second derivatives of the PPG signal to identify interest points. For most researchers, the major concentration has been on the use of peak locations, relative amplitudes, derivative-based slopes, and time intervals between the interests points [63, 64, 65, 66]. For example, Kavsaoğlu et al. [65] have used 40 such features and they applied a feature ranking algorithm with k-means clustering. Alternatively, Spachos et al. [67] used an Eigen space decomposition of the time-domain signal to obtain a template for identification. To the best of our knowledge, all of the previous efforts have relied heavily on time-domain analysis of individual PPG pulses.

A problem with such an emphasis is that the duration of a complete pulse depends on the instantaneous pulse rate, which is governed by the autonomic nervous system. Therefore, alignment and matching is inherently difficult for any two pulses, even from the same individual, with different instantaneous pulse rates (e.g., before and after exercise). Figure 3.3a presents an illustration of this phenomenon, in which the subject shows large variations in pulse rate. As a result, the duration of each pulse varies enough to introduce large variances into the positions of the dicrotic notch and the diastolic peak. This problem largely disappears when we use a limit cycle and angle-based alignment as described in Section 3.3.2. It has been well established that pulse rate and heart rate variability (HRV) vary significantly when an individual assumes different emotional states [32]. It is our conjecture that any PPG-based biometric system must demonstrate an ability to accommodate a wide range of instantaneous pulse rates.

Limitations in previous approaches have led us to introduce an analytical model that considers instantaneous heart rate while aligning multiple PPG beats. The technique transforms the signal from time domain to angular domain, where each beat starts at angle $\theta = 0$ and ends at $\theta = 2\pi$. Figure 3.3b shows an example of PPG beats that have been aligned in the transformed, angular domain. Clearly, the proposed alignment strategy shows better agreement, and it reduces intra-class variation. To validate these assumptions, we have experimented with a dataset that contains PPG signals of individuals who have experienced different emotional stimuli. These results will be presented in Section 3.4.3.

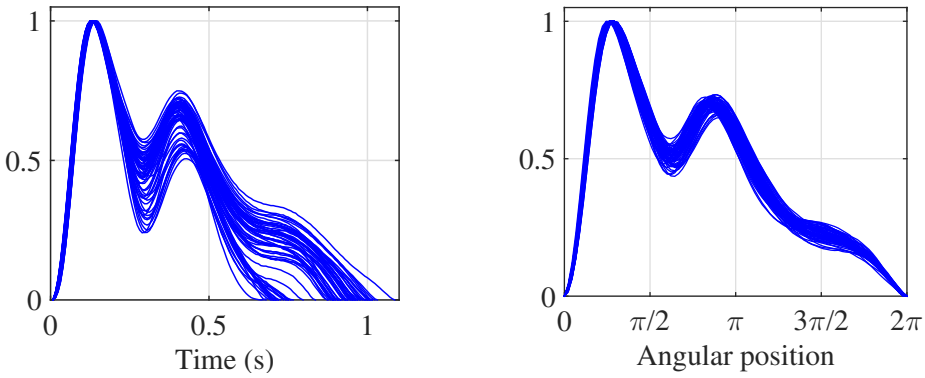


Figure 3.3: Morphology of the PPG signal. The primary fiducial points of the PPG signal are indicated: the foot, the systolic peak, the dicrotic notch, and the diastolic peak. The 1st, 2nd and 3rd temporal derivatives are also shown. Derivatives have been used for time-domain analysis by other researchers. We have used them for initial parameter estimation.

3.3 Feature Extraction

3.3.1 Preprocessing

Data acquisition of PPG signals often results in noise along with measurement artifacts. Low-frequency noise may be introduced into the measurement data by movements of the measuring site (e.g., a finger), and by actions such as coughing, laughter or even deep breaths. In some cases, the noise may affect the shape of the PPG signal. Therefore we have first used a moving-average low-pass filter with a 2-second window. High-frequency noise is eliminated using a 4th-order Butterworth filter.

The next step is to apply a peak detection algorithm to detect the feet of the PPG signals, to identify the starting and ending point of each pulse. We have used a custom built peak detection algorithm that scans every 0.4 seconds for a possible peak. This duration varies with different sampling frequencies. We then use the dynamic model to map each pulse onto a limit cycle. This step maps each pulse from the time domain to the angular domain. A single pulse beat of the PPG signal (one sinus cycle) is considered to traverse a full cycle of the limit cycle. Therefore it starts at 0 and ends at 2π. Finally we normalize each beat to the magnitude range [0, 1].

3.3.2 Dynamical Model

We have used the same dynamic model as Section 2.3 proposed by McSharry et al. [44] for ECG signals. In this section, we have generalized this model to represent any cardiac signal including ECG, ballistocardiography (BCG), and PPG, seismocardiography (SCG):

$$\dot{x} = \beta x - \omega y \tag{3.1a}$$

$$\dot{y} = \beta y + \omega x \tag{3.1b}$$

$$\dot{z} = -\sum_{i\in\mathcal{F}} a_i \Delta\theta_i \exp\left(-\frac{\Delta\theta_i^2}{2b_i^2}\right) - (z - z_0) \tag{3.1c}$$

Here, the dynamics of the third state defines the shape and characteristics of the time domain cardiac signal on a limit cycle traced on x and y plane. Each cardiac cycle (one heartbeat) completes one cycle on the $x - y$ plane with angular velocity ω (instantaneous

hear rate). The dynamics is defined for every instantaneous angular position $\theta = \tan^{-1}(y/x)$, distance from the fiducial points, $\Delta\theta_i = (\theta - \theta_i) \bmod 2\pi$, and $\beta = 1 - \sqrt{x^2 + y^2}$. $\theta_i$is the angular position of the fiducial point, a_i and b_i are other model parameters for the set of different fiducial points, $\mathcal{F}$. θ_i is measured from a reference point on the signal. For PPG it is measured from the starting foot of the signal (Figure 3.2). For ECG, it is measured from the R peak, so that $\theta_R = 0$ (Figure 2.2a).

$$z(\theta) = \sum_{i \in \mathcal{F}} \alpha_i \exp\left(-\frac{(\theta - \theta_i)^2}{2b_i^2}\right) \tag{3.2}$$

The analytical solution reduces $z(t)$ to $z(\theta)$, as a direct function of angular position, $\theta = \omega t$, instantaneous heart rate ω with the model parameters, $\{\alpha_i,\ b_i,\ \theta_i\}$. It is worth noting that the analytical solution is basically a sum of Gaussian functions with means of each Gaussian at θ_i, which are the fiducial point locations such as systolic and diastolic peaks.

3.3.3 Parameter Estimation

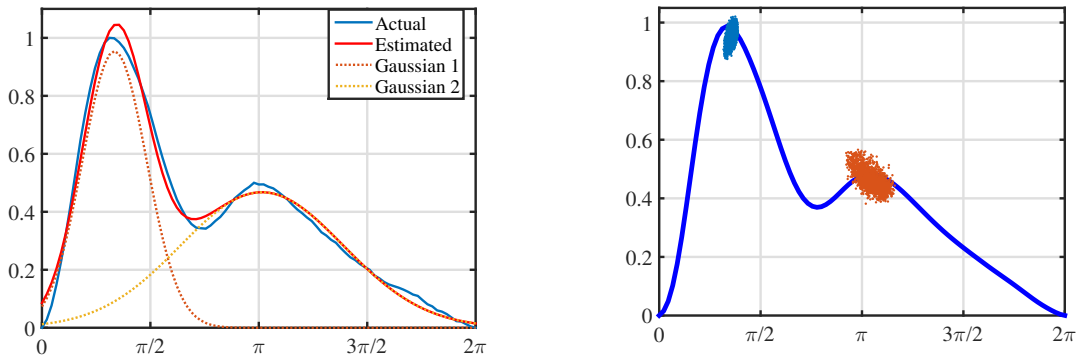
We quantize the angular position $(0 \to 2\pi)$ in equal parts, and formulate an optimization problem that minimizes the squared error between each observation $s(\theta_q)$ and model output $z(\theta_q)$ at instantaneous angular position θ_q :

$$\{\boldsymbol{\alpha}, \boldsymbol{b}, \boldsymbol{\Theta}\} = \underset{\alpha_i,\ b_i,\ \theta_i}{\operatorname{argmin}} \sum_{\theta_q=1}^{360} \|s(\theta_q) - z(\theta_q)\|_2^2 \tag{3.3}$$

The Levenberg-Marquardt algorithm was used to solve the non-linear least-squares optimization problem.

We have used two different models in our PPG experiments. The model differs based on the number of Gaussian functions, and on the choice of fiducial points in the PPG signal. The first model (referred as 2G hereafter) uses only 2 Gaussians to model two fiducial points at the systolic and diastolic peaks. The other model (referred as 5G hereafter) uses 5 Gaussians, and takes three additional fiducial points: the starting foot, the dicrotic notch, and the negative slope in the catacrotic phase. For the optimization process, initial parameter values are selected from the 1^{st}, 2^{nd} and 3^{rd} derivatives $(z'(\theta),\ z''(\theta),\ z'''(\theta))$ of the PPG signal with respect to the angle (Figure 2.2).

Figure 3.4a and Figure 3.7c show the original signal, the estimated signals and estimated Gaussian functions at each fiducial point for 2G and 5G respectively. Figure 3.4b and Figure 3.7d show the location and amplitude of the Gaussians on a set of PPG signals (total 1662 beats) from the same individual. The parameters from the Gaussians create the feature vector $x_b \in \mathbb{R}^{(3n_G)}$ for each pulse beat, where n_G is 2 or 5 depending on the model we choose. In case of 2G, the residual error from the estimated signal is higher than that for the 5G case. Only two Gaussians located at the systolic and diastolic peak often reconstruct the actual PPG signal with high residual error. Figure 3.4a shows the mismatch between the blue and the reconstructed red line. But, when we use five Gaussians in the case of 5G model, the PPG signal was properly estimated as shown in Figure 3.7c. In both models, the estimated parameters are seen to be clustered in close proximity for a particular individual. They show low intra-class variance, which is important for subsequent classification.



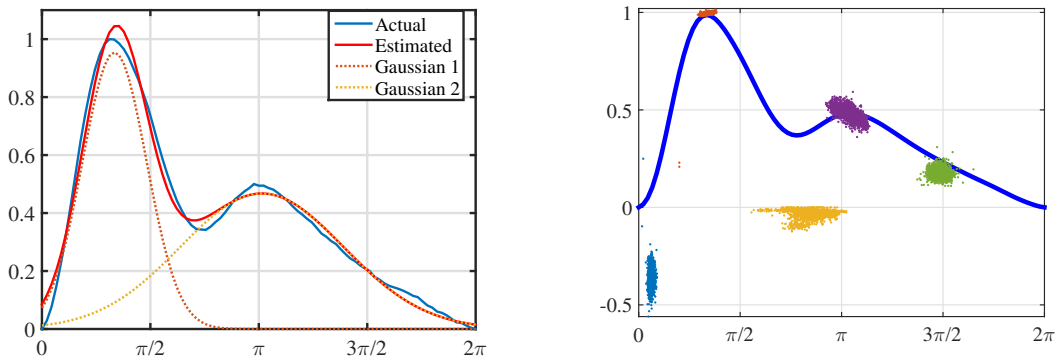
(a) (b)

Figure 3.4: Parameter estimation for the 2G model. (a) Two Gaussians have been used to estimate a single PPG signal, and are centered at the systolic and diastolic peaks. (b) Locations and amplitudes of the estimated Gaussian functions for 1921 PPG beats for a single individual.

3.4 Experiment and Results

3.4.1 Dataset

In the current scope of work, we have used a publicly available PPG dataset, DEAP [46], for our experiments. The dataset contains physiological signals from 32 individuals with



(a) (b)

Figure 3.5: Parameter estimation for the 5G model. (a) Five Gaussian functions have been used to model the signal foot, systolic peak, dicrotic notch, diastolic peak, and knee point for a single PPG signal. The final estimated signal (red) has been plotted on top of the actual signal (blue). (b) Locations and amplitudes of the estimated Gaussian functions for 1921 PPG beats for a single individual.

varying age and different gender. Unfortunately, data from only 23 individuals contains PPG recording that can be used for analysis. Data from each subject is available for 40 sessions, each session with duration of one minute. In each of the 40 sessions, a video based visual stimulus has been used to elicit a particular emotion. The emotional stimuli are well distributed in the arousal-valance map. Therefore it is expected that the acquired PPG signal covers a diverse excitation signal from the autonomic nervous system to generate pulse beats. Sessions for a particular subject were recorded on the same day, with breaks and baseline emotional stimuli between every session. To the best of our knowledge no other publicly available dataset contains high quality, larger and diverse range of PPG signals. The number of available pulse beats for each participant after noise elimination varies from 1500 to 2200.

3.4.2 Biometric Identification

Once the parameter for each pulse beat is estimated, we have used two discriminant classifiers: linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) for training. Discriminant analysis is simple and effective classification method that has been widely used by biometric researchers over the years.

Identification with Single Pulse.

We randomly choose n_{train} number of pulse beats for each subject and train the classifier, which learns a multivariate Gaussian function for each subject sub_k from the knowledge of the feature vector $x \in \mathbb{R}^{(3n_G)}$ estimated from Section 3.3.3:

$$P\left(x \mid sub_k\right) = \frac{1}{\left(2\pi \left|\Sigma_k\right|\right)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}\left(x-\mu_k\right)\right) \tag{3.4}$$

The terms μ_k and Σ_k are the learnt multivariate mean vectors and multivariate co-variance for subject class sub_k. Then we compute the posterior probability $P\left(sub_k|x_b\right)$, for each testing beat b, which quantifies the belief that the pulse beat comes from sub_k. Figure 3.4.2 shows results for LDA and QDA based classification accuracy for 2G and 5G models when we vary number of training samples n_{train}. The results show average classification accuracy of each single beat. We run 40 trials with different combination of training and testing samples and report the mean value. The average size of test samples is 1552. The detection accuracy increases with increasing number of training samples, but saturates after 200 training samples. The training size does not influence the LDA classification compared to its QDA counterpart. With larger training size, the QDA outperforms LDA, but it is interesting to note that with small training size (20 pulse cycles), LDA performs better.

Identification with Larger Template

Next, we evaluate our algorithm where we test n consecutive pulse beats and determine their joint probability. We calculate the posterior probability for each beat similar to the previous section. Then we select n consecutive pulses (constituting a template $\mathcal{T}$) and compute their joint probability

$$P\left(sub_i|\mathcal{T}\right) = \prod_{j=1}^{n} P\left(sub_i|b_j\right), \qquad b_j \in \mathcal{T} \tag{3.5}$$

$$\tag{3.6}$$

Figure 3.4.2 show results from the 2G and 5G models where we keep the training size fixed at 100 pulses (80 seconds) and change the number of consecutive beats. The accuracy

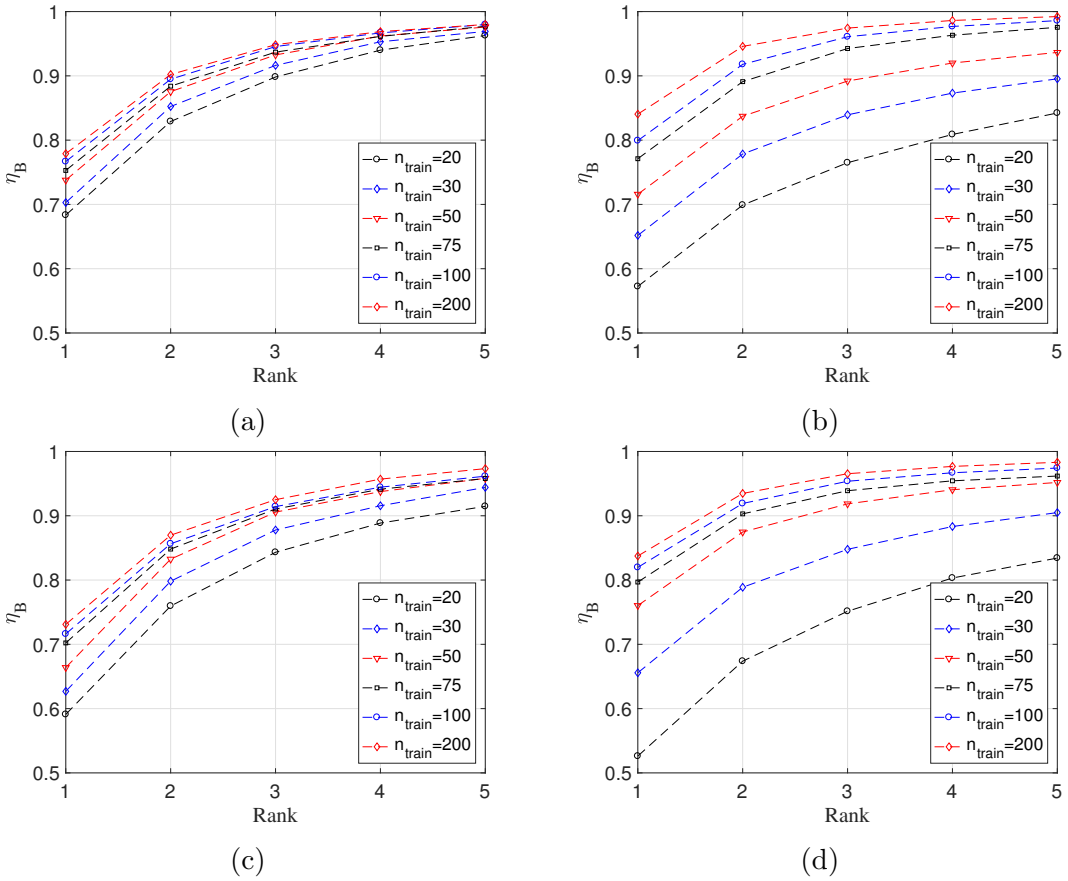


Figure 3.6: Average identification accuracies of individual pulse beats for with different training sample sizes, indicated by . Results for (a) 2G model with LDA, (b) 2G model with QDA, (c) 5G model with LDA, and (d) 5G model with QDA.

increases as we increase the number of test samples in the template. It shows that when training with 100 pulse beats and testing with only 10 pulse beat (8 seconds of testing sample considering average heart rate of a person is 75 BPM), our algorithm can achieve 95% rank-1 accuracy. We can achieve a commendable rank-1 accuracy of 92.5% when using only 5 pulses or 4 seconds of data.

A comparison in Table 3.1 and Table 3.2 show accuracy (in percentage) resulting from the 2G model and 5G models respectively, for two different training sample sizes, 100 beats (1.3 minute duration, Figure 3.4.2) and 75 beats (1 minute duration). In general, the QDA model performs better in most of the scenarios, except that the accuracy is very close with the LDA accuracies, when working with 2G model. When we consider training with 75 beats, the 2G model can achieve a rank-1 accuracy of 90% and rank 2 accuracy 97% while only considering

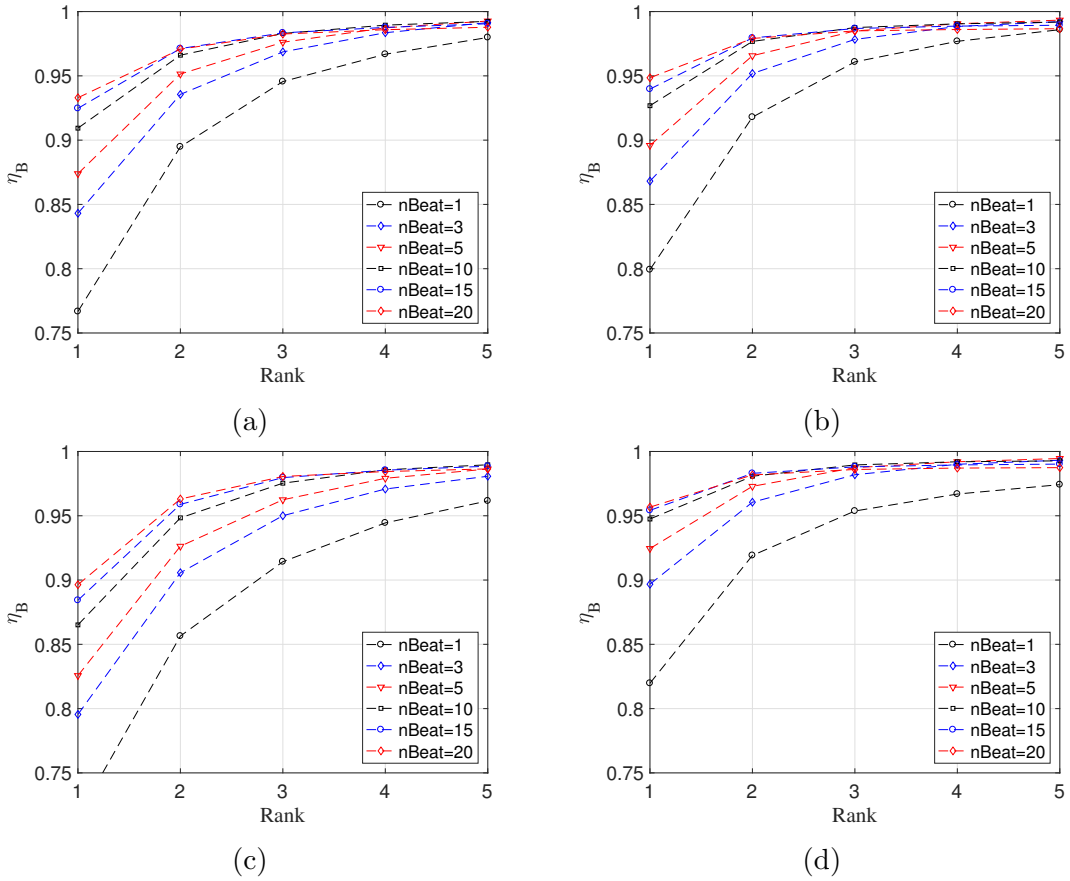


Figure 3.7: Average identification accuracies of pulse beat sequences, when we fix the training sample size at 100 pulse signals and test with n consecutive full cycles of PPG pulse signal. Results are shown for (a) 2G model with LDA, (b) 2G model with QDA, (c) 5G model with LDA, and (d) 5G model with QDA

4 seconds of PPG data. This is in comparison to the 5G model where we can achieve 93% in rank-1 accuracy and 98.5% in rank-2 accuracy.

Table 3.3 shows comparison of state the art methods with our results. We outperform the work of Gu et al. by 6% and the result from Kavsaoğlu et al by 2%. The feature ranking method proposed by Kavsaoğlu et al. uses a 40 dimensional feature space and may take higher learning time when compared to our reduced feature space of six feature and fifteen features. We compare our result with other cardiac signal based methods. The work from Odinaka et al. [68] report 92% rank 1 accuracy while using ECG and LDV together when performing analysis. All other work from ECG biometrics reports rank-1 accuracy of 94% to 99% [47, 56, 57]. But all previous analysis was performed in a single session at rest with

Table 3.1: A comparison if LDA and QDA for 2G model. The results stay almost same for both the classifiers.

	n	**1**	**3**	**5**	**10**	**15**	**20**
	n_{Train}						
LDA	**75**	75.29	82.63	85.95	89.66	91.32	92.31
	100	76.67	84.31	87.39	90.93	92.47	93.3
QDA	**75**	77.11	83.96	**86.95**	**90.53**	**92.12**	**93.32**
	100	79.93	86.81	**89.59**	**92.68**	**93.97**	**94.85**

Table 3.2: A comparison of accuracy from LDA and QDA for 5G model. QDA shows about 10% higher accuracy when compared with LDA in low test-sample size.

	n	**1**	**3**	**5**	**10**	**15**	**20**
	n_{Train}						
LDA	**75**	70.22	78.33	81.95	85.67	87.67	88.9
	100	71.62	79.55	82.57	86.51	88.43	89.64
QDA	**75**	79.69	87.50	**90.49**	**93.22**	**93.91**	**94.18**
	100	81.97	89.68	**92.45**	**94.75**	**95.42**	**95.67**

Table 3.3: Comparison with state of the art PPG based biometric systems.

Study	*No of subject*	*No of features*	η_{Rank1}
Gu et al. [61]	17	4	90%
Kavsaoğlu et al.[65]	30	40	94%
Ours (2G)	23	6	**95%**
Ours (5G)	23	15	**96%**

no external stimuli. Also, acquisition of ECG signal is more challenging that accessing PPG signal. Therefore, our result proves the viability of using PPG as a biometric modality.

3.4.3 Robustness against Emotional Excitation

Finally, we report our results for cross session authentication. It is well known that the behavior and shape of the physiological signal change with various intrinsic and extrinsic factors [60]. One of the major components that often influences the PPG morphology is the sympathetic activation from the autonomic nervous system. In the DEAP dataset, each person was exposed to 40 different emotional stimuli which altered their sympathetic and parasympathetic balance. These 40 emotions cover a large spectrum of the arousal-valance map which shows significant differences in the heart rate variability.

To investigate the robustness of our system, we took PPG data from only one session (say, session 1 for each participant) and train our classifier. Then we took test data sample from a different session and check the accuracy. For each training case, we tested signals from the rest of the 39 sessions. We repeated this 40 times, where in each case the classifier (QDA) was trained with data from session i, $i \in [1, 40]$, for all participants and test with other 39 sessions. Our results show that (Table 3.4) we achieve an average accuracy of 90% while testing with 20 beats. This is in consideration that for session based analysis each session comprises different number of training samples (sometimes as low as 45 beats average) due to variation of pulse rate and motion artifacts. We believe that larger training instances from each session will boost our result.

Table 3.4: A comparison if LDA and QDA for 2G model. The results stay almost same for both the classifiers.

n	**1**	**3**	**5**	**10**	**15**	**20**
	73.12	82.77	84.29	87.64	89.08	**90.53**

3.4.4 Conclusion

We have introduced a new technique for biometric authentication that utilizes photoplethysmography signals. The approach is based on a dynamical model that maps each cardiac cycle to a limit cycle, transforming the time-domain PPG signal into an angular domain. This representation aids in the registration of successive PPG pulses. To the best of our knowledge, this work is the first instance where instantaneous heart rate has been considered as a latent variable, thus including its effect on the shape of the PPG signal.

The model facilitates the decomposition of the PPG signal to a sum of Gaussian functions. We have considered cases of 5 Gaussians and 2 Gaussians separately for estimation. They both may be considered as a fiducial approach, as the resulting Gaussians are aligned with distinctive interest points of the PPG pulse, including the systolic peak, the diastolic peak, the dicrotic notch, and the pulse foot. The Gaussian parameters serve as a biometric feature template that can be used for recognition or authentication. Using 2 different discriminants classifiers (based on LDA and QDA), experimental results have demonstrated rank-1 accuracies of 90% with 2 seconds of PPG test signal data, and 95% using 8 seconds of test signal data.

Although a few researchers have considered biometrics modalities based on activity of the cardiovascular system, to our knowledge this work shows commendable results while using a test signal as small as 4 seconds. We have further shown that our algorithm is robust under different emotional excitations. This is particularly important because emotional stimuli often changes the behavior of the physiological signals and it is very unlikely to consider that a person will have the same emotional state that of training in all of the subsequent real life applications. Therefore, the proposed technique is fast, robust and holds potential for use with a variety of wearable sensors, with capability for continuous, noninvasive biometric authentication.

Chapter 4

Remote Plethysmography in Naturalistic Scenarios

4.1 Introduction

This chapter discusses challenges and promises of noninvasive measurement of HR and HRV from face video. As discussed in Section 1.1.2 and Section 1.1.3, these two parameters are important in assessments involving cognitive workload, stress level, overtraining, drowsiness, and fatigue. Although several instruments are available to measure these variables, to the best of our knowledge, no mechanism currently exists that is capable of non-intrusive measurement under natural (non-laboratory) conditions. Video monitoring offers a potential means for measuring BVP from a stream of video data. Every cardiovascular pulse wave sends a volume of blood through the body. This blood flow also causes slight changes in the color of the skin. These changes typically are not perceptible to the normal vision. Poh et al. [3], McDuff et al. [69] and Wu [70] have shown that with the help from advanced computer vision, heart rate information from this pulsating signal is possible to recover. Our work is mostly inspired by the work of Wu [70], who have used Eulerian video magnification to extract heart rate information. Most of the recent literatures concerning video based plethysmography have reported results from indoor experiments(e.g. Lam and Kuno [71], Li et al. [72]., [38]). In practice, the outdoor scenario poses additional challenges compared to controlled experiment in the laboratory. Again, most of them except [69] have reported average heart rate measurement accuracy. But, instantaneous heart rate plays more important role for critical applications in affective computing and medical applications [32]. Therefore,

this work is devoted to implement a suitable algorithm which can tackle practical problems posed by unconstrained natural environment. We have discussed each challenge step by step to adapt to remote measurement of HR. We have evaluated the performance of our algorithm with indoor and outdoor experiments to check:

- How accurately we can measure instantaneous HR?
- Can we preserve HRV and assess psychophysiological condition in naturalistic scenario (driving)?

4.2 Eulerian Video Magnification (VidMag)

Wu [70] has introduced a video magnification algorithm that combines spatial and temporal processing of videos. First, each frame is down-sampled using a Gaussian pyramid. The down-sampling reduces spatial noise in each frame. Then each pyramid level is temporally filtered using a band-pass filter to extract the frequency of interest. Finally, each pyramid level is concatenated with gain factor. The gain factors define the relative weight of the frequency component from each pyramid level. This helps to magnify the color for visualization. Figure 4.2 shows an example from [4] how the signal can be magnified.

In the present application, we use only one level of pyramid for spatial decomposition and use a adaptive bandpass filter. The passband is generally between $0.66Hz$ to $3Hz$ for normal person. This refers to 40 to 180 beats per minute of HR value. We amplify the output signal and analyze for heart rate.

4.3 Implementation and Testing of VidMag

The main process is divided in three major steps: pre-processing, video magnification, and post-processing. The following sections discusses the methods applied for each of these steps and is followed by detailed discussion of each method in the subsequent sections.

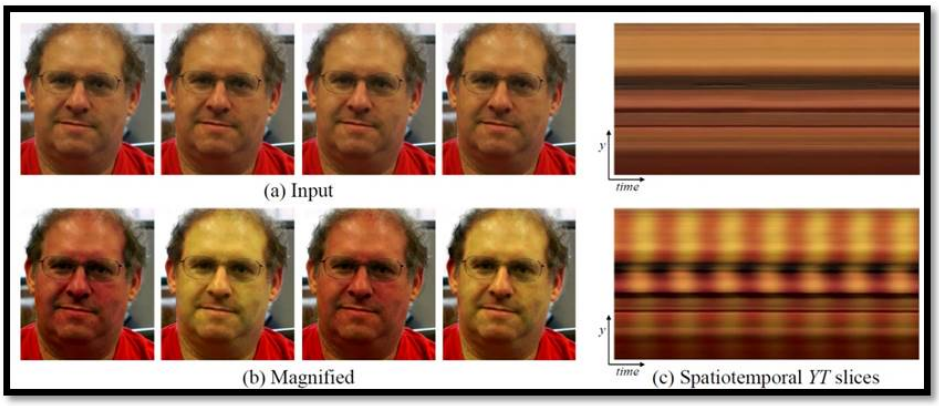


Figure 4.1: Eulerian Video magnification (image source - [4])

4.3.1 Preprocessing

VidMag needs stable video sequence to maximize performance of detecting the pulse. Therefore the videos need several preprocessing steps before they are fed to the VidMag algorithm. First, we apply a face detector to identify the face in the video. The face detector localizes the position of the face in each video frame. Standard face detectors, available in the literature, can also detect the position of fiducial points like eyes, mouth, nose and eyebrow. These features help us track the 3D orientation of the face in the 2D image coordinates. During normal conditions, a person moves his/her face in all three translational and three rotational directions. Therefore, a single point on face is not stable as it moves within the 2D image and may even disappear from view at times. Hence, we need to stabilize the face in order to track each pixel over time. This process of understanding the face orientation and tracking the areas of interest is known as image registration. We automatically select a reference frame in which all other frames are registered. This calibration process minimizes noise that would otherwise result from both voluntary and involuntary face movement. Now that the faces over different frames are registered to their reference location, the video is ready for the VidMag algorithm and the resulting signal magnification. Detailed parameter settings for VidMag like filter parameter, gain control for the magnification are discussed later in this chapter.

4.3.2 Post- processing

All parts of the face do not contain the information from which the pulse rate may be extracted. Only the skin provides such information. Features such as the eyes and facial hair introduce undesirable noise to our assessment. Therefore, before extraction of the pulse rate from the magnified videos, we implemented an automated skin detection algorithm. An effective algorithm was not identified within the existing body of work such that we developed a novel skin detection algorithm which can be applied to grayscale skin pixel localization. The generalized skin detection algorithm expand the scope of our algorithm to different types of images as well. We read pulse information only from the known locations of the skin pixels. Once we extract the temporal pulse signal from the skin patches, we implement a custom peak detection algorithm to find peaks. Then we use Savitzky-Golay filter, specialized for filtering non-uniformly sampled data to filter the heart rate data. Finally, these heart beat sequence can be analyzed in time domain or frequency domain to determine the physiological characteristics for drivers. Lomb periodogram is one of the most commonly used frequency analysis tool that helps understand the psychophysiological condition. In the following sections we shall discuss all the steps in more details.

4.4 Face Detection and Registration

The results reported for the [70] have considered heart rate measurements under restricted conditions, but naturalistic situations are very different from laboratory settings. VidMag tracks the changes of a pixel over time at a fixed location of a persons face (The face is well suited for analysis, because it typically provides a consistent region with exposed skin). In mobile settings, the subject frequently changes the position and orientation of the head. Any kind of natural vibration also adds to the random motion which is not well tolerated by VidMag. In we consider standard naturalistic driving videos for example, the drivers need to look at the mirrors and scan the surrounding to actively monitor the road and traffic conditions. Substantial head movement is also natural during specific maneuvers, such as lane changing action, turning in the intersections or interacting with passengers. The frequencies of such head movement of the driver are also largely governed by the traffic condition, the road condition, and individual driving patterns. For example, while driving in a city, the driver needs to scan the road, other vehicles and pedestrians more frequently than while driving in an interstate or freeway. In comparatively stable driving scenarios,

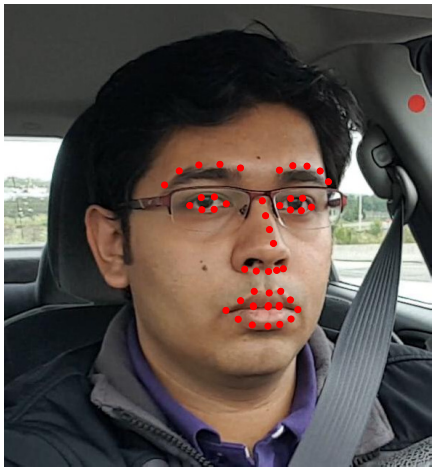


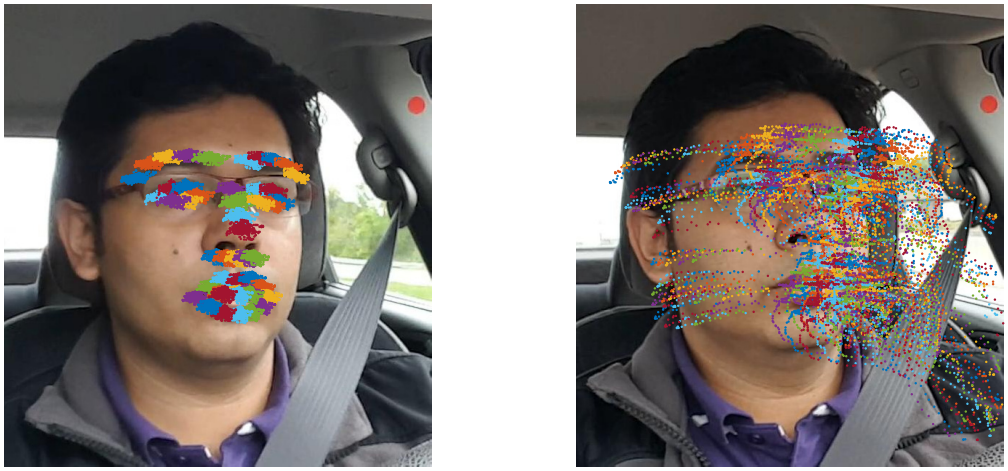
Figure 4.2: Face detection for the driver face video. It also detects 49 fiducial points as shown using red dots.

the drivers tend to look towards forward and often use eye movements to scan the roadway. These cases are particularly favorable for our application. Thus, we have implemented a face detector and tracker to identify scenarios where the face is comparatively stable and the driver is looking forward. In the next section we discuss how a state of the art face detector was leveraged and extended to this problem space.

4.4.1 Frame Selection

We have extracted landmark points from each frame using the face detectors from Zhu et al. [73]. Figure 4.2 shows 49 landmark points extracted for the face. This includes eyebrows, eyes including eyeballs, nose-bridge, lips and mouth. The VidMag tracks changes of color (intensities) of a fixed point on the face over time. Therefore we need to identify the same location on the face and track its changes. Because of the fact that drivers often change their head position and orientation with respect to the camera, tracking a single point needs exact registration of faces between frames. Linear methods like homography shows very good results for small movements but the error increases as the rotation angle increases. This is partly due to interpolation of the missing points and partly due to loss of weak perspective.

Therefore, we chose to find a sequence of frames where the head movement is limited.



(a) (b)

Figure 4.3: We show two sequence of video frames and plot the movement of the points. (a) shows a case when the face was considerably stable, and (b) shows spread of points during larger movement of face, in this case turning of head.

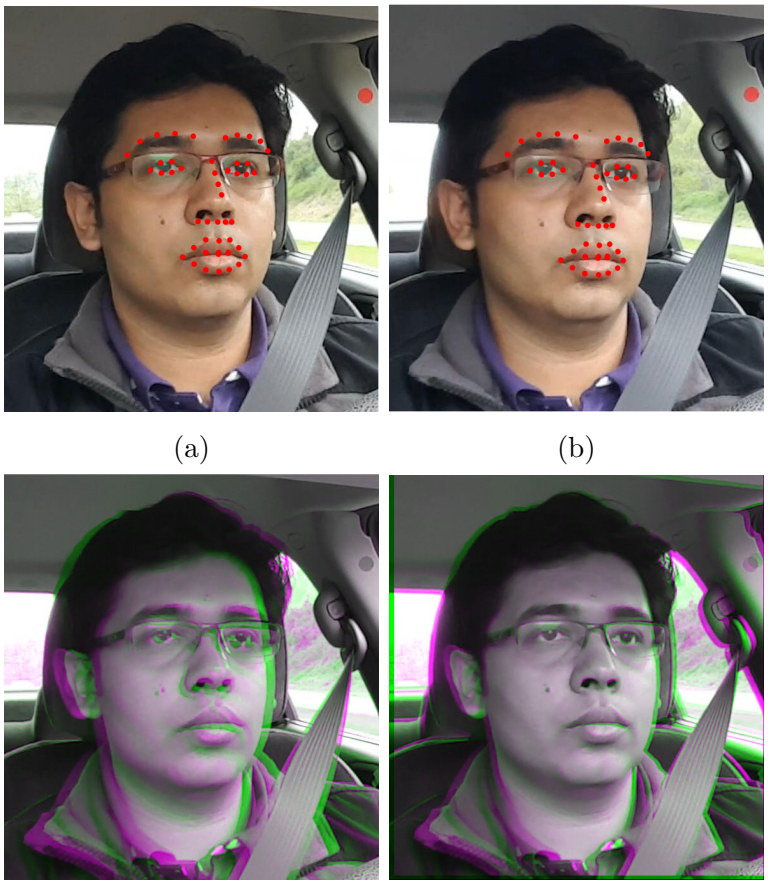
This limitation does not affect the current application scenario especially because drivers tend to look forward for most of times (roughly greater than 90% of the driving time). This gives us an option to estimate the heart rate without losing most of the frames due to head movement. In order to recognize a sequence of frames with limited head movement, we track the fiducial points for consecutive frames. We calculate the average shift in the Euclidean distance for the 49 points over time. If that value is within a limit, we accept that as an allowable sequence of frames, otherwise we discard that sequence. In this case, we have selected such frame sequences where mean-square shift of the landmark points are limited to 17 pixels, for an average face dimension 80×80 pixels. Figure 4.3 plots the fiducial points for two sequences of frames. It shows the spread of the points over time. The face images are for illustration purpose. Figure 4.3(a) shows a compact plot of the points. We qualify this as potential frame sequence to apply VidMag. The points in Figure 4.3a have very high scatter and we discard these frames.

4.4.2 Face Registration

Natural vibrations of the face can be reduced by image registration. Image registration helps to align two or more images containing the same scene or object. The appearances of the objects often differ depending on their orientation and scale. Image registration method estimates the linear translation, rotation of an object or scene from one image to the next. In our case, we align faces from consecutive frames using their fiducial points. In many of these cases, registration by translation only may suffice the purpose. This reduces the complexity of image registration. As an example, we have extracted landmark points from consecutive frames and have registered them by translation [74] . Figure 4.4 shows registration of two randomly chosen frames from a video sequence. Figure 4.4a and 4.4b show the frames with their corresponding landmark points. Figure 4.4c shows two superimposed frames before the registration process and Figure 4.4d shows the results after linear sub-pixel registration [74]. The purple and green colors indicate registration mismatches from the two frames. In a more challenging video sequence where the user is very active, the complexity of face registration increases. We can use direct homography for registration or we may first register faces by linear shift and then use homography on the shifted image. In general, we have used an incremental registration process which first align the fiducial points on face for consecutive frames by translation and then use this estimation to align with a single frame in the video sequence. Our algorithm automatically identifies the frame to which all the other frames should be registered by computing a distance matrix. Homography calculates a linear transformation between the two frames on the basis of the 49 landmark points. Figure 4.5a shows the points in the original frame sequence. Figure 4.5b show result homography. This method facilitates us to improve the mean deviations of pixels from 10 pixel unit to 1.49 pixel unit.

4.5 Peak Detection and Re-sampling

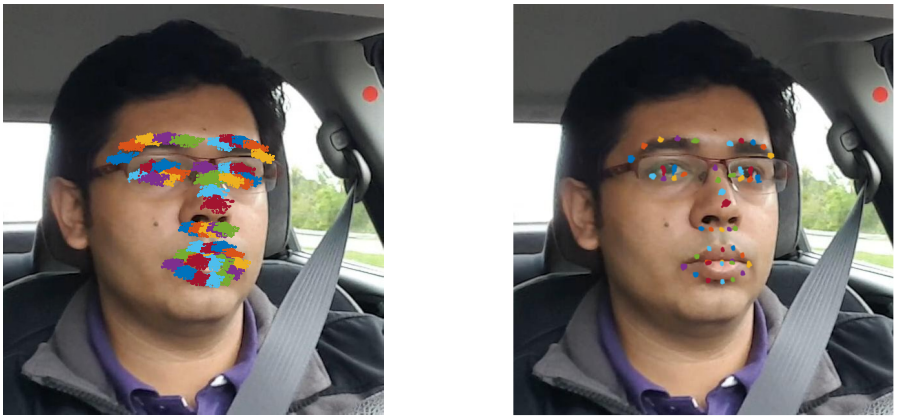
In most real-life applications of interest, video data with very high frame rate is not available. Recordings from standard mobile devices and surveillance cameras often range from 10 frames per second (fps) to 30 frames per second. The frequencies of interest for heart rate lie within the range of 40 to 200 beats per minutes (bpm), which is equivalent to 0.667 to 3.33 Hz. The rate of 10 fps satisfies the Nyquist criterion for this frequency range, but imposes limitations on the accuracy and resolution of the estimated pulse rate signal.



(a) (b)

(c) (d)

Figure 4.4: Face registration: a, b) Two randomly chosen faces with automatically detected landmark points. Superimposed frames: c) before registration, d) after registration.



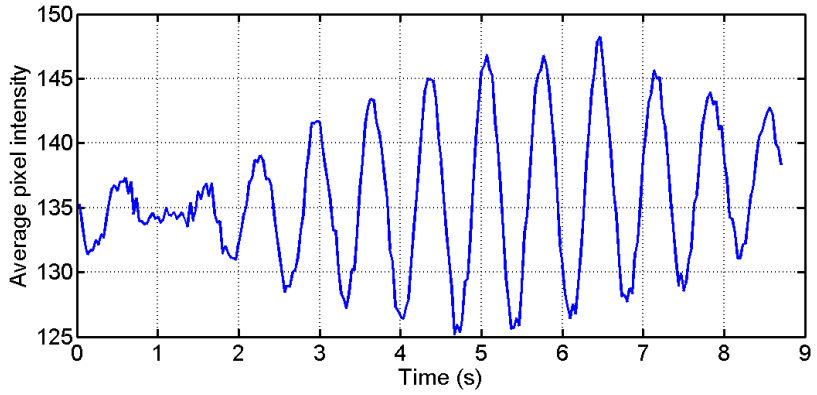
(a) Original points (b) After homography

Figure 4.5: Face registration: As the head moves, the locations of the fiducial points change. We show movement of 49 fiducial points over a sequence of 20 seconds in (a). We use homography to register the face. The fiducial points are used as the feature point locations to compute homography. The homography matrices are used to align the faces from each frame. (b) shows the locations of the fiducial points after registration.

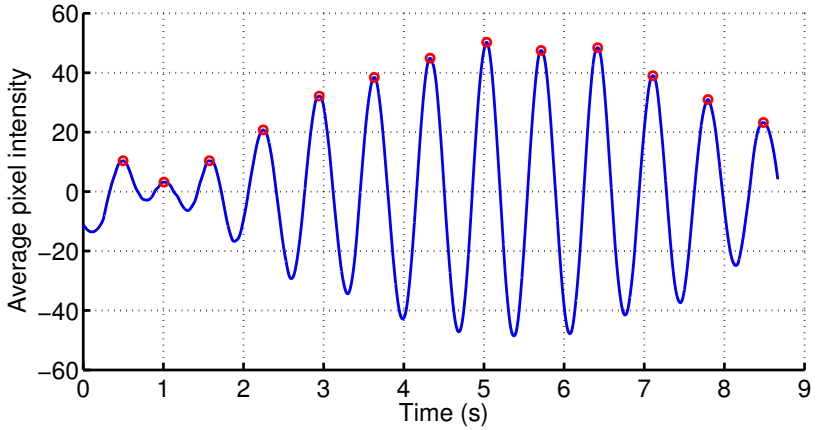
Each registered video sequence is fed to the VidMag algorithm, which attempts to extract the temporal BVP signal. The BVP signal extracted from the magnified video is noisy, partially due to artifacts from motion and illumination, as well as from the low frame rate. Figure 4.6a shows one such sample, which includes spurious peaks. Additionally, the mean intensity varies over time. Even after smoothing and peak detection, the result suffers from low frame rate. For a 30 fps video, the HR estimator cannot detect a heart rate that is known to be between 90 and 94.7 bpm. To address these problems, the BVP signal is re-sampled by a factor of twenty before the peak detection algorithm is applied. This helps to increase the resolution of consecutive heart rate estimates, and heart rate can be calculated with a resolution less than 0.2 bpm.

4.6 Skin Detection and Improved Beat Accuracy

After we apply the video magnification algorithm to the registered face video, it is necessary to identify the skin pixels. This is necessary because all the pixels in face do not contribute to the pulse rate information. For example, pixels from eyebrows or nostrils include noise

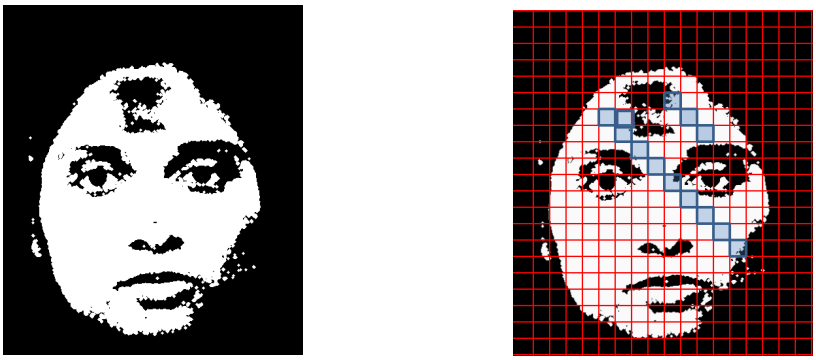


(a)



(b)

Figure 4.6: The peak detection and smoothing steps are shown here. (a) shows the raw data extracted from the magnified video. This signal is re-sampled at a higher rate with spline smoothing such that the resolution of the heart rate is improved. (b) shows the result after peak detection after the re-sampling step.



(a) (b)

Figure 4.7: Skin detection and patch selection. We use a custom skin detection method (details in Chapter 5) to select skin pixels. Then we divide the the skin pixels in square patches of size 10×10 pixels. We select Patches that shows good SNR values in the frequency of interest.

in the measurement. Therefore we need to eliminate them during the post processing. It is to be noted that, there were no standard algorithm available which detect skin pixels in grayscale images. We develop a novel skin detection algorithm.

For detecting in grayscale, we automatically sample some pixels which we believe to belong to skin. Then we learn a statistical model with the grayscale and texture information from the sample space. Finally, applying a prior distribution of face location from the face detector, we apply the statistical model in a Bayesian framework to identify skin pixels. Figure 4.7 shows an example of skin detection and subsequent processing. Once the skin pixel locations are known, we divide them in small patches. In most of our work, we have used square patches of size 10×10 pixels. We extract heart rate information from each of the small patches, detect the pulse location and align them together (Figure 4.8).

4.7 Patch Selection

In practice, each pixel position from the detected skin may contribute noise of different magnitude and phase to the temporal pulse rate signal extracted from them. Therefore we divide the skin pixels of the face in number of small square patches, in this case pixels patch and extract pulse rate signal from each of them. During the registration process, we have seen that the mean error was 2 pixels. Therefore we have chosen a size of pixels for the

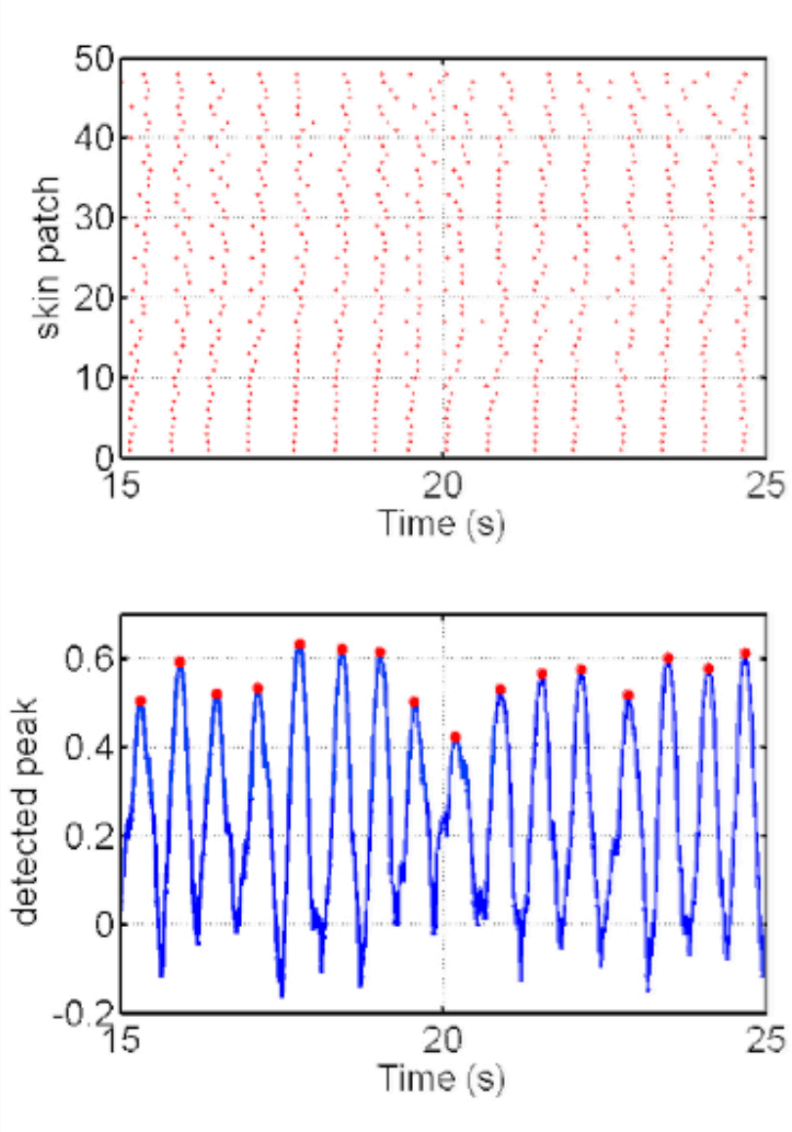


Figure 4.8: The peak locations from each patch are aligned and the average heart rate is calculated.

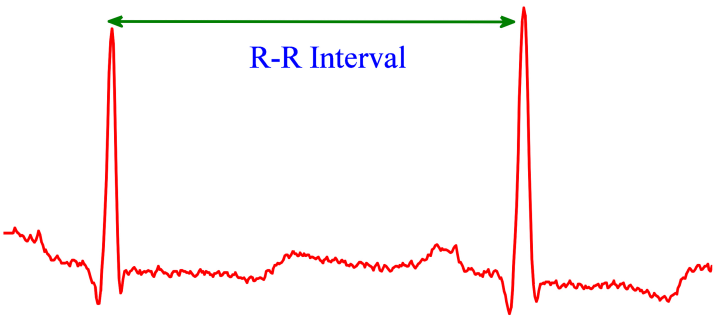
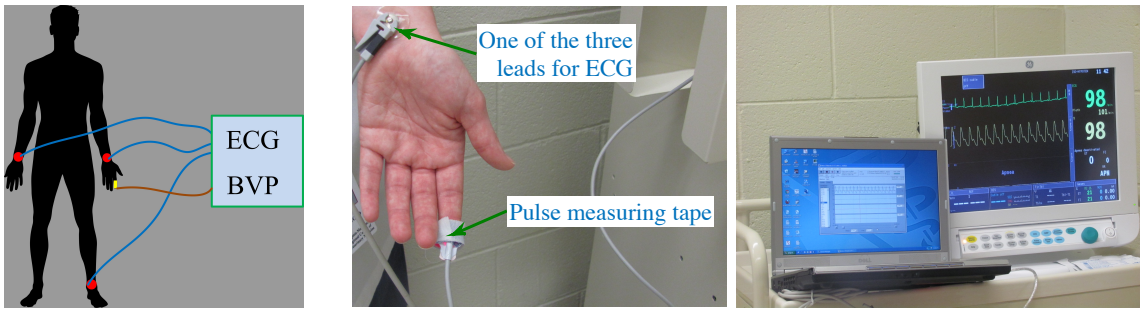


Figure 4.9: R-R interval. A sample electrocardiogram (ECG) signal from the MIT-BIH arrhythmia dataset [5]

squares to minimize the effect of noise from registration error. Formally, we have divided the skin pixels into 10×10 skin patches. Each patch from the magnified video independently estimates the pulse signal . We detect the pulse positions from each of the patches by using a peak detection algorithm. The pulse positions are then aligned together to determine the real signal and eliminate the outliers. Here the basic assumption is that all the patches should ideally produce exactly same signal without any noise and phase delay.

Before we describe the methods for post processing of the extracted heart rate, we would like to introduce the idea of heart rate variability. Consideration and conservation of HRV is the biggest motivation for our post-processing algorithm development. HRV represents the variability of time between two consecutive pulses (or R-R interval) as shown in Figure 4.9. HRV reflects the state of the autonomous nervous system and the balancing of the sympathetic and parasympathetic nervous systems. A detailed review by Acharya et al. [34] shows examples of time domain analysis and frequency domain analysis of HRV. Healey and Picard [33] use Lomb periodogram to study the stress level of a driver on the road. It shows that the power spectral density in different frequency ranges is correlated with the stress level of the driver. Their work also shows that frequency analysis can provide information concerning the respiration of the driver. Noelia et al. [75] show the potential of using HRV for drowsiness detection using R-R intervals of ECG data. Their technique calculates different time domain and frequency domain properties, and they show that specific HRV properties indicate distinct evidence of drowsiness. Acharya et al. [34] further state that alcohol induces sympathetic activation and/or parasympathetic withdrawal. This in turn affects the measured HRV parameters. These characteristics of HRV could be crucial in assessing psychophysiological conditions of a person.



(a) (b) (c)

Figure 4.10: Experimental setup for data collection phase III. (a) Schematic of the setup. (b) Photo-plethysmography pulse meter is worn at the fingertip and one of the three leads of ECG measurement kit attached to the left wrist of the participant. (c) The S5 machine with the data recording software.

4.8 Static Data Collection for HRV Validation

We have seen that heart rate variability plays a crucial role in defining the physical and psychological condition of a person. Therefore it is necessary to evaluate how accurately the VidMag algorithm holds the beat-by-beat accuracy. We have used an FDA certified medical instrument GE S/5-Collect to measure the ECG and pulse signal at a resolution of 100 Hz while recording video at a rate of 30 fps using Canon Powershot on a tripod. Figure 4.10a shows the schematic of the experimental set up. The S5 uses a 3 point ECG measurement where the pulse is measured from the PPG pulse-meter attached to the finger.

Figure 4.10c shows sample signals from the ECG and pulse meter reading. As discussed in previous sections we are interested in measuring the accuracy of the beat to beat interval as well as the actual traces of the signals. We have recorded two participants in a total of five sessions. The duration ECG readings are approximately one minute and PPG readings were 150 seconds. The participants were at their rest position but no restrictions on the head movement were imposed. We shall discuss the post processing methods for rPPG with the help of the collected data.

4.8.1 Pulse signal extraction

This part explains post processing of the raw heart rate signal extracted from the magnified face video. Estimation of actual heart rate is often challenging, particularly when the range

of heart rate in a given time sequence is unknown. Human heart rate can vary from 40 bpm (0.667 Hz) to 240 bpm (4 Hz) depending on the physical and cognitive condition, age, gender and various other factors. As the probability of 240 bpm (4 Hz) is very low in most of the real life scenario, we first use a wide band-pass filter in the range of 0.667 2.5 Hz and magnify any temporal signal in this range in the evaluation test. Due to the use of a wide band pass signal, the extracted heart rate is often corrupted by noise from frequencies other than true embedded heart rate. This noise can be eliminated by learning the passband frequencies, or in other words, using a narrow band filter. At this point, we assume that the signal to noise ratio in the initially magnified video is high and most of the true heart rate signals have been captured. Therefore if we perform a histogram analysis of the extracted heart rate, we shall get an estimate of the actual range of HR for that specific time sequence. The 5 percentile and 95% frequency values are assigned as the lower and upper cutoff frequencies for the updated band pass filter. Analysis of MIT-BIH arrhythmia dataset [59] shows that instantaneous human heart rate may vary within a range of $20-50$ bpm in one minute. This justifies our choice of new frequency band. Once the passband is defined, we again run the video magnification algorithm with the new band-pass filter. In more challenging scenario, this can be performed in an iterative way until convergence. Table 4.1 shows the algorithm to find the final narrow band pass frequencies.

4.8.2 Savitzky-Golay filter analysis

Once the frequency range of the passband filter is determined, the raw signal from the magnified video is extracted and heart rate is estimated for every single beat. The blue line in Figure 4.11 shows the raw output of the extracted heart rate. Clearly, the output has considerable mismatch with the ground truth measurements (green trace) from the pulse meter. This is often incorporated due to noise from sampling and peak detection error. To minimize this error we have used a Savitzky-Golay filter (S-G or SGolay filter) [76, 77]. The heart beats of human does not appear at a constant interval. Therefor heart beat data is a classic example of non-uniformly sampled data. Because of this fact, S-G filter runs a locally concentrated least square method to smooth the raw data. The filtered result is shown in Figure 4.11. It can easily be noticed that the filtered result (HR_{vm}) is very much close to the ground truth (HR_{gt}). We further use a moving average filter with a window of five second to get the average heart rate (HR_{avg}).

Table 4.1: The algorithm to Select narrow band pass frequencies for the band-pass filter of the VidMag.

Algorithm 4.1: Select narrow band pass frequencies
Step 1: Select a face video
Step 2: Initialize passband frequencies $f_l = 0.667Hz$, $f_h = 2.5Hz$
Step 3: Perform video magnification frequencies f_l,f_h
Step 4: Calculate instantaneous heart rate
Step 5: Perform cumulative histogram analysis
Step 6: Assign new passband frequencies f_l,f_h, from 5 and 95 percentile value from Step 5
Step 7: Repeat step 3 step 6 until convergence

$$HR_{avg}(i) = \frac{1}{2n_i + 1} \sum_{k=-n_i}^{n_i} HR_{vm}(i + k) \tag{4.1}$$

When we perform the error analysis, the output HR from Figure 4.12 has a high correlation of 80.4% with the ground truth (Figure 4.13). Table 4.2 shows the statistics of the error in HR estimation. We can achieve 1 percentile error of 1.35 BPM for average HR and 5.9 BPM for instantaneous HR. Table 4.2 also demonstrates the benefit of using skin detection and patch selection. For cases (a)-(b), we manually selected area of interest and patches to calculate the error bound for instantaneous HR (HR_{inst}) and average HR. With inclusion of more non-skin areas, or exclusion of important skin patches, the error rate increases in comparison to case (c), where, we use our skin detection algorithm to select skin patches.

4.8.3 Lombs Periodogram

Finally, we compare our results in frequency domain. The sampling rate of heart rate signal and the $R - R$ intervals are non-uniform in nature. Therefore, standard frequency analysis

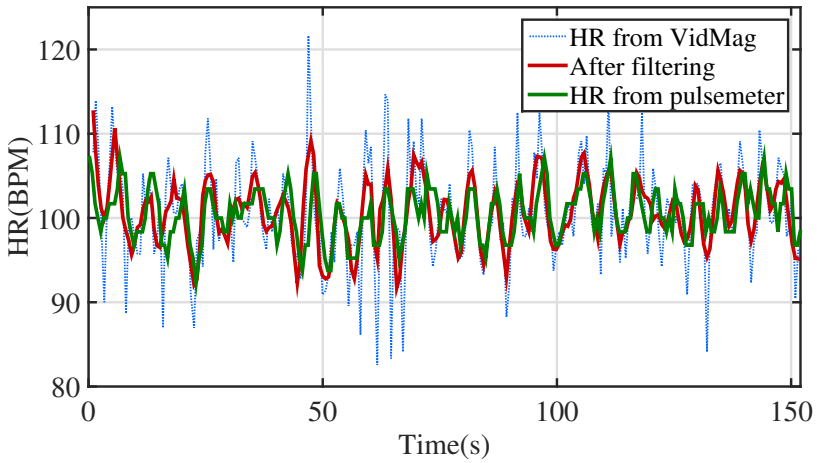


Figure 4.11: Heart rate estimation after filtering

Figure 4.12: Average heart rate estimation after filtering

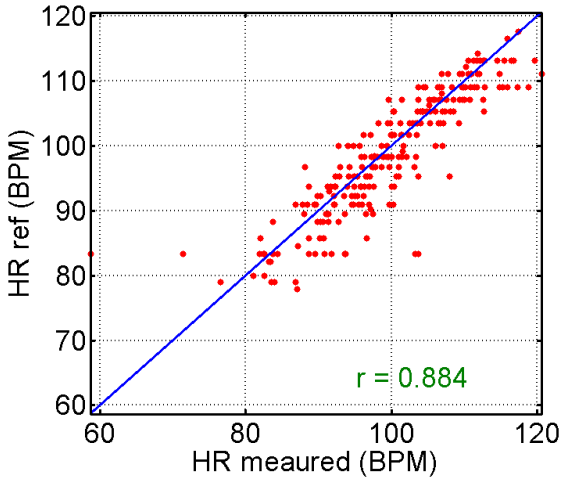
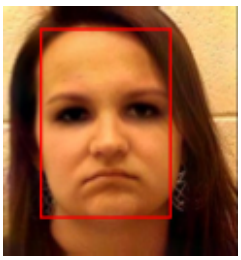


Figure 4.13: Average heart rate estimation after filtering

Table 4.2: Comparison for the effect of skin detection and patch selection. (a) Manually selected region of interest which includes most part of the face including eyes, mouth, (b) Only forehead area (c) Selected patch and then use of weight for each patch.

	 (a)	 (b)	 (c)
$HR_{inst}error$	8.4 BPM	9.1 BPM	**5.9 BPM**
$HR_{avg}error$	1.8 BPM	1.7 BPM	**1.35 BPM**

methods using standard periodogram often fails to provide correct analysis. If we have any discrete time series data

$$\boldsymbol{x}\left(\boldsymbol{t}\right)=\left\{x\left(t_i\right)\right\},\ i=1,2,\ldots \tag{4.2}$$

Then the classical periodogram calculate the spectral power as

$$P_{\boldsymbol{x}}\left(\omega\right)=\frac{1}{N}\left|\sum_{i=1}^{N}x\left(t_i\right)e^{-i\omega t_i}\right|^2,\qquad \omega=2\pi f \tag{4.3}$$

When the samples are equispaced, then FFT algorithm can easily calculate the spectral power at each frequency components. But, when the samples are taken at non-uniform time intervals, this algorithm does not apply. Re-sampling the data to uniform sampling is a simplified solution to this problem, but, we run into the risk of losing data as well incorporate error introduced by interpolation. Lomb [78] and Scargle [79] addresses this problem by introducing a least square spectral analysis method

$$P_{Lomb}\left(\omega\right)=\ \frac{1}{2\sigma^2}\left[\frac{\left(\sum_i\left(x_i-\overline{x}\right)cos\left(\omega\left(t_i-\tau\right)\right)\right)^2}{\sum_i\cos^2\left(\omega\left(t_i-\tau\right)\right)}+\frac{\left(\sum_i\left(x_i-\overline{x}\right)sin\left(\omega\left(t_i-\tau\right)\right)\right)^2}{\sum_i\sin^2\left(\omega\left(t_i-\tau\right)\right)}\right] \tag{4.4}$$

where

$$tan\left(2\omega\tau\right)=\frac{\sum_i sin\left(2\omega t_i\right)}{\sum_i cos\left(2\omega t_i\right)} \tag{4.5}$$

$$\overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i,\ \sigma=\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\overline{x}\right)^2,\ x_i=x\left(t_i\right). \tag{4.6}$$

The parameter τ is a time delay that Scargle introduced. He has proved that the τ is selected by the algorithm in such a way that it makes the expression of $P_{Lomb}\left(\omega\right)$ equivalent to estimating the harmonic content at any angular frequency by least square fit to the model

$$f\left(t,\omega\right)=Acos\ \omega t+Bsin\ \omega t \tag{4.7}$$

This makes the algorithm a better suited for any non-uniformly sampled data than conventional FFT methods. It emphasizes on each point, so all the observation points are used for the periodogram. Laguna et al. [80] has shown the usefulness of Lombs method for analysis of the HR. Reader may particularly consult the work in [78, 79, 80]. We have used Lomb periodogram to compare our results and ground truth. Figure 4.14 shows that our result matches closely with the reference heart rate. Although there are some noises in the higher range of the periodogram, most of the high power frequency components are there.

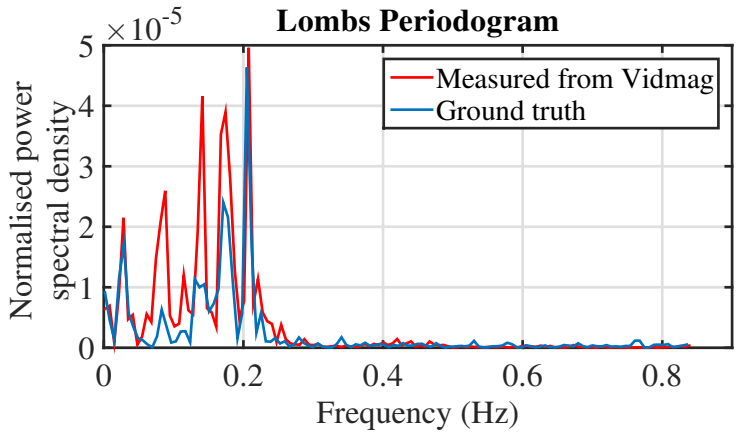


Figure 4.14: Comparison of instantaneous heart rate in frequency domain. The power spectral density is computed using Lomb's periodogram.

We further analyze the whole sequence of heart beat with spectrogram analysis. We select a small window (T_w) in the given time and then perform the Lombs periodogram. Then we shift the window by (T_sh) and perform it again. We have used $T_w = 30sec$ and $T_sh = 1sec$. Figure 4.15 shows the comparison for the spectrogram.

4.9 Validation with Large Scale Dataset

Experimental results from Section 4.8 show that under static lighting condition and restricted head movements, we can achieve good correlation between the estimated heart rate from face video and the ground truth heart rate measured using PPG device. This results demonstrate the feasibility of rPPG in estimating beat-to-beat accuracy of the R-R intervals (instantaneous heart rate). As discussed in Section 1.1.2, analysis of HRV from instantaneous heart rate gives good indication of human psychophysiology, in particular the balance between sympathetic and parasympathetic nervous system, that reflects the emotional condition of a person. Hence these experimental results proves that RGB face video data preserves the HRV information that can be used for emotional analysis.

4.9.1 DEAP dataset

Now, as we know that HRV information can be retrieved from face videos, we extend our experiment to large scale video data with more challenging indoor scenarios. We have used

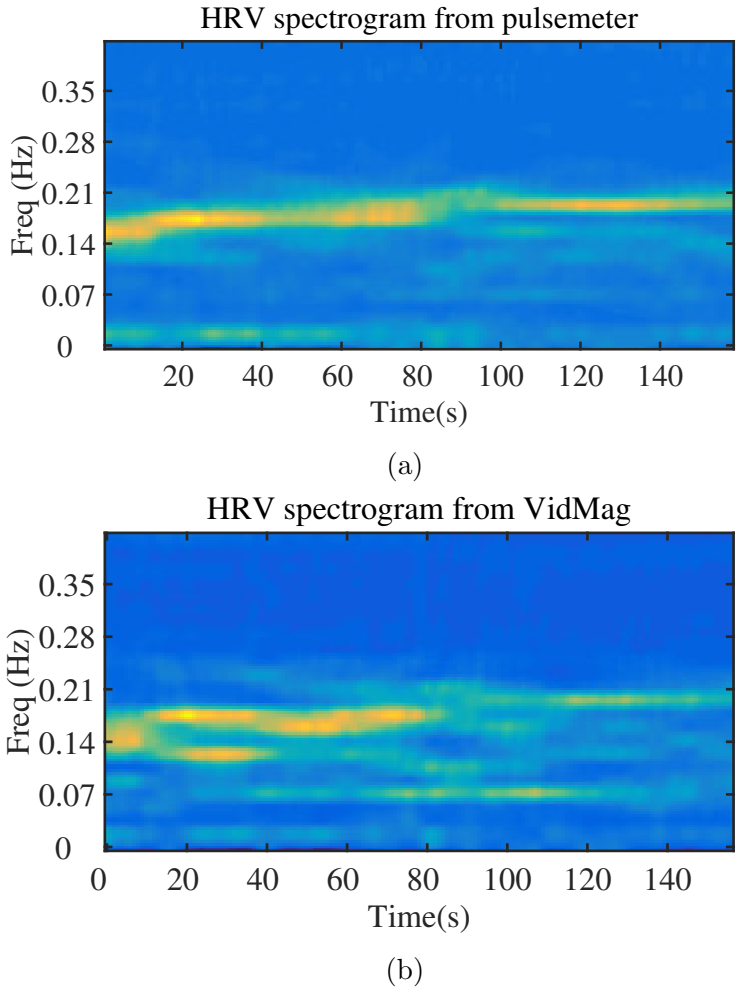


Figure 4.15: comparison of video magnification algorithm by spectrogram analysis. (a) shows the spectrogram from the pulse meter and (b) shows the results from the video magnification algorithm

the DEAP dataset (same dataset used for PPG biometrics). DEAP is a publicly available dataset which comprises face video of individuals sitting in an indoor condition. Each subject is seated in front of a computer monitor and videos are played on the monitor. As the video is played on the monitor, the intensity of the monitor changes. This in turn changes the incident illumination on the subjects' face with time. This changes often not visible in normal eye, but considering the magnitude of the rPPG signal, this changes produce significant noise. We evaluate accuracy of rPPG under this dynamic indoor illumination condition. The video is recorder at a frame rate of 50 fps and with a resolution of 576×720. For each subject, forty sessions on an average is recorded with each session spanning for 60 seconds. The PPG data is recorded in sync of the video data with Biosemi ActiveTwo system [46]. The dataset contains video data from 22 subjects. In our study we have used the first 17 subjects. Figure 4.16 shows the experimental setup.

Figure 4.16: The experimental setup for DEAP dataset. The participant seats in front of the a monitor. Video is played on the monitor and the physiological variables are recorded synchronously. *image source:* `www.eecs.qmul.ac.uk/mmv/datasets/deap/`.

4.9.2 Patch Weight Determination

As shown in Figure 4.7, we initially divide the face in 10×10 skin patches and then track the average intensity from each patch over time. But, it is interesting to ask if signal from

each of the patches are reliable?Researchers have used different region of interest from the face area to extract rPPG signals. In most cases, the researchers have chosen the skin pixels from the forehead area, cheek area or a combination of these areas. In most of the cases this selection is performed heuristically. Lempe et al. [81] have used a discrete Fourier transform (DFT) based metric to find patches of skin that has strong rPPG signal. Their analysis was mainly performed for restricted condition where the illumination were fixed and the faces were almost static. Also, they concluded that cheek and forehead area are the most reliable area.

For more diverse situations like the current test case, choosing a fixed face area like forehead and cheek can be erroneous. Depending on the ambient illumination and shading, a certain patch may not exhibit good rPPG signal. To test this, we first select a rectangular area around face as shown in Figure 4.17. Then we divide the area in small non-overlapping patches of 10×10 pixels. Then, we follow the method described in Section 4.8.1 for each patch and compute heart rate for a single session of 60 seconds. Finally, we compute an rms error of the estimated heart rate compared to the ground truth PPG measurement. Figure 4.19 shows heat-map of the RMS error values for all of the 40 sessions. Each sub-figure represents the rectangle around face (as shown in Figure 4.17). Blue color represents low rms error and yellow represents high rms error. Hence, one should select the blue patches in order to compute the rPPG. We have observed from 4.19 that the heat map for different sessions are different from each other. In most cases, the skin area is highlighted, but the relative locations of the patches are not consistent. Also it is interesting to see that the right cheek (perspective of the viewer) area shows lower error than the left cheek. In ideal condition, they should be identical.

The main reasons for this kind of differences are the effect of ambient illumination, shading and occlusions. The shading effect of incident light is clearly visible from Figure 4.17. The right part of the face (from the perspective of viewer) and the left part of the face is exposed to different amount of ambient light. hence their behavior can be different. To test this, we chose two rectangular patches outside face area (see Figure 4.17) and compare the signal to noise power for each skin patches. The yellow rectangle is taken from a surface that is at a similar distance as the face is from the computer screen. The other rectangle (violet) is taken from the background wall which is comparably distant from both the face and the screen. We record the temporal signal from both of the patches and compare their power in heart rate frequency band.

We have computed the power spectral density (PSD) of the background signals and the

patch signals. Next, we computed total power of all the signals within a fixed frequency of interest $[f_l \quad f_h]$. As the heart rate determines the power distribution from an rPPG signal, we have restricted the frequency range from $f_l = 0.9Hz$ (equivalent to 54 BPM) to $f_h = 1.6Hz$ (equivalent to 96 BPM). Next, we computed the ratio of the total power in a skin patch, i to the total power in the background patch.

$$R_i = 10 \log_{10} \left(\frac{\sum_{f_l}^{f_h} P_i(f)}{\sum_{f_l}^{f_h} P_{bg}(f)} \right) \tag{4.8}$$

Here, P_i and P_{bg} represents the PSD of the signals from the i^{th} patch and the background. Figure 4.18 shows heat map indicating the relative power in each patch compared to the background patches. The patch suitable for rPPG estimation should have high R_i value. As seen from the heat map, all skin patches does not show reliability. Also, It is interesting to note that the map does not conform to the symmetry of the face. One advantage of this analysis is that, inadvertently, we have included the information of the changing illumination in our analysis. The temporal sequence of any pixel either in the foreground or in the background contains the change of illumination intensity coming from the computer screen. The light intensity emitted by a sequence of frames in a video is generally smooth if not, it has jump cut. So, if a video contains too many jump cuts, too frequently, then the emitted light intensity will contain high frequency component, otherwise, most of its components will be at the low frequencies. If the change in illumination has more frequency components in a frequency range less than $0.5Hz$ (jump cuts every 2 seconds), it should not interfere with the rPPG frequency of interest.

4.9.3 Entropy Based Patch Selection

In the previous section,we have seen that the signal characteristics from each patch may be different. Even if they contain the same rPPG signal, the strength and the accuracy of the available rPPG signal largely depend on different factors including illumination and ambient noise. In practice, we need to automatically select skin patches that will guarantee good HR estimation and eliminate the ones which contains more noise. Ideally, a good patch should agree in both time domain and frequency domain with the actual signal. The one that does not contain HR information or the one with low SNR is considered as a bad patch and should be discarded from the consideration. Therefore, we have investigated all the good patches and bad patches from the previous section and looked at their frequency response.

Figure 4.20 shows the power spectral density (PSD) of a typical 60 seconds long PPG

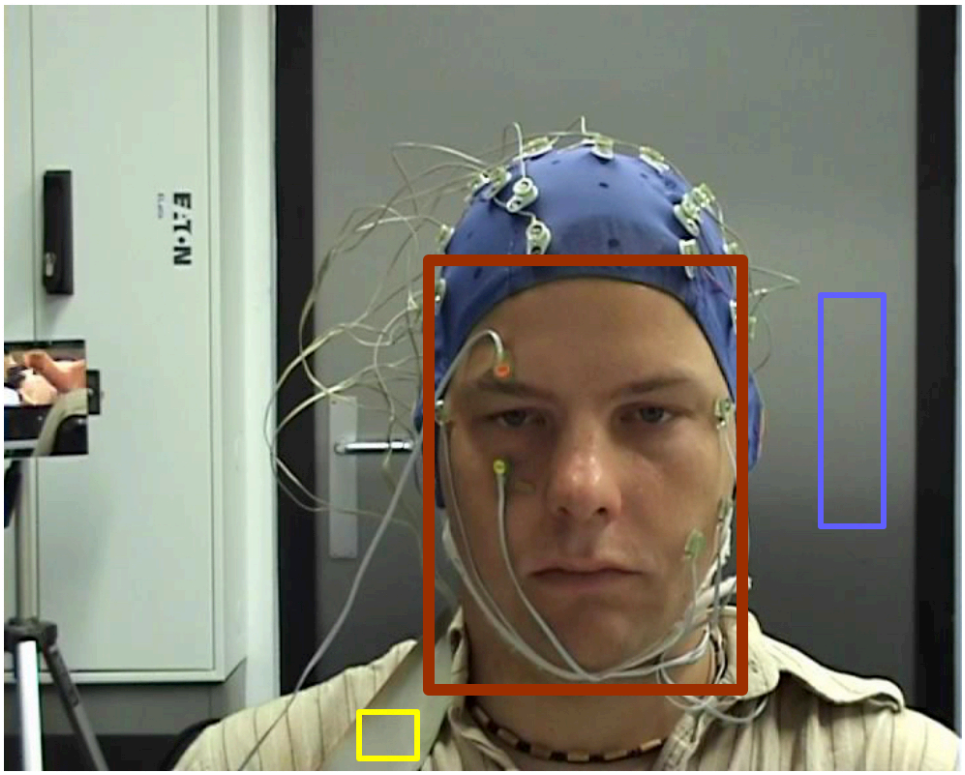
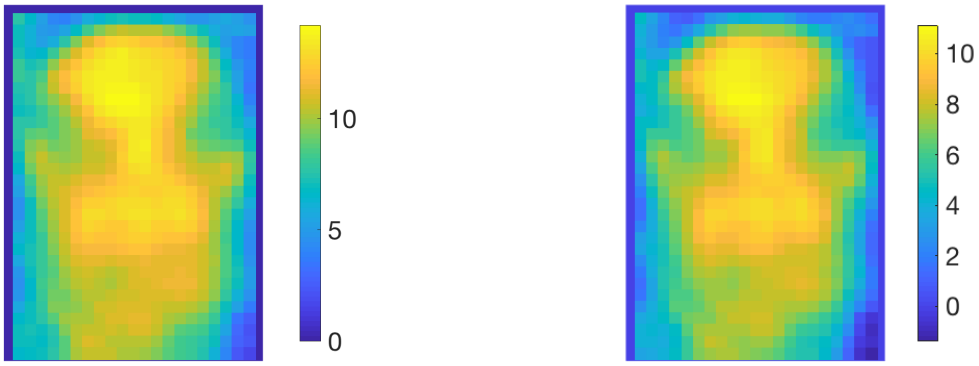


Figure 4.17: A rectangular section (shown in red) is selected as foreground where rPPG can be measured. This section is divided in 10×10 patches and rPPG is extracted from each of the patches. We select two non-face rectangles, where rPPG information is not available. The yellow rectangle selects non-face patches that are closer to the computer screen, and at a similar distance as the face. The violet rectangle is non-face patch from the background wall which is far from the computer screen (frame taken from DEAP dataset)

signal. The signal is taken from one of the sesions of DEAP dataset. In this particular case, the signal carries fundamental frequency components ranging from $0.95Hz$ and $1.2Hz$. This is equivalent to 57 BPM to 72 BPM. The heart rate in that interval varies around a mean value which is reflected as the peak of the signal. Next we looked at the PSD of the good patches and bad patches. Figure 4.21a and Figure 4.21b shows two examples of a good and bad patch respectively. Evidently, the PSD of a good patch conforms with the PSD of a PPG signal in Figure 4.20. But more importantly we observed that a good patch has its power concentrated around a small frequency range of 15 or 20 BPM, where the power of a bad patch is distributed all over the frequency range. In other words, the bad patch has uniform distribution or higher entropy in comparison to a good patch. We can use this information



(a) (b)

Figure 4.18: The figure shows the ratio of total power in the frequency of interest for rPPG. It shows the relative power in dB for each face patch when compared with (a) total power in the patch from background shown by violet rectangle (b) total power in the non skin patch shown by yellow rectangle (4.17)

to differentiate between a good patch and a bad patch.

Hence, we compute the entropy of PSD for each patch and select those patches which have entropy values below a threshold. Figure 4.22 shows entropy map for the same subject as in Figure 4.19. We have used a threshold entropy value of 2.9 to select patches. Once the patches are selected we compute final heart rate in the same way described in Section 4.6 and in Figure 4.7b. The rms error and 5 second mean error for each of the 40 sessions are shown in Figure 4.23. In some of the cases, the session fails to find at least 5 patches whose signal satisfy the entropy threshold. We consider these sessions as not reliable to compute video based rPPG. For 27 sessions out of 40 sessions from the particular subject, entropy based patch selection estimates instantaneous HR with an error less than 6 BPM.

To compare our result, we further select patch from Figure 4.19 where the error is less than 8 BPM and compute heart rate. The result of entropy based method shows only 0.3 BPM increase in instantaneous heart rate error. In most of the cases the errors are comparable (as shown in Figure 4.23a and 4.23b). In total we tested for 674 video sessions from 17 subjects and 223 videos (33%) were qualified as reliable. The instantaneous HR error were recorded as 8.85 BPM and 5 second average HR error were recorded as 7.93 BPM.It has been recorded, that in many sessions, the total instantaneous errors are recorded below 5 BPM. This result is particularly significant as most of the researchers concentrate on average

(sometimes for as long as 30 seconds) HR measurements.

4.10 Conclusion

In this chapter we have demonstrated the feasibility of non-contact monitoring of heart rate in naturalistic situations using RGB face video data. We have used video magnification approach, which exaggerates small, nearly imperceptible changes of blood volume over time reflected in pixel intensities. The use of the video magnification algorithm in real-life situations is challenging, due to head movement, changes in illumination, low spatial resolution, low frame rate, and image noise. This work has shown that a step by step approach to addressing several of these limitations for the purpose of heart-rate monitoring, is effective. With our algorithm, we can achieve very good estimation of instantaneous heart rate. We have also shown that an entropy based analysis can help us to differentiate between good patch and bad patch in a dynamic illumination condition. In the process we have developed a novel skin detection algorithm that will be discussed in the next chapter.

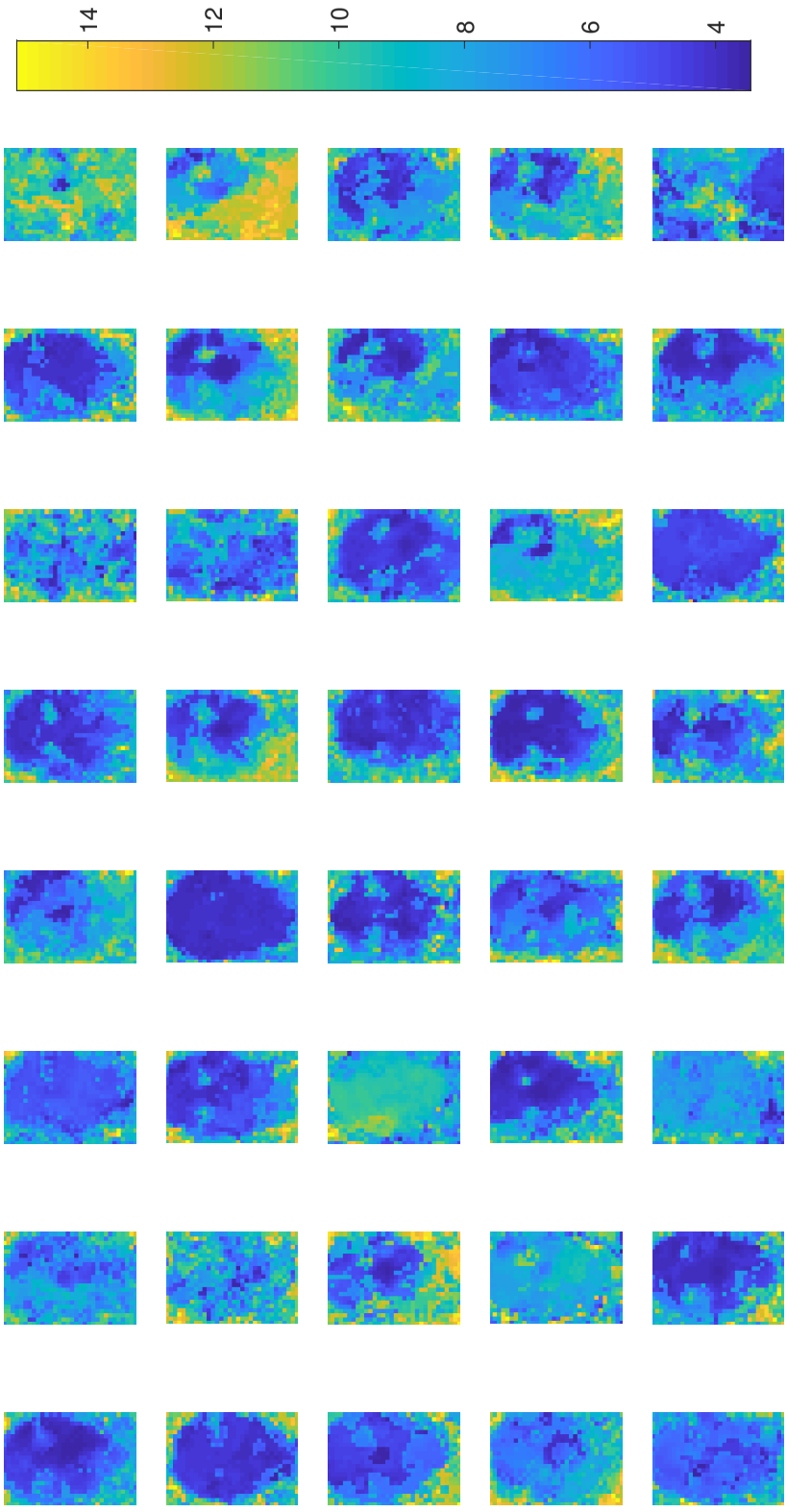


Figure 4.19: Heat map representing RMS error for estimating HR. Each of the 40 sub-images are results of 40 different sessions for the same subject. The error map for each of the cases are distinct. In most cases part of right cheek (perspective of the viewer), forehead is seen to be a good patch.

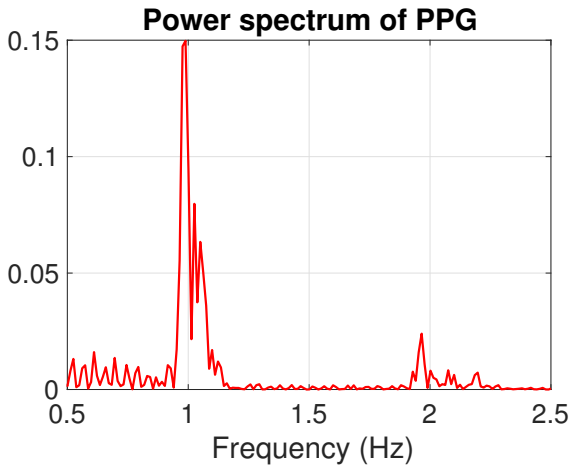
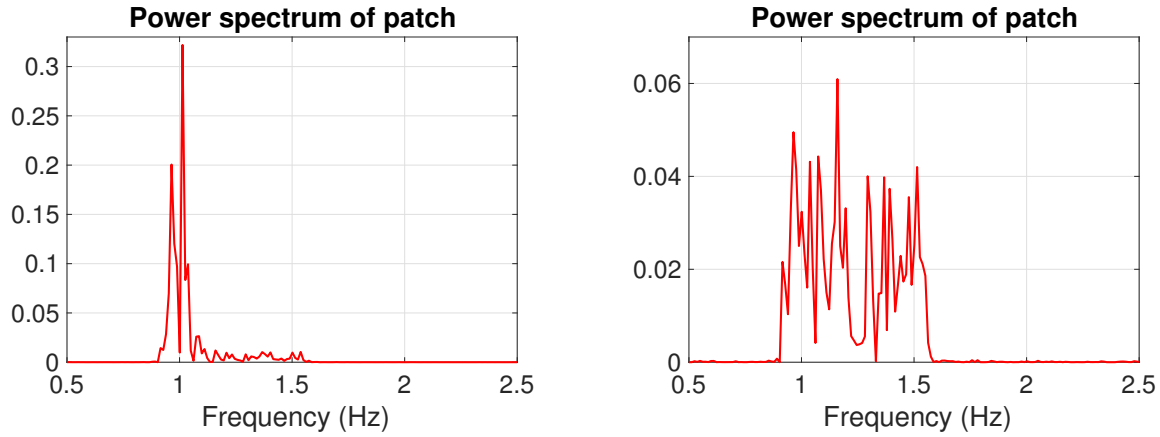


Figure 4.20: Power spectral density of 60 seconds PPG signal. Maximum power is seen at the fundamental frequency (around 1 Hz, equivalent to 60 BPM). The second harmonic is also seen to be significant.



(a) (b)

Figure 4.21: Power spectral density of signals extracted from skin patches. (a) PSD of the patch signal that gives low RMS error. (b) PSD of the patch signal that have significantly high error. Clearly, the PSD of (a) closely matches the PSD of the actual signal (Figure 4.20. Also, PSD of (b) is uniformly distributed over the full frequency range of interest.

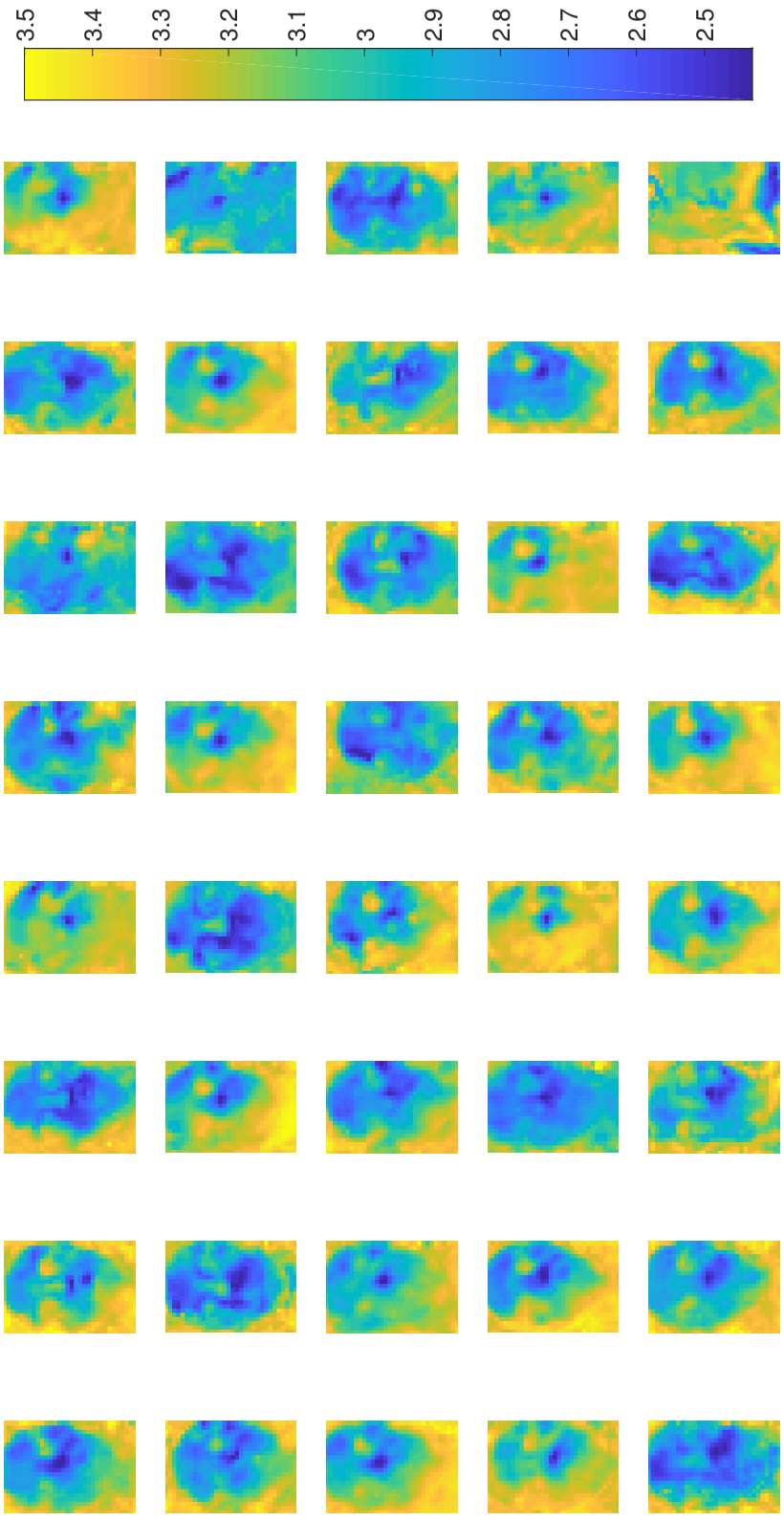
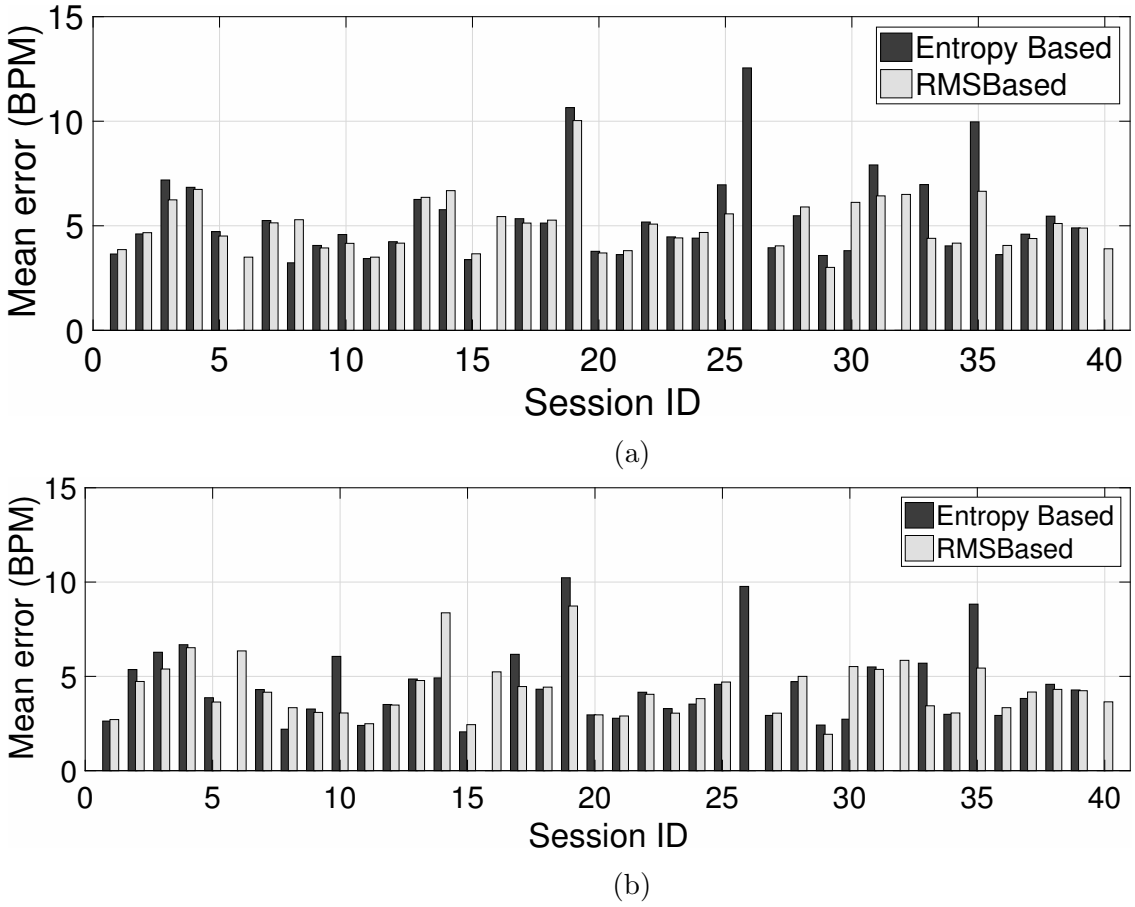


Figure 4.22: Heat map of the entropy value. Blue color shows low entropy and yellow shows high entropy values. The patches with low entropy values generally agrees with rPPG signals. Patches with higher entropy values carry more noise and have low SNR to compute rPPG. Part of forehead, cheeks, area between nostril and upper leap often qualifies as a good area for rPPG.



(a)

(b)

Figure 4.23: HR estimation error for entropy based method. This is compared with patch selection method when the patches are selected knowing the ground truth. (a) It shows the rms error in estimation of instantaneous heart rate. (b) It shown the rms error in estimation of average heart rate averaged over 5 beats. The entropy based method perform similar to the supervised method.

Chapter 5

Skin Detection Without Color Information

5.1 Introduction

The detection of human skin in images is a topic that has received considerable attention in the research community. The primary motivations for detecting skin in images and videos have been face detection (e.g., [82, 83]), gesture and action recognition, surveillance, and adult content screening [84]. Researchers have devoted considerable effort to investigate different color spaces to detect skin under different illumination conditions and for different ethnicities. For example, Vezhnevets et al. [85], Kakumanu et al. [86], and Phung et al. [6] have provided comprehensive surveys of skin color modeling and skin segmentation techniques using color cues within different color spaces. An interesting result from [6], using 4000 images captured "in the wild," is that skin-detection accuracy is not overly sensitive to the choice of color space, nor to color quantization bin sizes.

In order to refine the results of skin segmentation, a few researchers have also used texture [87, 88, 89] or shape [90] information in combination with color-based methods. Wavelets, contourlets [88, 91], and textons are among the textural features that have been used in conjunction with color cues. For all of this previous work, however, color cues have provided the dominant source of information.

In spite of the emphasis on color-based analysis, a considerable number of applications will benefit from a system that can perform skin detection in the absence of color cues. Surveil-

lance cameras, for example, often provide monochromatic video data in the visible band or in the near-infrared (NIR) band. Notably, color cameras often operate under conditions of very poor illumination, resulting in low-contrast images that suffer from large levels of thermal noise and from an absence of useful color information. Furthermore, many image and video archives exist in grayscale formats only, and cannot benefit from existing skin-detection techniques. As a recent example, the Strategic Highway Research Program (SHRP2) has obtained grayscale videos of more than 3400 drivers in naturalistic conditions, in order to investigate the behavior of drivers on the road [92]. That dataset will become much more useful to researchers if skin detection can be performed automatically. Other applications that will benefit from this capability include movie colorization [93], which requires skin patch detection in grayscale images.

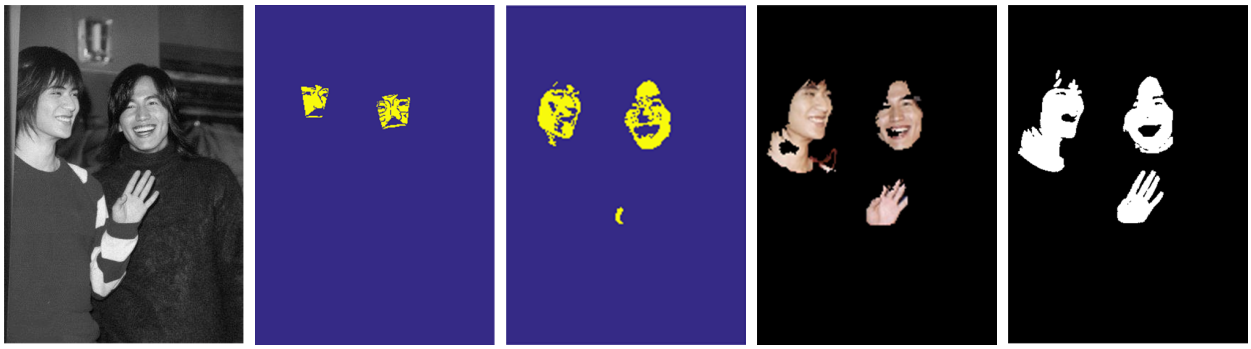


Figure 5.1: Major steps in skin detection without color. (a) Original grayscale image. (b) Initial skin detection, based on prior knowledge from a face detector. (c) Seeds for region growing, chosen based on textural cues from the face regions. (d) Final result after region growing. (Color is shown for the reader's convenience.) (e) Ground truth, for comparison.

We present a technique for skin detection in images without the use of color cues. The approach benefits from recent work involving human face detection (e.g., [94, 95, 73]), which provides a spatial prior that our system uses to guide the search for skin. The main steps of our approach are illustrated in Fig. 5.1. Our system exploits the spatial prior map to perform one-shot learning of textural and intensity-based cues for presumed skin patches. The system uses those cues to seed a region-growing step that can detect skin from the entire body, including face, torso, and limbs.

To the best of our knowledge, this system is the first to attempt skin detection without the aid of color cues. This algorithm learns textural cues that are image-specific, and uses

contextual cues to address the large diversity of skin appearances. Our results demonstrate good performance under different illumination conditions, age, ethnicity, and for multiple subjects in the field of view. Lastly, the proposed method does not require a lengthy training procedure or a large memory.

To motivate the work further, an example is given in Fig. 5.2 for which color saturation has been reduced manually for a sample color image. As the saturation is dialed down, the recall value also falls gradually. Precision remains high until saturation falls to approximately 30% of the original level. Below this threshold, the precision values fall dramatically, and very poor performance would be expected from any color-based skin detection algorithm. By comparison, the new technique presented here detects skin pixels quite well for the grayscale version of this image (precision and recall are 82% and 79%, respectively).

The next section describes the skin detection method in details. Section 5.2.1 demonstrates the importance of contextual information for skin detection, and Section 5.2.2 describes the region-growing approach that we have developed. Section 5.3 describes different features and their characteristics for skin detection, and section 5.4 presents experimental results for two different databases.

5.2 Skin Detection with Grayscale Features Only

The new skin detection method is adaptive to a given image, and it works in two steps. First, it localizes faces, and builds a statistical model to find skin areas for those faces. In the second step, the system learns local texture characteristics from the face pixels and uses a region-growing approach to search for similar textures within the rest of the image.

5.2.1 Skin detection using contextual information

In this first step, our system applies a face detector to localize faces in the image, and then uses a Bayesian model to obtain grayscale statistics and to detect skin pixels within the face region [96]. The schematic of the algorithm is shown in Fig. 5.3. We have used the face detector by [95], although in principle our technique could be used with any face-detection method. Depending on the locations of fiducial points on the face, the system selects a small patch near the center of the face that we call the *mask*. This mask consists of carefully but automatically selected skin patches located around the eyes, forehead, cheeks, and nose.

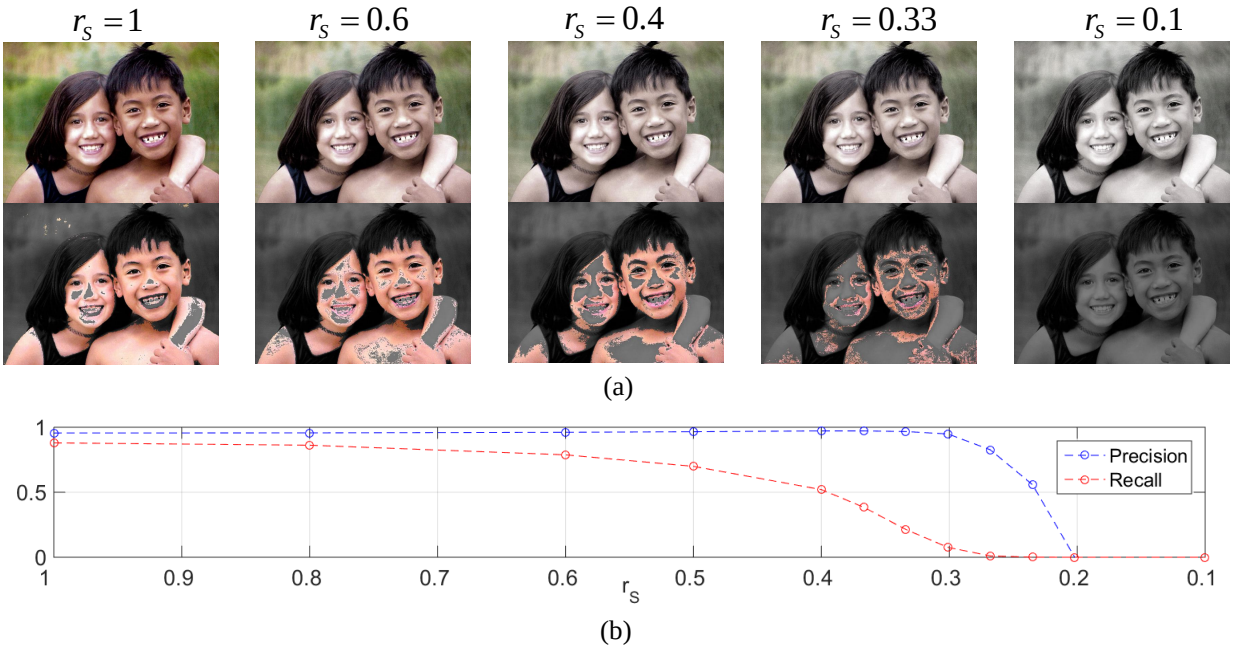


Figure 5.2: The dependence on color for skin-detection algorithms. (a) In the top row, an original image is shown at the left. Proceeding to the right, the color saturation level (indicated by factor r_s) is gradually reduced. For each case, the second row shows the results from the skin-detection algorithm of [6] using a Bayesian color model. Detected skin pixels are shown in color. (b) Precision/recall statistics for different saturation levels. Detection performance for this color-based method is dramatically affected for saturation levels below $r_s = 0.5$.

The mask has been specified so that it possesses low probability of being occluded by facial hair, clothing, or other common obscurants. Implicitly, through the use of this mask, the system eliminates eyes, mouth, nostrils, dark edges, and other non-skin regions from further consideration. We tested our mask-selection procedure using using more than 2000 images from the ECU dataset [6], and achieved a precision of 98.2%.

Because of the mask's location near the center of the face, and thanks to facial symmetry, mask pixels are expected to provide a good representation of illumination variations over the face. For this reason, statistics derived from pixels within the mask, along with the face location prior, can be used to extend the face region and find a additional skin patches within the image. First, the system constructs a histogram of intensity values from the mask and normalizes it to approximate a discrete probability distribution:

$$P\left(I \mid s\right) = \frac{n\left(I\right)}{N_s} \tag{5.1}$$

$P\left(I \mid s\right)$ represents the conditional probability of occurrence of intensity value $I \in \{0, 1, ..., 255\}$ given that it is a skin pixel. The term $n\left(I\right)$ represents the pixel count in bin I, and N_s is the total number of skin pixels in the mask.

We also introduce a *distance-based prior probability* $P_{face}\left(s\right)$ to approximate the expected locations of skin pixels with respect to a given face mask. This term is defined empirically so that $P_{face}\left(s\right)$ is high for pixels near the center of the detected face, with values decreasing gradually as a function of increasing image distance from the detected fiducial points. The probability map, $P_{face}\left(s\right)$, is specified so that it includes pose information. With this framework, an initial set of skin pixels can be detected automatically using

$$P\left(s \mid I\right) \propto P\left(I \mid s\right)\left[P_{face}\left(s\right)\right]^{\alpha} > \lambda \tag{5.2}$$

where λ is an empirically chosen threshold and α is a relative weighting, also selected empirically. We have tested 771 images from the SFA dataset [7], for which each image contains a single face only. (Examples are shown in Figure 5.4a.) Figure 5.4b shows precision-recall curves with a varying set of hyperparameters λ and α. We chose an operating point with high precision (91%) and moderate recall (75%) value.

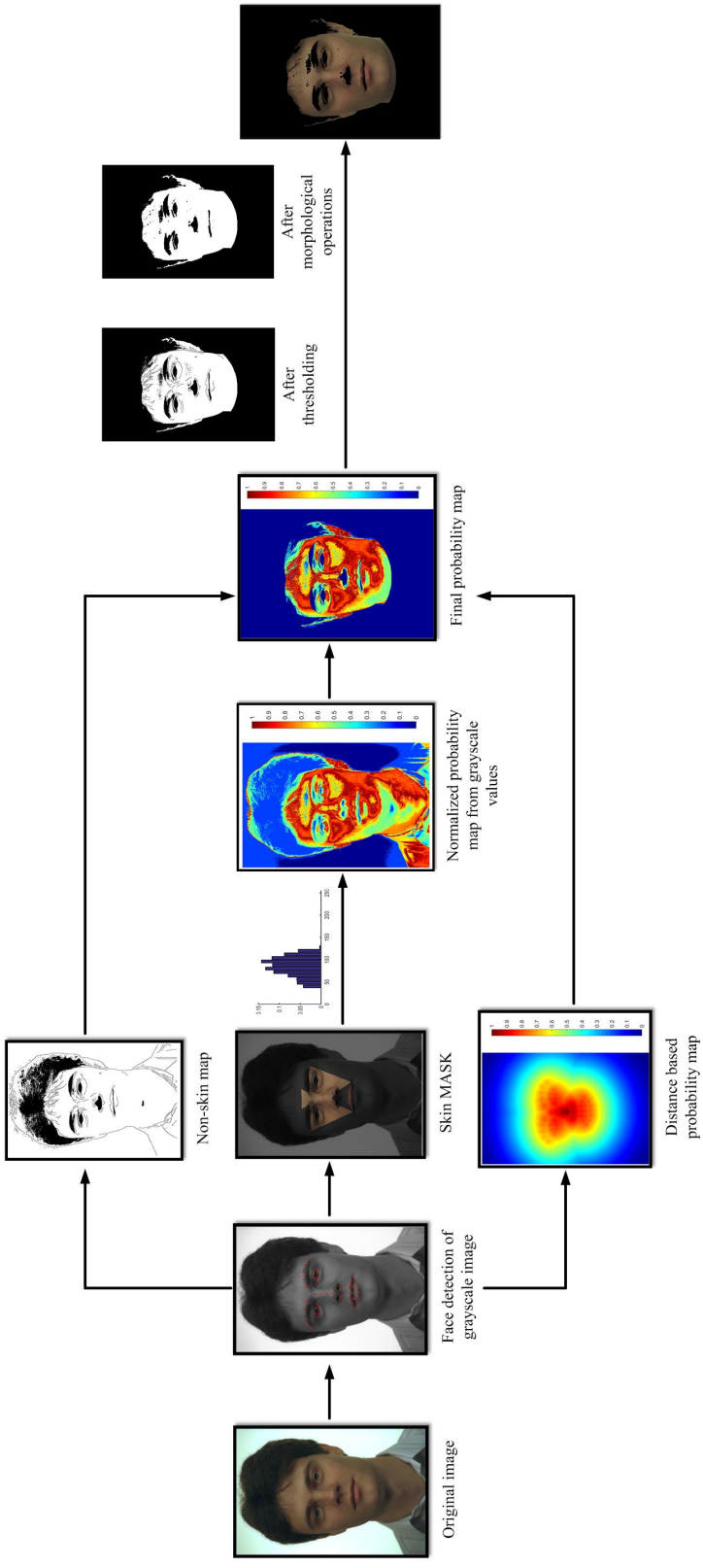


Figure 5.3: Diagram of the first step for skin detection without color. A geometric prior and an intensity distribution are used to select candidate skin pixels from a face.

This discussion has considered grayscale statistics alone. We note parenthetically that this framework could be easily extended for any feature type, including explicit texture cues or even color cues.

5.2.2 Skin detection beyond the face

The technique that was introduced in the previous section is conservative in nature, in order to avoid false positives. It limits the search for skin pixels to image regions near previously detected faces only. In this section, we present a relaxation strategy that formulates a region-growing procedure starting from the skin patches that were selected in the previous section. We assume that the characteristics of skin pixels for a person in an image do not vary significantly. Let Ω_s refer to an image region that represents a skin patch from section 5.2.1. In this section, we shall use this skin region to extract representative texture descriptors for skin throughout the rest of the given image. The system learns textural features from Ω_s, and automatically selects a set of pixels (seed points) in the image that are statistically similar to Ω_s. The hypothesis is that these additional seed points represent skin with high probability. These seed points are used in a region growing algorithm to search for remaining skin pixels in the image.

For computational efficiency, the system utilizes superpixels as seed points instead of pixels. Superpixels are clusters of spatially connected pixels that exhibit similar characteristics. We have used the SLIC superpixel algorithm [97], which is fast and can be implemented for grayscale images. For each superpixel i, we estimate the probability of skin conditioned on a set of selected textural features as follows:

$$P_i\left(s|f_1^i, f_2^i, ..., f_n^i\right) = \prod_{k=1}^{n} P_i\left(s|f_k^i\right) \tag{5.3}$$

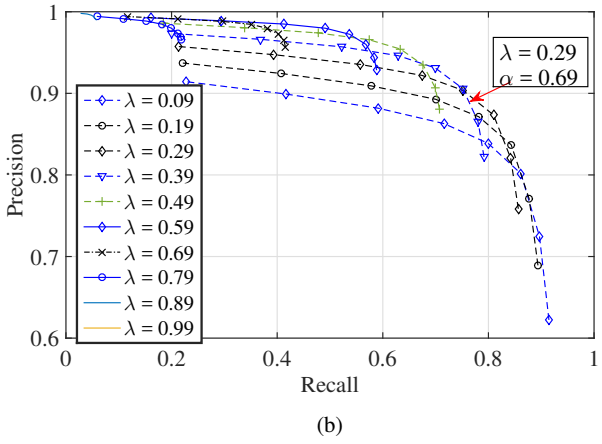
where

$$P_i\left(s|f_k^i\right) = \Phi_k\left(d_{i,\Omega_s}^{f_k}\right) \tag{5.4}$$

is the learned probability of each feature, f_k, computed from the local statistics of Ω_s. This probability is a function of distance metric in the feature space. The distance metric, $d_{i,\Omega}^{f_k}$, is defined as:

$$d_{i,\Omega_s}^{f_k} = \begin{cases} KL(f_k^{\Omega_s} \;||\; f_k^i) & \text{if} \quad f_k \in \mathbb{P} \\ \left\| f_k^{\Omega_s} - f_k^i \right\|_2 & \text{if} \quad f_k \in \mathbb{R} \end{cases} \tag{5.5}$$

(a)



(b)

Figure 5.4: (a) Sample images from the SFA dataset [7]. This database consists primarily of head shots taken indoors. The images exhibit wide variations in age, skin tone, head pose, facial hair, illumination, and wearables such as eyeglasses. (b) Precision-recall curves for skin detection from the SFA dataset for different hyperparameters λ and α. The gray arrow indicates the final operating point.

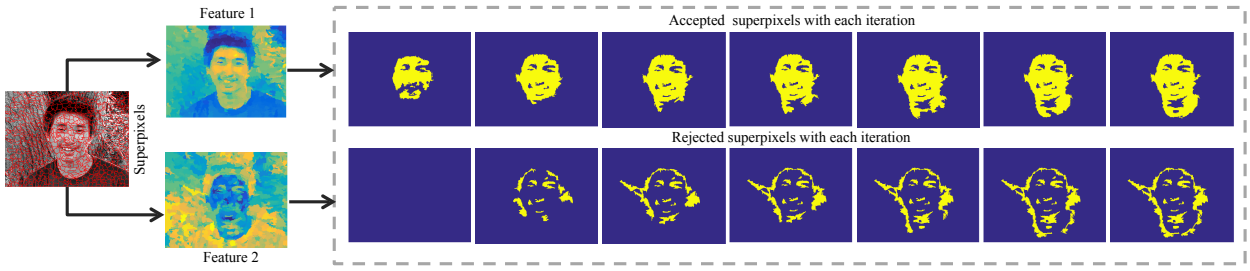


Figure 5.5: Example of region growing algorithm. First the image is segmented into superpixels. Next, texture features are computed. Following (5.4) and (5.5), probability maps are generated and candidate seed superpixels are selected using the threshold from (5.11). Then, with every iteration, the neighboring superpixels are classified as skin or non-skin, until the stopping criteria are satisfied. The top and bottom rows show the selected and rejected superpixels with each iteration.

If the feature represents a probability distribution defined in $\mathbb{P}$, we compute the KL (Kullback $-$Leibler) divergence of the testing superpixels with the average feature distribution in Ω_s. On the other hand, if the feature is a scalar quantity we simply compute the Euclidean distance. We use $\Phi_k(\cdot)$ to map a distance value x to a probability value:

$$\Phi_k(x) = \frac{1}{1+e^{-1/x}} \tag{5.6}$$

5.2.3 Region-growing algorithm

After the probability map $P_i\left(s|f_1^i, f_2^i, ..., f_n^i\right)$ has been computed for all the superpixels, the system uses a threshold to select seed superpixels in the image. For each seed superpixel sp_s we select candidate neighbors N_{sp} and for each neighboring superpixel sp_i, we recalculate the class probability with the belief from the neighboring skin superpixels, N_i^{skin}:

$$P_i^{new}\left(s|F^i\right) = P_i^{node}\left(s|F^i\right) P_i^{edge}\left(s|F^i\right) \tag{5.7}$$

$$P_i^{node}\left(s|F^i\right) = P_i\left(s|F^i\right) \tag{5.8}$$

$$P_i^{edge}\left(s|F^i\right) = \prod_{k=1}^{n} \Phi_k\left(\frac{1}{\mathrm{N}_i^{skin}} \sum_{j \in \mathrm{N}_i^{skin} \setminus i} d_{i,j}^{f_k}\right) \tag{5.9}$$

$$F^i = \left\{f_1^i, f_2^i, ..., f_n^i\right\} \tag{5.10}$$

Again, comparing the new probability with a threshold τ_F, the system decides the candidacy of the superpixel sp_i. The algorithm stops when there are no further candidate superpixels near the seeds. Additional stopping criteria are an empirically selected maximum number of iterations, and maximum deviation of the feature probability value from that of its original set of seed points. Algorithm 1 describes our approach in pseudocode form.

We have developed a general formulation for a combination of features. However, the current implementation is restricted to evaluating a single feature at a time. Therefore, the value $n = 1$ is used in (5.3). The threshold parameter for new candidate superpixel, τ_F, is calculated as the average probability of the seed superpixels sp_s,

$$\tau_F = \kappa \frac{1}{N_s} \sum_{i \in sp_s} P\left(s|F^i\right) \tag{5.11}$$

where N_s is the total number of the seed superpixels. The κ is a hyperparameter and has been selected from precision-recall curve as shown in Fig. 5.8 for individual feature.

Fig. 5.5 shows an example to illustrate algorithm operation. After computing probability maps from grayscale intensity features, the system selects a set of superpixels that satisfy the threshold condition. The top row in Fig. 5.5 represents the accepted superpixels as skin and the bottom row shows the rejected superpixels. At the beginning of the algorithm, the bottom row candidate is empty. With every iteration the algorithm either includes or rejects the neighboring superpixels. Both sets of superpixels grow until no candidate superpixels remain, as shown in the last column.

5.3 Feature Selection for Grayscale Skin Detection

Skin regions tend to exhibit a smooth texture in most of the images, compared to its surroundings. This section describes grayscale-based texture features that can be used to detect smooth regions that may be associated with skin. The emphasis is to make use of grayscale distributions and local binary patterns (LBP) and lacunarity [98].

5.3.1 Grayscale intensity

It has been shown in [99, 86, 6] that skin pixels tend to cluster in distinctive patterns within different color spaces. Therefore, a color-based skin detector can be trained relatively easily.

Algorithm 1: Region growing algorithm

Initially compute class probability of each superpixel and select set of seed superpixels $\{sp_s\}$; from contextual information, define set of non-skin superpixels $\{sp_{ns}\}$;

For $l = 1$ to max iterations **do**
    For $k = 1$ to number of seeds sp_i **do**
        Select candidate neighboring superixels, N_{sp}
        For $i = 1$ to number of superpixels in N_{sp} **do**
            Calculate $P_i^{new}(s|F^i)$
            If $P_i^{new}(s|F^i) > \tau_{F^i}$ then
                Update $\{sp_s\} = \{sp_s\} \cup \{sp_i\}$
            else
                Update $\{sp_{ns}\} = \{sp_{ns}\} \cup \{sp_i\}$
            end If
        end For
    end For
    Check stopping criteria
end For

On the other hand, grayscale intensities for skin pixels in different images do not necessarily cluster within predictable upper and lower bounds. This is often due to diverse variations in skin tone and illumination. Therefore, we do not attempt to develop a universal detector, but instead focus our effort to a single image specific features of skin.

We have assumed that the grayscale statistics within a local, restricted region (in our case, a superpixel) are relatively uniform. We have used mean intensity value of each superpixel, and the discrete probability distribution of grayscale values (similar to (5.1)) for the given superpixel as features. We have also used the standard deviation and entropy in a local neighborhood as additional features. For a neighborhood ($\mathbb{N}_i$) of size $r \times r$, where $r \in \{3, 5, 7\}$, we computed standard deviation and entropy around each pixel i, and the new feature map for the full image is represented as:

$$\begin{aligned} M^r_{STD} &= \bigcup_i \text{std}\left(\mathbb{N}_i\right) \\ M^r_{ENT} &= \bigcup_i \text{entropy}\left(\mathbb{N}_i\right) \end{aligned} \tag{5.12}$$

where for any pixel intensity x_i, and mean $\bar{x}_i$ in $\mathbb{N}_i$,

$$\text{std}\left(\mathbb{N}_i\right) = \sqrt{\frac{1}{r^2} \sum_{j \in \mathbb{N}} \left(x_j - \bar{x}_i\right)^2} \tag{5.13}$$

$$\text{entropy}\left(\mathbb{N}_i\right) = -\sum_i P\left(x_i\right) \log\left(P\left(x_i\right)\right) \tag{5.14}$$

These maps are used to compute composite features for each superpixel, either by the superpixel's mean value or by the distributions of features within the superpixel. Fig. 5.6 shows a qualitative example of the characteristics of these features. The heat map from Fig. 5.6(b) and (c) shows that skin areas from face, arms and legs have similar feature values. They generally have low standard deviation and entropy in local neighborhoods. It also shows that skin pixels often exhibit less textural variation than the surroundings (roadway, buildings, etc.). However, it is interesting to see from Fig. 5.6 that there are several non-skin areas (windows, for example) that also exhibit low variation in texture. The use of contextual information (section 5.2.1) and interconnectivity (through region growing) helps eliminate these outliers.

5.3.2 Local binary patterns

Local binary patterns (LBP) [98] represent a well-established feature type within images. We have used LBP to learn local textural features of skin. Given a pixel p_i, the $LBP_{(P,R)}$ selects P sample points from a circular symmetric neighborhood of radius R. Then a thresholding function $s(x)$ is used to compare the pixel intensity (I_p) with the intensities of the sampled points and create a binary pattern.

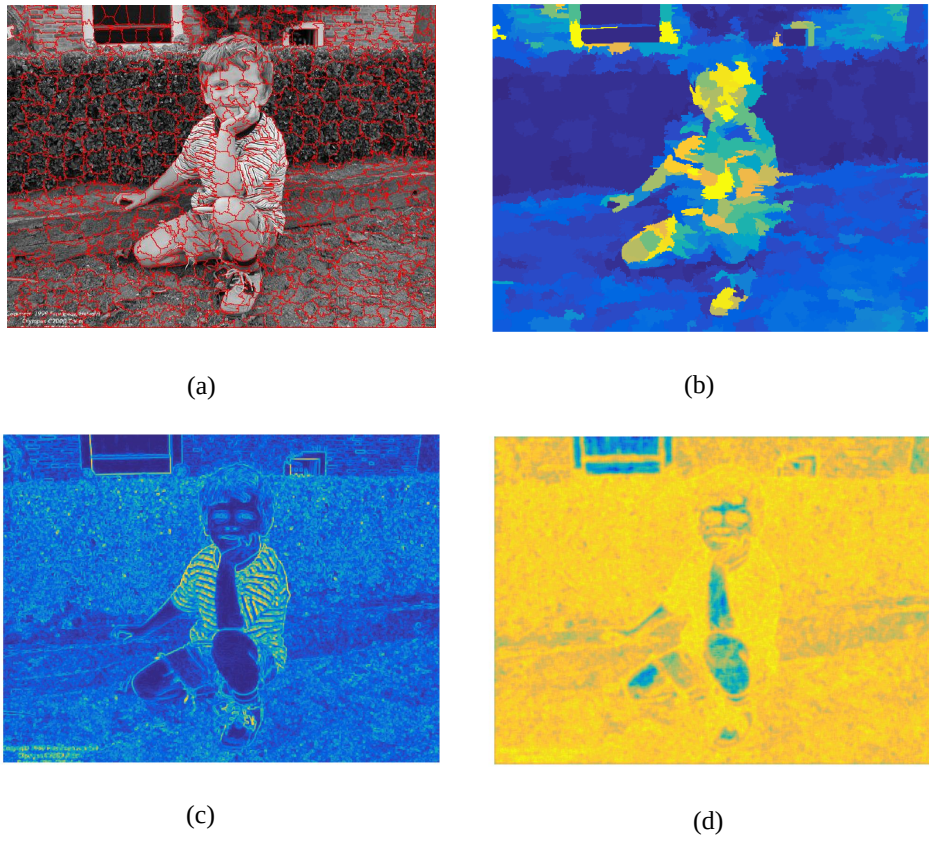
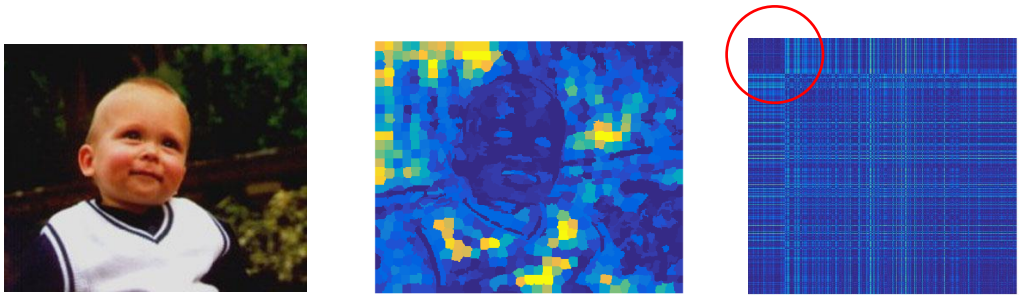


Figure 5.6: Examples of different grayscale feature types: (a) superpixel segmentation, (b) grayscale mean of superpixels, (c) standard deviation (5.13), (d) entropy (5.14). Notice that the standard deviation and entropy maps exhibit similar values for face as well as arm and leg regions.



(a) (b) (c)

Figure 5.7: Feature map of LBP-Lac. (b) shows an example of the intermediate map, where (c) shows the Euclidean distance between all the superpixels. The circled part shows all the superpixels corresponding to skin.

$$
\begin{aligned}
LBP_{P,R} &= \sum_{p=0}^{P-1} s\left(I_p - I_{p_i}\right) 2^p \\
s(x) &= \begin{cases} 1 & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x < 0 \end{cases}
\end{aligned}
\tag{5.15}
$$

We have used a uniform rotation invariant operator $LBP^{riu2}_{(P,R)}$ for different $[P, R]$ values, to include the skin patch behavior over different scales. For this study, we have selected twelve $[P, R]$ combinations over six different scales:

$$
\left\{ \begin{array}{l} [4,1], [4,3], [8,1], [8,3], [16,2], [16,5] \\ [4,2], [4,4], [8,2], [8,4], [16,3], [16,7] \end{array} \right\}
\tag{5.16}
$$

For each combination of $[P, R]$, we get an LBP image map, $I_{[P,R]}$. Next we compute a *lacunarity map* for each LBP map. Lacunarity is a property introduced by Mandelbrot [100], particularly in terms of explaining fractals. In a general sense, lacunarity represents gaps in a binary image using a box counting algorithm. Recent work by Quan et al. [40] has shown significant performance of LBP based lacunarity value for texture classification. Following their method, we computed a LBP lacunarity vector of 44 bins for each superpixel in an image.

We transform each of the lacunarity maps (each $[P, R]$) into multiple binary images. For any pixel p_i, in the LBP map $I_{[P,R]}$, the binary image is defined as:

$$
J^k_{[P,R]}\left(p_i\right) = \begin{cases} 1 & \text{if} \quad I_{[P,R]}\left(p_i\right) = k \\ 0 & \text{otherwise} \end{cases}
\tag{5.17}
$$

where

$$k = \begin{cases} \{1,2,4\} & \text{if} \quad P = 4 \\ \{1,4,7,8\} & \text{if} \quad P = 8 \\ \{1,8,15,16\} & \text{if} \quad P = 16 \end{cases} \tag{5.18}$$

Therefore, we have in total $4 \times 3 + 4 \times 4 + 4 \times 4 = 44$ binary maps. For each superpixel sp_i and binary map $J^k_{[P,R]}$, we have calculate the average lacunarity, which is the number of 0s in the superpixel.

$$Lac^k_{[P,R]}(sp_i) = \frac{1}{N_i} \sum_{p_i \in sp_i} I(p_i = 0) \tag{5.19}$$

In the equation, $I(p_i = 0)$ is the indicator function and N_i is the number of pixels in superpixel sp_i. The 44-dimensional feature vector is used for training of skin vs. non-skin. Fig. 5.7 shows an example image along with its LBP-Lac map and Euclidian distance map between all superpixels in the feature space. The circled portion indicates all of the skin pixels. Because the skin superpixels share a close neighborhood in this 44-dimensional Euclidean space, this feature is very useful for detecting skin.

5.4 Experimental Results

We have evaluated our results on two databases: the ECU skin database [6], and the CASIA-NIR [101] face database. The ECU dataset contains images of men, women and children with different skin tones, head poses, style of facial hair, scale, occlusion and background conditions. We use this database to test the performance of our algorithm in diverse naturalistic conditions. The CASIA dataset was used to test images taken under low light conditions and near infrared (NIR). We selected 2000 images from the ECU dataset for evaluation. The face detector successfully found faces in 1693 images. The main goal of this experiment was to test if the system can learn skin texture from the selected face patches and extend to the other parts of the body (e.g. torso, limbs). We tested the images for each of the the features described in section 5.3. Fig. 5.8 presents precision-recall curves for skin detection using the ECU database. For comparison, also show results that were obtained using the standard color-based Bayesian classifier describe in [6]. It is worth noting that Phung et al. [6] have shown that results from several color spaces (RGB, YCbCr, Lab, HSV) were almost identical, and RGB-Bayes outperformed other classification algorithms including Gaussian mixture model and multilayer perceptron. The grayscale approaches, using different feature

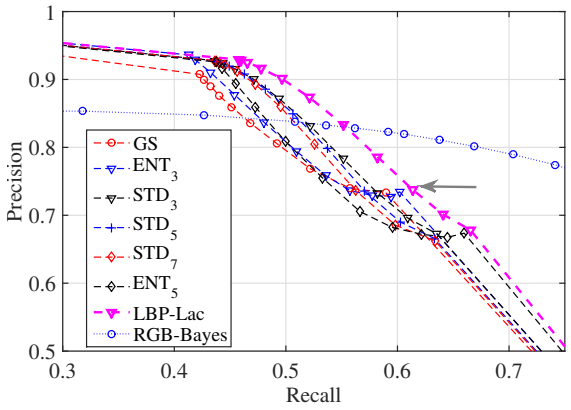


Figure 5.8: Precision-recall curves for grayscale images from the ECU database. We compared all the features described in section 5.3. We found that results from LBP-Lac is marginally better than the other feature descriptors, as indicated by the gray arrow. ($GS =$ indicates grayscale intensity feature, ENT_r and STD_r signify entropy and standard deviation map features with radius $r = \{3, 5, 7\}$ and $LBP - Lac$ indicates LBP based lacunarity)

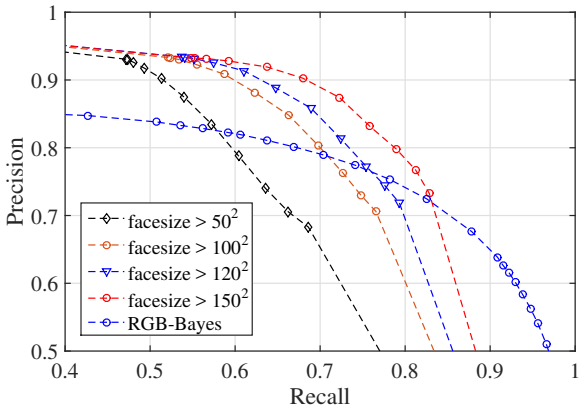


Figure 5.9: Precision-recall curves for grayscale images from the ECU database. An analysis by face size shows that our algorithm performs better for images with higher resolution. In addition, our grayscale results match state-of-the-art results from RGB-based detection methods.

types, did not produce dramatically different results. Overall the LBP-Lac feature yielded slightly better results than the rest.

Most cases where our algorithm seems to fail involve faces that are very small in the images, presumably because the skin's textural appearance is not prominent. This is logical, as the appearance of skin varies with scale, and smaller faces may fail to provide textural features of the skin in sufficient quantity. Hence, we reran our analysis for increasing face sizes. Fig. 5.9 shows a comparison of performance using LBP-Lac features with different face sizes. As the face size threshold increases, we notice that the performance of our descriptor improves as well. For faces larger than 150×150 pixels (more than 20% of test cases), we observed a precision of 78% at 81% recall. For face larger than 100×100 pixels (more than 36% of test cases), we achieved a precision of 73% at 75% recall. Table 5.1 shows comparison of F1 scores [102] for each feature type. For face sizes larger than 100×100, our method achieved performance comparable to color based methods.

Fig. 5.10 shows some sample results of skin detection from our grayscale LBP-Lac algorithm. Fig. 5.10 (a), (b),(d),(f),(g), and (j) - (m) demonstrate that the algorithm can successfully learn skin features from local cues (face), and then utilize those features to find skin in other parts of the body. The algorithm performed well in both indoor (d, m, n) and outdoor (b, g) unrestricted scenarios, with multiple faces (c, l, o), skin tones (i,j,k,l), and scales (l). The algorithm adapts to the texture of the skin, as Fig. 5.10 (d), (m), and (o) show cases involving older subjects with wrinkled skin.

Example results of our grayscale approach using the CASIA-NIR database are shown in Fig. 5.10 (p) - (s). This dataset was not designed for skin detection, and therefore no ground truth was provided. We tested 400 images, especially cases that were taken under low lighting conditions and NIR. The qualitative results are quite good. We have also tested our algorithm on grayscale video from SHRP2 example data (this was publicly available from the SHRP2 website) in Figure 5.11.

5.5 Conclusion

This work has introduced a framework for detecting human skin from grayscale images. Traditionally, automated skin detection has relied almost exclusively on color information, even though human observers are adept at using textural and contextual cues alone. To our knowledge, this is the first attempt to develop an automated approach for skin detection

Table 5.1: Comparison of F1 scores for different skin-detection methods. For reference, the last column is the conventional color-based technique [6]. The remaining columns represent different grayscale approaches that were assessed in Fig. 5.8 and Fig. 5.9. Two of the grayscale techniques with large face size thresholds (boldface) are comparable to the RGB method.

GS	ENT	STD	LBP-Lac	$> 100^2$	$> 150^2$	RGB
0.653	0.662	0.648	0.671	**0.746**	**0.795**	0.771

without the use of color. This method is adaptive to each image.

After using a face-detection system to obtain prior knowledge, our approach applies a novel texture-sampling technique to "learn" visual skin characteristics for the current image. These characteristics are then used to extend the search to other portions of the image. Our system is therefore capable of detecting skin from multiple subjects in an image, and from parts of the body beyond the face. Experimental results have demonstrated that our algorithm can achieve precision of 78% at 81% recall with high-resolution images. These results, obtained using a single monochromatic channel, are comparable to state-of-the-art algorithms that utilize three-channel color information. Extensions of the algorithm could include additional texture and context information.

(a) $P = 89, R = 84$

(b) $P = 89, R = 80$

(c) $P = 88, R = 75$

(d) $P = 90, R = 90$

(e) $P = 88, R = 92$

(f) $P = 93, R = 81$

(g) $P = 85, R = 78$

(h) $P = 96, R = 93$

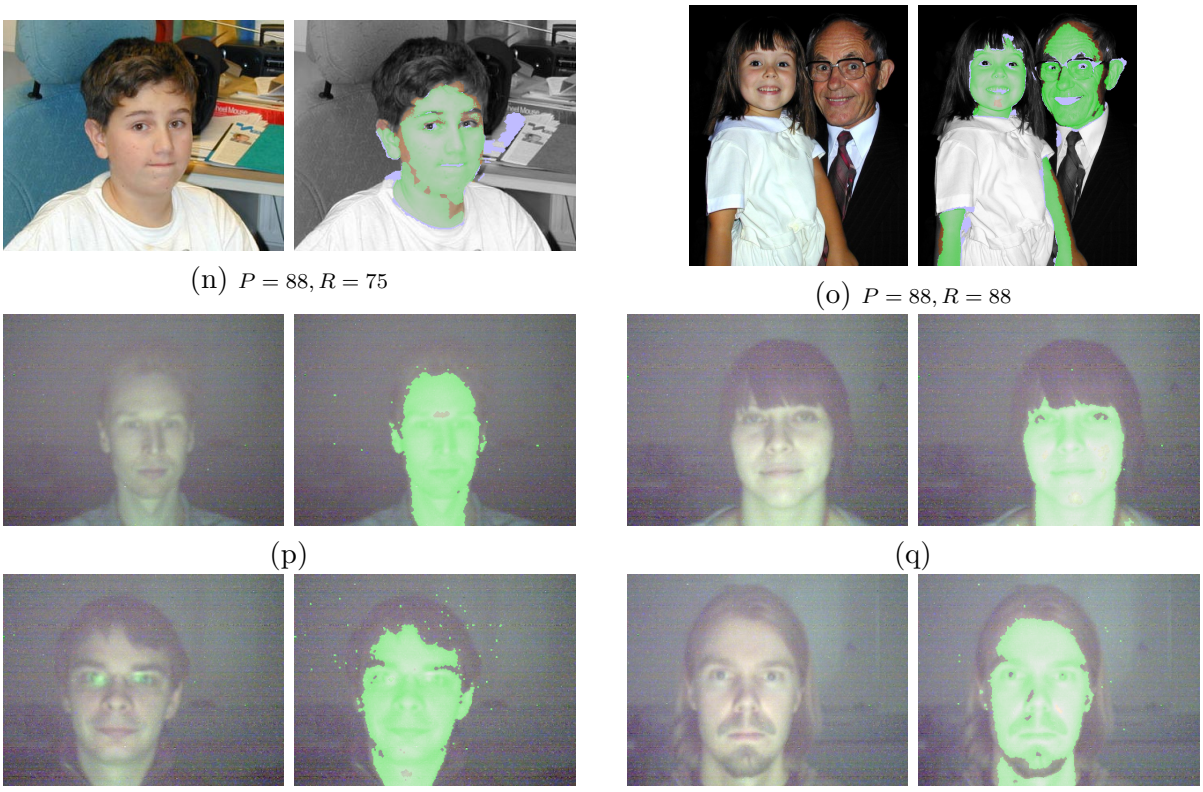
(i) $P = 85, R = 92$

(j) $P = 88, R = 80$

(k) $P = 97, R = 90$

(l) $P = 85, R = 79$

(m) $P = 87, R = 90$



(n) $P = 88, R = 75$ (o) $P = 88, R = 88$

(p) (q)

(r) (s)

Figure 5.10: Sample results from ECU database (top three rows) and CASIA database (last row) using our LBP-Lac algorithm. For each example, the left image is the original and the right image shows the result of skin detection. Although originals are shown in color, skin detection was performed using grayscale information only. For ECU results: TP = green, FP = purple, FN = original skin color, TN = grayscale. P and R signify the precision and recall values as percentages. For CASIA, only qualitative results are shown. These results demonstrate that the proposed algorithm can successfully detect disjoint skin patches from all parts of the body for different ages, skin tones, genders, and lighting conditions.

Figure 5.11: Results from the SHRP2 face video data. We tested on the publicly available video sequence for on road driver. We can successfully detect the face area.

Chapter 6

Micro-texture-based Face Anti-spoofing with Partial Occlusion

6.1 Introduction

The ability to detect and recognize faces automatically is important for several applications, including biometric authentication and access control [103, 104, 105]. Because of the popularity and convenience of face-based biometrics, the ability to detect and resist spoof attacks has also grown in importance. This work introduces a new technique for face antispoofing based on microtexture analysis of static images.

Several types of spoofing attacks have been attempted and studied for authentication systems based on face recognition. One of the most common types of attack occurs when an imposter presents a photographic print in place of a person's actual face. A photograph might be downloaded from a social media site and printed for this purpose, or a stolen photo-identify card could be used. As more advanced alternatives are available these days, the imposter could also use a video display, a 3D model of a face, or a printed mask. Although a human observer may spot the difference between a photograph and a real person, many face-recognition systems have been developed without considering the need to resist spoofing attempts [39].

Examples of simulated spoofing attempts are shown in Fig. 6.1. These were taken from the benchmark NUAA dataset [106] and Replay-Attack dataset [9]. The top row shows images of actual faces, and the bottom row shows images of printed photographs (photo attack), and a

snapshot of a video screen (video replay attack) that were presented as spoofs. In cases with the printed photo (NUAA) the distortion may be easily visible (fourth case). In either cases, characteristics features of the real faces are lost. Many existing face anti-spoofing systems are designed to identify these artifacts.

Two major directions in designing face antispoofing algorithms can be described: appearance-based and behavioral. For appearance-based models, researchers have used Fourier methods [107], texture descriptors [8] or illumination models of the face [106]. For the detection of photographic prints during presentation attacks, there are two primary considerations. First, the printing of an image is expected to introduce artifacts that may trigger the detection algorithm. Second, the more compelling consideration, is that a real face is a nonrigid 3D surface, and anti-spoofing methods can be developed that utilize shape-based analysis to distinguish an actual 3D face from the near-planar shape of a printed photograph. Tan et al. [106] and Bai et al. [108], for example, have exploited shape-from-shading reflectance models to distinguish the surface of a face from the surface of a photographic print. As expected, the specular reflection components are substantially different. Wen et al. [109] have used four appearance based features to identify distortion.

Behavioral models often rely on a sequence of images, rather than a single image. For example, Pan et al. [110] have detected the blinking pattern of the eyes as evidence of liveness. Their work makes use of the fact that humans normally blink at least once in every 2-4 seconds. They used an undirected conditional random field to track the opening and closing of the eyelid. Similarly, Kollreider et al. [111] and Bao et al. [112] have used optical flow to characterize movement of different parts of the face. They have shown how the movement of a 3D nonrigid object (real face) can be distinguished from movement of a planar surface (printed photo, or video display device). As a liveness detection methods, Li et al. [107] have suggested that vein map retrieval using ultraviolet imaging can be the most secure mode for liveness detection, but this approach is often infeasible due to the associated cost.

For both single-image and image-sequence methods, micro-texture analysis based on local binary patterns (LBP) has recently shown promising results in characterizing skin texture. Maatta et al. [8] have used three different scales of LBP features and encoded them in a concatenated histogram. In one of the LBP maps, they divided the image into nine overlapping blocks (3×3) to include the spatial dependency. Pereira et al. [113] have extended the use of LBP features into the spatio-temporal domain. They extracted LBP features from the x-y, y-t, and x-t planes, where (x, y, z) represents 3D metric space and t

refers to the temporal dimension. A summary of most of these methods can be found in [114].

In all of the previous efforts, features have been extracted from an unoccluded face in a fully frontal pose. However, in real-life scenarios, part of the face may be occluded by wearables like sunglasses, or by clothes such as scarves. Hence the question arises: Does any part of the face contribute more to anti-spoofing effectiveness than other parts? A robust system should utilize information related to occlusion, expecially if such occlusion may hamper the performance of the anti-spoofing algorithm. Therefore it is desirable to test which parts of the face provide information that is most relevant to appearance based distortion. In this work, we have experimentally evaluated the contribution of different parts of the face by creating synthetic occlusion. Following the success of appearance based models, we have restricted our scope to micro-texture analysis using a single image. We have shown, with this new feature descriptor, that static micro-texture analysis is capable of characterizing distortion as needed by anti-spoofing analysis systems.

In section 6.2 we describe a new texture descriptor for face anti-spoofing that uses a lacunarity based model on top of the LBP texture map to generate a feature vector spanning multiple scales [40]. Then in Section 6.3 we use this feature in a binary classification framework. The classifier has been trained and tested on the NUAA imposter dataset [106] and Replay-Attack database [9]. Finally, in Section 6.3.2 we present results of systematic tests that rate the relative importance of different parts of the face to determine the sensitivity of the new feature descriptor.

6.2 Micro-texture Feature Generation

We have used LBP in the previous chapter for skin detection. Here we use the similar LBP-Lac texture feature, only the computation of lacunarity is different.

6.2.1 Local Binary Patterns

LBP captures patterns of intensity changes within a local neighborhood of a given pixel (p_i). If the intensity of the given pixel is I_{p_i}, LBP first compares it to the intensity of P pixels selected from a neighborhood of radius R. For each comparison, it uses a thresholding function $s(x)$ which assigns either 1 or 0 to that neighboring pixels depending on whether

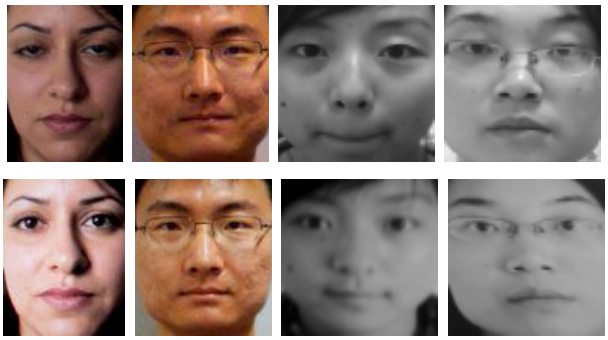


Figure 6.1: Example from the NUAA dataset. The top row shows examples of real (client) faces, and the bottom row shows images of printed photographs that are being presented as imposters from Replay attack and NUAA dataset. The first two images are particularly difficult for even human to distinguish. These images are chosen from video playback attack. The two cases at the right may be easier to spot because of blurring or geometric distortion.

their intensity (I_p) is *lower* or *higher* than I_{p_i}. Next the binary sequence is converted to its corresponding decimal value. For any P and R, the calculated value is

$$\begin{aligned} LBP_{P,R} &= \sum_{p=0}^{P-1} s\left(I_p - I_{p_i}\right) 2^p \\ s(x) &= \begin{cases} 1 & \text{for} \quad x \geq 0 \\ 0 & \text{for} \quad x < 0 \end{cases} \end{aligned} \tag{6.1}$$

LBP values can be calculated at different scales by varying R. A common practice has been to use a set of $[P, R]$ values to cover different scales and a desired neighborhood of the pixel. Ojala et al. [98] have proposed many variations of the basic LBP method. We particularly include properties of uniformity ($u2$) and rotation invariance (ri) for our experiments. Uniformity in the LBP helps to reduce the sensitivity of the feature from random noise, and rotation invariance provides universality for different testing scenarios. Hence the new LBP descriptor is denoted as $LBP^{riu2}_{P,R}$. When we compute this feature for the full image, the LBP image is denoted as $\mathbb{I}_{[P,R]}$.

6.2.2 LBP-Lacunarity

Lacunarity was introduced by Mendelbrot [100, 115] for measurement of texture in fractals. Since then it has been used for texture analysis of images in many instances (e.g., [116, 117, 118, 119]). The term derives from a Latin word meaning "lake" or "gap," and it refers to

the sizes or quantity of holes that are present in a binary image. Recently, Quan et al. [40] have shown that lacunarity derived from LBP maps yields good results in texture-based classification. Following their method, we compute local lacunarity values from each of the LBP feature images, $\mathbb{I}_{[P,R]}$. First we convert $\mathbb{I}_{[P,R]}$ to $P+2$ binary images by searching for discrete LBP values present in the feature map:

$$J_{[P,R]}(k) = \begin{cases} 1 & \text{if} \quad \mathbb{I}_{[P,R]}(p_i) = k \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

where

$$k = (0, ..., P+1) \tag{6.3}$$

Next, our approach uses a box counting method [115] to compute the overall lacunarity value for each binary image. The box counting method uses a square box of size $r \times r$, which slides over the full image to count the number of ones, at each position (n). After scanning the whole image, the procedure returns the probability for each value of n, denoted as $\mathbb{P}_r(n)$. Lacunarity for a given binary image J and for box size r is defined as the ratio of the first and the second moments of $\mathbb{P}_r(n)$:

$$\mathcal{L}(r, J) = \frac{E\left[\mathbb{P}_r^2(n)\right]}{E\left(\left[\mathbb{P}_r(n)\right]\right)^2} \tag{6.4}$$

Lacunarity computed using this method follows the power law [120], and can be rewritten in linear form [40] as follows:

$$\ln \mathcal{L}(r, J) = C_1(J) \ln r + C_2(J) \tag{6.5}$$

$C_1(J)$ and $C_2(J)$ are the parameters of the line, to be found using linear regression. We use different values of $r \in [2, ..., 14]$, and we use least-squares fitting for estimation.

Therefore, for each binary image $J_{[P,R]}(k)$, we compute two parameters that capture the distribution of LBP values in an LBP map generated with P points collected from radius R of each pixel. Once all the Lacunarity parameters are computed, they are concatenated into a single feature vector. The full approach is illustrated in Fig. 6.2.

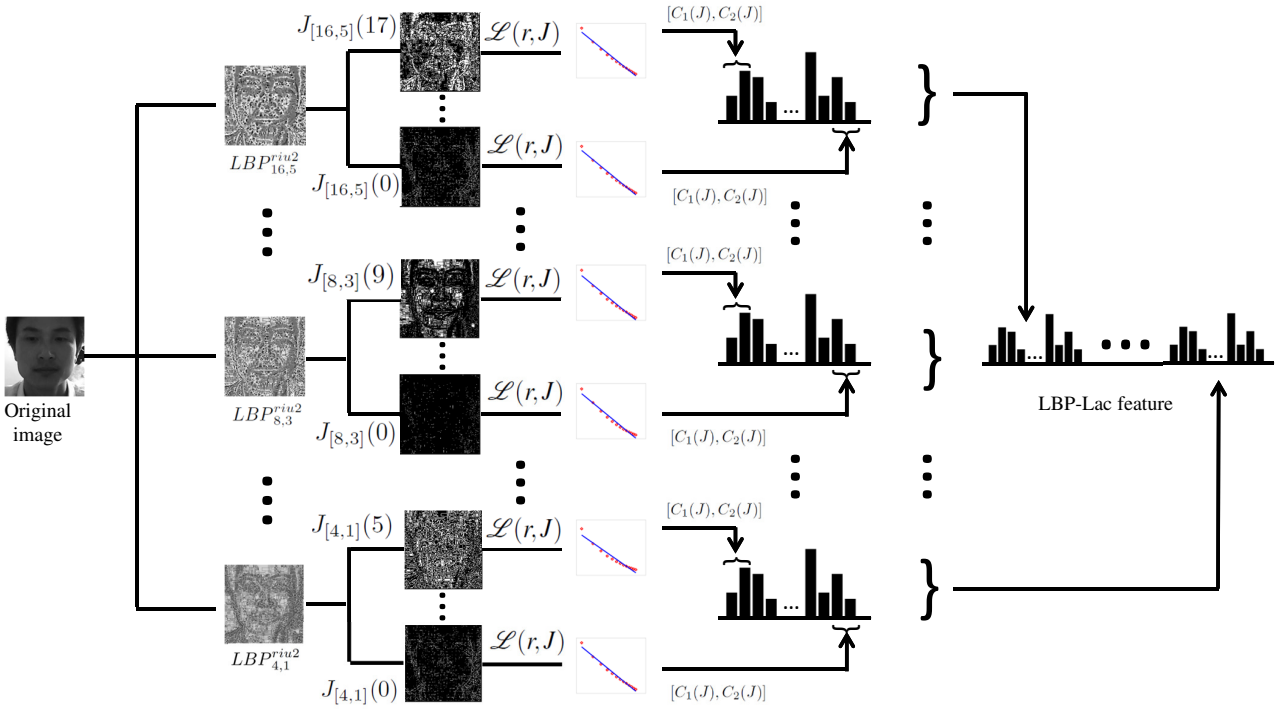


Figure 6.2: Illustration of LBP-lacunarity computation for a single image. First, different LBP maps are generated from the original image using the $[P, R]$ pairs given in (6.6). Then each LBP map is binarized using (6.2). For each binary image, lacunarity values are calculated with different box sizes ($r \times r$) using (6.4). Then least-squares mappings from r to lacunarity (blue lines) are found from the computed lacunarity values (red dots). Each binary map generates two line-fit parameters, to be used as feature values. Finally all the features are concatenated to represent the final LBP-lacunarity feature of dimension 212.

6.2.3 Feature Dimension

For each image, we have used the following set of $[P, R]$ values, which were chosen based on the convention in [98]:

$$[P, R] = \left\{ \begin{array}{c} [4,1], [4,2], [8,1], [8,2], [8,3] \\ [8,5]\, [16,2], [16,3], [16,5] \end{array} \right\} \tag{6.6}$$

Each LBP map $I_{[P,R]}$, has $P+2$ discrete values, hence, we can get total of $P+2$ binary maps Also each binary map gives two LBP-lac feature values, following (6.5). Hence for each set of $[P, R]$, we generate $2(P+2)$ LBP-lac feature values. Following all of the combinations in (6.6), as operated on the original image (using grayscale or a single color channel), the final feature dimension of the LBP-lac feature becomes $\sum_i 2(P+2) = 212$. .

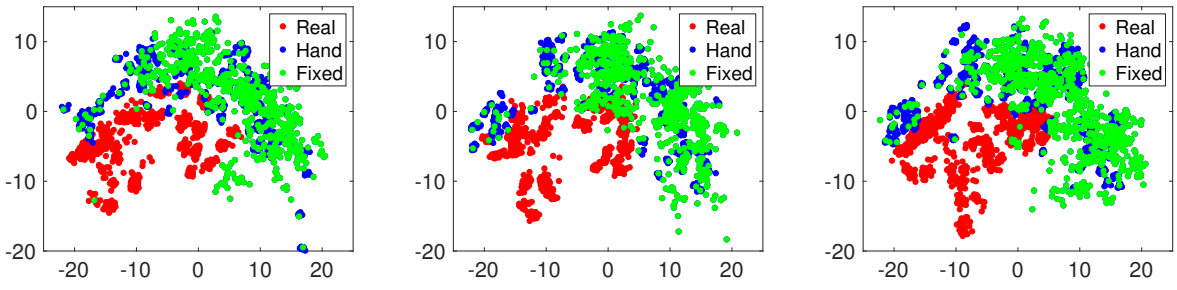
6.3 Experiment and Results

6.3.1 Full-face classification

We have used two standard databases for our experiment: NUAA dataset [106] and Replay-Attack database. The NUAA dataset consists of faces from 15 subjects in two different sets. In one set, the camera has captured the real person, and henceforth the result is called a *client* image. In the other set, a photograph of the same person is held in front of the camera to serve as a spoof attempt, and the result is called an *imposter* image. For the Replay-Attack dataset, 1300 videos are recorded for total of 50 subjects. The attack is either through a printed photo or though a recorded video of the client. The photo or video is presented to the camera either by fixing it to a stand (*Fixed*) or the imposter holds it by hand (*Hand*). Both datasets represent people of both genders, in different poses and lighting conditions. The NUAA dataset is divided into training and testing sets, and the Replay-Attack dataset has an additional development set. All cases are selected in such a way that no overlap in the sets. (If data from subject i is used during training, no images of subject i are used during testing or development).

For NUAA, we used 1743 client images and 1748 imposter images for training. For testing, we used 3362 client images and 5761 imposter images. The Replay dataset has 360 training and development video (60 for real and 300 for attack), and 480 testing videos (80 for real and 400 for attack). Because we are using an appearance based model, we select 1 image per second of video, so for a 15 second video we select 15 images. As most of the videos are static and consecutive frames does not vary too much, this sampling minimizes repetition of the feature points. We used the 212-dimensional LBP-Lac feature vector, with each feature normalized by its mean and variance of the training sample. We reduced the dimensionality of the feature vector using principal component analysis (PCA) and tested with LDA, QDA, and SVM (linear and RBF kernel). For both datasets, SVM gave better results with RBF kernel. We have tested our system with increasing number of PCA components. The first 40 PCA components gave us the best results for intra and cross-database performance evaluation.

Fig. 6.3 illustrates the separability of classes made possible by the proposed features. This figure shows a scatter plot for all of the training, development and testing data by their projection on the first two major principal components. Apart from a small fraction of the training sample, the client and imposter sets are distinct. Moreover the pattern is seen in fully disjoint sets, hence suggesting discriminative power of the LBP-Lac descriptor.



(a) Train (b) Development (c) Test

Figure 6.3: Projection of the LBP-Lac features along the first two principal components, indicated by PCA-1 and PCA-2. For all the three cases, the projection looks similar. This shows the consistency of the LBP-Lac feature.

We report the performance in terms of Half Total Error Rate ($HTER$). $HTER$ is the mean of False Acceptance Rate (FAR) and False Rejection Rate (FRR). For intra dataset experiment of NUAA dataset our experiments resulted in a classification $HTER$ of 5.8% when using SVM classifier trained with five fold cross validation. Although this result does not surpass the $HTER$ result of 4.2% reported by [9], our result is comparable as shown in Table 6.1. LBP-Lac features also provided good results when we used a linear discriminant analysis (LDA) classifier. Our results surpass previous efforts where LDA was used for classification. This observation suggests that the new LBP-Lac feature possesses relatively strong discriminative power.

For Replay-Attack dataset, we have used 40 major PCA components and the provided training dataset to train our binary classifier. The hyper-parameters for the RBF kernel was chosen using the development set. The testing result shows that we can achieve an $HTER$ of 0.3%, with $TPR = 99.7\%$ at $FAR = 0.01$. The result is summarized in comparison to other methods in Table 6.2. Our method perform better than most of the previously reported results and in particular to the HOOF based method proposed by Bharadwaj et al. [121]. The QDA based method also outperforms the result from [121].

We have also performed experiments to perform cross validation of the trained model. We train the model on the Replay-Attack dataset using an SVM with RBF kernel. When we tested the trained model on the NUAA face database, it gave a $HTER$ of 33.5% with $TPR = 99.46\%$ at $FAR = 0.01$. This result is comparable to Wen et al. [109], who have done cross dataset validation and in most of the cases, our result surpasses their general trend.

Table 6.1: Comparison of classification using LBP-Lac features with results by other researchers using the NUAA dataset. LBP-Lac seems to perform comparable with both the LDA and SVM classifier. (* indicates the reevaluation of the LBP result of [8] by [9].)
.

Method	$HTER$ (%)
$LBP^{u2}_{3\times3}$ + SVM [9]	19.0
$LBP^{u2}_{3\times3}$ + LDA [9]	18.3
LBP + SVM [8]	2.9
LBP + SVM [8]*	4.2
LBP-Lac + LDA (ours)	**8.1**
LBP-Lac + SVM (ours)	**5.8**

Table 6.2: Comparison of classification using LBP-Lac features with results by other researchers using the Replay-Attack dataset. LBP-Lac seems to perform comparable with both the QDA and SVM classifier.
.

Method	$HTER$ (%)	$HTER$ (%)
	Dev	**Test**
$LBP^{u2}_{3\times3}$ + SVM [9]	14.84	15.16
$LBP-TOP$ +SVM [113]	7.88	7.60
IDA + SVM [109]	–	7.41
HOOF + LDA [121]	0	3.75
HOOF + LDA (NN) [121]	0	1.25
LBP-Lac + QDA (ours)	4.22	**0.84**
LBP-Lac + SVM (ours)	0.7	**0.3**

6.3.2 Face area analysis

The spoofing images in the both the dataset are taken either from a printed photograph or from a screen of a mobile device. The printed photo is a two dimensional object, and so is any mobile device, hence it is expected that these images miss specific texture cues and specular characteristics that we may have from a picture of a 3D non-rigid object. But while using only a micro-textural descriptor, it is often unclear which part of the image contributes to the discriminative characteristic between classes. As we look at a typical face image (6.4a), we can easily look at different texture patterns. The hair, the skin, the eyes or the nose has specific texture and depth characteristics. Therefore, to further understand the contribution of each part, we divide the image in 9 parts as shown in Fig. 6.4a. A closer look shows that each one of the 9 blocks covers a distinct part of the face that carries vital face recognition information. To test it, we have followed similar strategy of Zeiler and Fergus [122]. Zeiler and Fergus have blocked a certain part of the image and then studied its reflection in the feature generation. This helps to understand the construction of different feature components. We formulated five scenarios and in each case we eliminated a distinct part of the face before computing the LBP-Lac feature:

- Case 1: Eyes and eyebrows are eliminated (Fig. 6.4b).
- Case 2: Part of the cheek is eliminated. This part mostly constitute plain skin pixels (Fig. 6.4c).
- Case 3: Nose is blocked (Fig. 6.4d)
- Case 4: Lower part of the cheek are blocked. This case also includes sharp edge transition of the face, hence the perception of depth relative to the background. Also this part has varying surface normal (Fig. 6.4e).
- Case 5: Mouth was blocked (Fig. 6.4f).

In this formulation we have assumed that the face is symmetric. For each of the five cases, we synthetically made the grayscale values of the area of interest equal to zero. We also eliminate this area from calculation of the binary image and lacunarity value. Hence, for a choice of blocked area $\mathcal{B}$ (6.2) and (6.4) modify to :

$$J_{[P,R]}(k) = \begin{cases} 1 & \text{if} \quad \mathbb{I}_{[P,R]}(p_i) = k \\ 0 & \text{otherwise} \end{cases} \qquad \forall p_i \notin \mathcal{B} \tag{6.7}$$

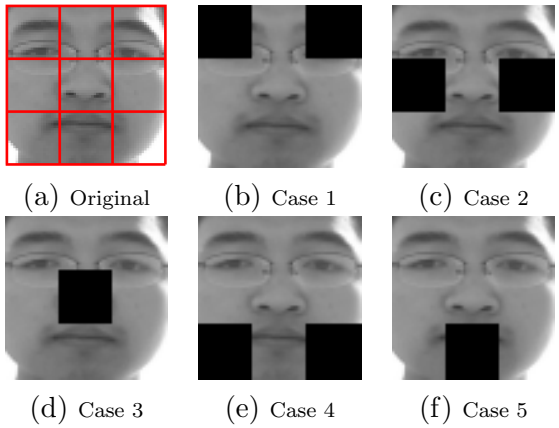
$$\mathcal{L}(r, J) = \frac{E\left[\mathbb{P}_r^2(n)\right]}{E\left(\left[\mathbb{P}_r(n)\right]\right)^2} \quad \forall p_i \notin \mathcal{B} \tag{6.8}$$

Next we train the SVM with the newly generated feature vector $LBP_{Lac}(\mathcal{B})$ for all the training images. During testing we use the original images (i.e. without the blocking). We perform this for both datasets. Table 6.4 shows results for each of the cases. We compare the cases in terms of FAR, FRR and $HTER$. For case 2 and case 4 the performance degrades significantly. The effect is more prominent for NUAA dataset. The accuracy decreases at least 10%. Fig 6.5 shows the comparison of the HTER for all the cases. Clearly case 2 and case 4 stand out with 18.67% and 21.13% error rate for NUAA. For Replay-Attack dataset, the errors are rather uniform except for case 3, where the FRR is 62.6%. It is noticeable that the FRR increases significantly for all the test cases. We further tested the Replay-Attack dataset with a counter experiment. Here we keep the classifier trained on the original dataset. We assume that a good dataset was used for training purposes. However, during testing we used the occluded version. Table 6.5 shows the results. The false acceptance rate increases by 2% on average.

Hence we can conclude that LBP-Lac feature extracted from different areas of the face plays a pivotal role. So, in case the training or testing data includes artificial occlusions like glasses, facial hair, they can significantly effect the classifier and the detection results subsequently. As this experiment reveals the sensitivity of different parts of the face anti-spoofing countermeasure, it also shows the limitation of the LBP-Lac feature vector in certain application scenarios. In real life scenarios, if the test condition experience occlusion similar to the cases presented here, the performance of the classifier may degrade.

6.4 Conclusions

This work has addressed the problem of face anti-spoofing using a single image. We have introduced a novel micro-texture descriptor, known as LBP-lacunarity, that has proven effective at distinguishing genuine faces from photographic prints and video replay of faces, which represent the common form of presentation attack. Experimental evaluation on two standard datasets have shown that imposter detection using LBP-Lac features is effective when using two different classification methods (SVM and LDA). For the case of a support vector machine, our technique resulted in half total error rates ($HTER$) of 5.76% for



(a) Original (b) Case 1 (c) Case 2

(d) Case 3 (e) Case 4 (f) Case 5

Figure 6.4: Six cases where different part of the face were blocked before deriving the LBP-Lac feature. (a) Original image: no part were blocked (b) Case 1: both the eyes and eyebrows were blocked (c) Case 2: Part of the cheek was blocked. This part mostly constitute plain skin pixels. (d) Case 3: Nose was blocked. (e) Case 4: Lower part of the cheek were blocked. This case also includes edge of face, hence depth perception with background. (f) Case 5: Mouth was blocked.

Table 6.3: Comparison of FAR (%), FRR (%) and $HTER$ (%) for five cases as shown in Fig. 6.4 fro NUAA. The boldfaced numbers in the HTER column indicate the cases with maximum deviation from the original accuracy: Case 2 and case 4. These cases mostly includes skin pixel with different surface normal and depth cue from background.

Cases	**FAR**	**FRR**	**HTER**
Orig.	7.60	3.96	5.78
Case 1	14.72	4.19	9.46
Case 2	9.58	27.75	**18.67**
Case 3	9.25	5.92	7.59
Case 4	7.71	34.60	**21.13**
Case 5	11.8	6.01	8.91

Table 6.4: Comparison of $FAR(\%)$, FRR (%) and $HTER$ (%) for five cases as shown in Fig. 6.4 for Replay Attack dataset. The boldfaced numbers in the HTER column indicate the cases with maximum deviation from the original accuracy: Case 2 and case 4. These cases mostly includes skin pixel with different surface normal and depth cue from background.

Cases	**FAR**	**FRR**	**HTER**
Orig.	0.30	0.25	0.28
Case 1	15.1	52.3	33.74
Case 2	14.9	56.5	**35.71**
Case 3	14.2	62.6	**38.43**
Case 4	14.8	51.9	33.34
Case 5	15	53.7	34.37

Table 6.5: Comparison of $FAR(\%)$, $FRR(\%)$ and $HTER$ (%) for five cases as shown in Fig. 6.4 for Replay Attack dataset. Here we train the classifier with original dataset and then test it on the occluded dataset.

Cases	**FAR**	**FRR**	**HTER**
Orig.	0.30	0.25	0.28
Case 1	2.12	0.08	1.11
Case 2	2.75	0.6	1.68
Case 3	1.63	0.51	1.06
Case 4	1.64	0.51	1.08
Case 5	2.78	0.51	1.64

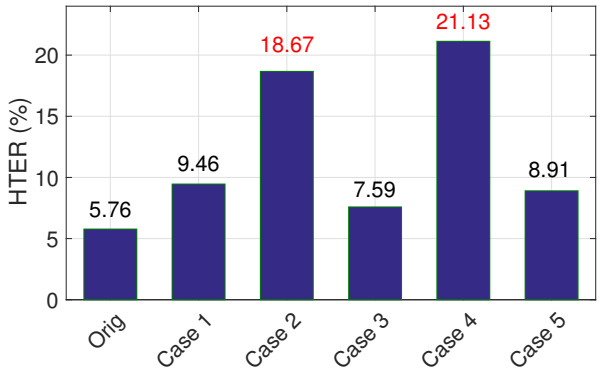


Figure 6.5: Comparison of different cases using HTER as the performance measurement. When either the upper or lower part of the cheek (case 2 and case 4 respectively) is eliminated from computing the feature descriptor, the performance of the classifier degrades.

the NUAA dataset, and 0.3% for the Replay-Attack dataset. To ensure that the technique will generalize, it has also been trained on Replay attack dataset and tested on the NUAA dataset.

This work has also investigated the role of partial occlusion during anti-spoofing. The motivation has been to determine whether different portions of the face contribute different levels of information toward imposter detection, and test the robustness of micro-texture based methods. We observed that false-rejection rates and false-acceptance rates increased substantially when some of the areas are occluded, particularly when the upper or lower cheeks and nose were obscured. In summary, this chapter has provided new insights into single-image anti-spoofing of faces that successfully detect video and photo based spoof attacks. The proposed micro-textural approach can be combined with other approaches, including behavioral methods.

Chapter 7

Conclusion

7.1 Contributions

In this dissertation, we have explored the scope of using cardiac signals in day-to-day life. Cardiac signals are widely used in medical facilities. Unfortunately, most of the equipments are costly and unsuitable for transportation. This has restricted access of basic health monitoring from remote areas and underprivileged societies. As an alternative we have investigated video based rPPG for non-contact measurement of heart rate. We have also looked into other applications of cardiac signals apart from medical applications. This includes biometrics, affective computing, and face anti-spoofing. We summarize our main contributions as follows:

- We have shown that using a dynamical model of cardiac system, ECG and PPG can be used as biometric modalities. Our result shows that the PPG based authentication system is fast and can be used in mobile devices. With increasing use of biometrics in the devices connected through internet, cardiac biometrics will help to reduce the threat of identity theft. The presented model gives a theoretical framework for cardiac biometrics which may help to analyze the problem from a theoretical point of view.
- We have discussed the possibilities and challenges for measurement of heart rate using a standard RGB camera. We have shown that under static lighting condition, the heart rate variability in time domain and frequency domain may be preserved. We have also shown that the instantaneous heart rate can be estimated with good precision even when the ambient light is not static. If the change in illumination happens with a

frequency that does not interfere with the frequency of interest of the heart rate, the heart rate can be measured with good accuracy.

- Remote PPG can be used as a countermeasure to detect face spoof attack. As a precursor to the application, we have demonstrated how traditional texture based methods may fail in case of occlusion. We have also tested a state of the art texture descriptor to understand the behavior of the spoof attacks in a multi-dimensional feature space.

- Skin detection is an integral part of the video based rPPG algorithm. In Chapter 5, we have developed a novel skin detection algorithm which shows that use of contextual information helps us to locate skin pixels in an image without using any color information. To the best of our knowledge, this is first work to report a skin detection algorithm without using any color cues. This work has significantly extended the scope of skin detection to any type of image. The proposed algorithm can be used for RGB images, NIR images, grayscale images, and images taken under low light. Due to the universality of the algorithm, it can work for cases where the skin is illuminated with artificial lights, or any of the color channels have been digitally altered.

7.2 Future Scope of Work

In this section we discuss possible extension of the work presented in this dissertation . We also discuss the research questions that have emerged from our research.

7.2.1 Large scale authentication of cardiac biometrics

In Chapter 2 and Chapter 3, we have shown that ECG and PPG hold good potential as biometric modalities. But to prove universality of those modalities, we need to test on a larger dataset spanning over different sessions and different days. Unfortunately, these type of public dataset are not available to the best of our knowledge. The Physionet [5] dataset comprises multiple datasets with ECG data from different subjects, but for most of them the signal resolution and quality are not suitable for large scale authentication. Recently, Wahabi et al. [123] has released a large scale ECG database. This has ECG recordings from 1020 subjects each recording five sessions. A supporting dataset from the same research group in University of Toronto ([124]), has ECG data from 52 participants over one month

span. For each session the ECG of the participants are recorded in 'rest' or on a treadmill. These dataset may be useful for future research. For PPG, unfortunately, no such dataset exists to the best of our knowledge.

7.2.2 PPG-based authentication within a multimodal framework.

Over the past two decades, researchers have shown that multimodal biometrics can play a pivotal role in improving authentication and identification in challenging scenarios (e.g., [42, 125]). Multimodal biometric refers to the case where templates are obtained from two or more modalities for combined analysis. An example of such an case could be use of fingerprint and face biometric. Ideally, the different modalities are able to complement the drawbacks and limitations of each other, and they allow for decision-level fusion (e.g. [126]) to improve overall performance. However, most of today's biometric authentication systems are unimodal, relying on information for a single biometric trait obtained from a single source. One reason for unimodal systems is *convenience*; the gains that result from additional biometric modalities have to be weighed against any added inconvenience to the user. An attractive aspect of the contact and noncontact PPG-based modalities being developed in this project is that they are well suited for use within a multimodal authentication framework.

In Chapter 3, we have shown that finger based PPG signal is a reliable modality for biometric authentication. In view of the presented work along with promises of rPPG presented in Chapter 4, two particular multimodal sensor arrangements are possible. For each case, *biometric templates of different modalities are obtained from a single part of the body:*

- Fingerprint recognition combined with contact PPG biometrics (from the fingertip).
- Face recognition combined with remote PPG biometrics (from the face).

Because each of these multimodal approaches depends on just one part of the body, they can provide convenience to the user (in principle) along with improved performance that is expected from two biometric modalities.

7.2.3 Illumination Correction for rPPG

Remote PPG using face video has shown promising results over the last decade. But, there are many unanswered question that still remains to answer. We list the most common

questions that we have found most relevant during the course of this dissertation:

- The signal strength is a function of the surface normal and the intensity of the incident light. None of the previous work has theoretically or experimentally evaluated the effect of skin patches with different surface normals. It will be interesting to see how surface normal effects the SNR of a patch.
- The major problem for video based rPPG has been the change in illumination. Li et al. [72]has shown a model to correct the ambient illumination using a bi-linear filter. They have modeled the effect of illumination as a linear time domain additive component to the PPG signal. The actual effect of illumination is highly non-linear and therefore the model presented in [72] may not be generalized in diverse conditions.

7.2.4 Skin Detection

In Chapter 5, we have shown that use of contextual information helps us to locate skin pixels in an image without using any color cues. This work is mainly devoted to extract texture features from single grayscale channel of the image. While this is useful for images and videos that are available only in grayscale, we can extend this algorithm to multi-channel texture descriptors where more than one channel is available. We can use contextual information to learn local texture descriptor in the image and extend it to find skin pixels in the whole image. We shall still maintain the universality of the algorithm but use information from all channels to increase recall.

7.2.5 Face Anti-spoofing

As discussed in Chapter 6 biometric authentication using face recognition is vulnerable to spoof attacks. photographs, videos and 3D masks can be use to fool a face authentication system. A Temporal features combined with texture descriptors have been proved to be effective in detecting spoof attacks. But, as we discussed in Chapter 6, face occlusion might significantly fail texture based systems. Liu et al. [127] have shown that rPPG is an effective measure for 3D mask based face anti-spoofing. It will be interesting to see how rPPG performs for other mode of attacks.

Bibliography

[1] Y. Lin, H. Leng, G. Yang, and H. Cai. An intelligent noninvasive sensor for driver pulse wave measurement. *IEEE Sensors Journal*, 7, 2007.

[2] M.-Z. Poh, N. C. Swenson, and R. W. Picard. Motion-tolerant magnetic earring sensor and wireless earpiece for wearable photoplethysmography. *IEEE Trans. on Information Technology in Biomedicine*, 14(3):786–794, 2010.

[3] M.-Z. Poh, D. J. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. on Biomedical Engineering*, 58(1):7–11, 2011.

[4] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH)*, 31(4), 2012.

[5] G. B. Moody and R. G. Mark. The impact of the MIT-BIH arrhythmia database. *Engineering in Medicine and Biology Magazine*, 20:45–50, 2001.

[6] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai. Skin segmentation using color pixel classification: analysis and comparison. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(1):148–154, 2005.

[7] Joao Paulo Brognoni Casati, Diego Rafael Moraes, and Evandro Luis Linhari Rodrigues. SFA: A human skin image database based on FERET and AR facial images. In *Proceedings: IX Workshop de Visao Computacional*, 2013.

[8] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), International Joint Conference on*, pages 1–7. IEEE, 2011.

[9] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, 2012.

[10] Alexandros Pantelopoulos and Nikolaos G Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):1–12, 2010.

[11] Kevin Patrick, William G Griswold, Fred Raab, and Stephen S Intille. Health and the mobile phone. *American Journal of Preventive Medicine*, 35(2):177, 2008.

[12] Changzhi Li, Victor M Lubecke, Olga Boric-Lubecke, and Jenshan Lin. A review on recent advances in doppler radar sensors for noncontact healthcare monitoring. *IEEE Transactions on Microwave Theory and Techniques*, 61(5):2046–2060, 2013.

[13] Rafael Gonzalez-Landaeta, Oscar Casas, and Ramon Pallas-Areny. Heart rate detection from plantar bioimpedance measurements. *IEEE Transactions on Biomedical Engineering*, 55(3):1163–1167, 2008.

[14] Lorenzo Scalise, Ilaria Ercoli, Paolo Marchionni, and Enrico Primo Tomasini. Measurement of respiration rate in preterm infants by laser doppler vibrometry. In *Medical Measurements and Applications Proceedings (MeMeA), IEEE International Workshop on*, pages 657–661. IEEE, 2011.

[15] Sergey Y Chekmenev, Aly A Farag, and Edward A Essock. Multiresolution approach for non-contact measurements of arterial pulse using thermal imaging. In *Computer Vision and Pattern Recognition Workshop*, pages 129–129. IEEE, 2006.

[16] Jure Kranjec, S Beguš, G Geršak, and J Drnovšek. Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, 13:102–112, 2014.

[17] S. A. Israel and J. M. Irvine. Heartbeat biometrics: a sensing system perspective. *International Journal of Cognitive Biometrics*, 1(1):39–65, 2012.

[18] Fitbit, Inc. Web page: https://www.fitbit.com. Accessed Nov. 2015.

[19] G. Lin, T. Nakajima, P. Rahul, and A. Hodge. *Seamlessly embedded heart rate monitor*. U.S. Patent 8,615,290.

[20] AliveCor, Inc. Web page: https://www.alivecor.com. Accessed Nov. 2015.

[21] Nymi, Inc. Web page: https://www.nymi.com. Accessed Nov. 2015.

[22] P. K. Baheti and H. Garudadri. An ultra low power pulse oximeter sensor based on compressed sensing. In *Proc. 6th Intl. Workshop on Wearable and Implantable Body Sensor Networks*, pages 144–148, June 2009.

[23] Amy Nordrum. Popular internet of things forecast of 50 billion devices by 2020 is outdated. *IEEE Spectrum*, 18, 2016. Web Page: https://goo.gl/DgaUJg, Accessed July 2017.

[24] Aditi Roy, Nasir Memon, and Arun Ross. Masterprint: Exploring the vulnerability of partial fingerprint-based authentication systems. *IEEE Transactions on Information Forensics and Security*, 2017.

[25] Kia Kokalitcheva. *Fingerprint Spoofing Is Much Easier Than You Think*, Feb 2016. [Online] http://fortune.com/2016/02/24/fingerprint-spoofing-easy/. Retrieved in June 2017.

[26] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):864–879, 2015.

[27] Marc Garbey, Nanfei Sun, Arcangelo Merla, and Ioannis Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418–1426, 2007.

[28] R. Hoekema, G. J. H. Uijen, and A. van Oosterom. Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE Trans. on Biomedical Engineering*, 49(5):551–559, 2001.

[29] Hugo Plácido Da Silva, Ana Fred, André Lourenço, and Anil K Jain. Finger ecg signal for user authentication: Usability and performance. In *Biometrics: Theory, Applications and Systems (BTAS), IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[30] Anil Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*, volume 479. Springer Science & Business Media, 2006.

[31] Ikenna Odinaka, Po-Hsiang Lai, Alan D Kaplan, Joseph A O'Sullivan, Erik J Sirevaag, and John W Rohrbaugh. Ecg biometric recognition: A comparative analysis. *IEEE Transactions on Information Forensics and Security*, 7(6):1812–1824, 2012.

[32] Rosalind W. Picard, Elias Vyzas, and Jennifer Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, 2001.

[33] Jennifer A Healey and Rosalind W Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems*, 6(2):156–166, 2005.

[34] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006.

[35] National Highway Traffic Safety Administration et al. 2015 motor vehicle crashes: overview. *Traffic safety facts research note*, pages 1–9, 2016.

[36] B. Dolan. *Ford Research unveils heart rate monitor seats*, May 2011. [Online] http://mobihealthnews.com/11014/ford-research-unveils-heart-rate-monitor-seats/. Retrieved in June 2017.

[37] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 3430–3437, June 2013.

[38] Daniel McDuff, Sarah Gontarek, and Rosalind Picard. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 36th Annual International Conference of the IEEE*, pages 2957–2960. IEEE, 2014.

[39] Nguyen Minh Duc and Bui Quang Minh. Your face is not your password face authentication bypassing lenovo–asus–toshiba. *Black Hat Briefings*, 2009.

[40] Yuhui Quan, Yong Xu, Yuping Sun, and Yu Luo. Lacunarity analysis on image patterns for texture classification. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, pages 160–167, 2014.

[41] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[42] Arun Ross and Anil K Jain. Multimodal biometrics: An overview. In *Signal Processing Conference, 12th European*, pages 1221–1224. IEEE, 2004.

[43] Rudi Hoekema, Gérard JH Uijen, and Adriaan Van Oosterom. Geometrical aspects of the interindividual variability of multilead ecg recordings. *IEEE Transactions on Biomedical Engineering*, 48(5):551–559, 2001.

[44] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Trans. on Biomedical Engineering*, 50(3):289–294, 2003.

[45] J. Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), 2007.

[46] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.

[47] Lena Biel Lena, Ola Pettersson, Lennart Philipson, and Peter Wide. Ecg analysis: a new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*, 50(3):808–812, 2001.

[48] Steven A Israel, John M Irvine, Andrew Cheng, Mark D Wiederhold, and Brenda K Wiederhold. Ecg to identify individuals. *Pattern recognition*, 38(1):133–142, 2005.

[49] M. Kyoso and A. Uchiyama. Development of an ecg identification system. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 3721–3723. IEEE, 2001.

[50] John M Irvine and Steven A Israel. A sequential procedure for individual identity verification using ecg. *EURASIP Journal on Advances in Signal Processing*, (1):243215, 2009.

[51] H. Singh, J. S. Bhatia, and J. Kaur. Eye tracking based driver fatigue monitoring and warning system. In *India International Conference on Power Electronics (IICPE)*, pages 1–6, Jan 2011.

[52] G. Wübbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, and C. Elster. Verification of humans using the electrocardiogram. *Pattern Recognition Letters*, 28(10):1172–1175, 2007.

[53] Peter Sam Raj, Sukanya Sonowal, and Dimitrios Hatzinakos. Non-negative sparse coding based scalable access control using fingertip ecg. In *Biometrics (IJCB), IEEE International Joint Conference on*, pages 1–6. IEEE, 2014.

[54] Yongjin Wang, Foteini Agrafioti, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing*, 2008(1):1–11, 2007.

[55] Adrian DC Chan, Mohyeldin M Hamdy, Armin Badre, and Vesal Badee. Wavelet distance measure for person identification using electrocardiograms. *IEEE Transactions on Instrumentation and Measurement*, 57(2):248–253, 2008.

[56] John M Irvine, Steven A Israel, W Todd Scruggs, and William J Worek. eigenpulse: Robust human identification from cardiovascular function. *Pattern Recognition*, 41(11):3427–3435, 2008.

[57] Ikenna Odinaka, Po-Hsiang Lai, Alan D Kaplan, Joseph A O'Sullivan, Erik J Sirevaag, Sean D Kristjansson, Amanda K Sheffield, and John W Rohrbaugh. Ecg biometrics: A robust short-time frequency analysis. In *IEEE International Workshop on Information Forensics and Security*, pages 1–6. IEEE, 2010.

[58] R. Sameni, M. B. Shamsollahi, C. Jutten, and G. D. Clifford. A nonlinear Bayesian filtering framework for ECG denoising. *IEEE Trans. on Biomedical Engineering*, 54(12):2172–2185, 2007.

[59] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101:e215–e220, 2000.

[60] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, 8(1):14–25, 2012.

[61] Y. Y. Gu, Y. Zhang, and Y. T. Zhang. A novel biometric approach in human verification by photoplethysmographic signals. In *4th International IEEE EMBS Special*

Topic Conference on Information Technology Applications in Biomedicine, pages 13–14. IEEE, 2003.

[62] J. Yao, X. Sun, and Y. Wan. A pilot study on using derivatives of photoplethysmographic signals as a biometric identifier. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4576–4579. IEEE, 2007.

[63] Angelo Bonissi, R Donida Labati, Luca Perico, Roberto Sassi, Fabio Scotti, and Luca Sparagino. A preliminary study on continuous authentication methods for photoplethysmographic biometrics. In *Proceedings of the IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS)*, pages 28–33, 2013.

[64] Anthony Lee and Younghyun Kim. Photoplethysmography as a form of biometric authentication. In *SENSORS, IEEE*, pages 1–2. IEEE, 2015.

[65] A. Kavsaoğlu, K. Polat, and M. R. Bozkurt. A novel feature ranking algorithm for biometric recognition with ppg signals. *Computers in Biology and Medicine*, 49:1–14, 2014.

[66] Yongbo Wan, Xiaodong Sun, and Jianchu Yao. Design of a photoplethysmographic sensor for biometric identification. In *Control, Automation and Systems, International Conference on*, pages 1897–1900. IEEE, 2007.

[67] P. Spachos, J. Gao, and D. Hatzinakos. Feasibility study of photoplethysmographic signals for biometric identification. In *17th International Conference on Digital Signal Processing*, pages 1–5. IEEE, 2011.

[68] Ikenna Odinaka, Joseph A OSullivan, Erik J Sirevaag, and John W Rohrbaugh. Cardiovascular biometrics: combining mechanical and electrical signals. *IEEE Transactions on Information Forensics and Security*, 10(1):16–27, 2015.

[69] Daniel McDuff, Sarah Gontarek, and Rosalind W Picard. Improvements in remote cardiopulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, 61(10):2593–2601, 2014.

[70] H.-Y. Wu. *Eulerian video processing and medical applications*. Master's thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2012.

[71] Antony Lam and Yoshinori Kuno. Robust heart rate measurement from video using select random patches. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3640–3648, 2015.

[72] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4264–4271, 2014.

[73] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, pages 2879–2886, 2012.

[74] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.

[75] Noelia Rodriguez-Ibañez, Miguel A García-Gonzalez, Maria Aurora Filigrana de la Cruz, Mireya Fernández-Chimeno, and Juan Ramos-Castro. Changes in heart rate variability indexes due to drowsiness in professional drivers measured in a real environment. In *Computing in Cardiology (CinC)*, pages 913–916. IEEE, 2012.

[76] Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[77] William H Press and Saul A Teukolsky. Savitzky-golay smoothing filters. *Computers in Physics*, 4(6):669–672, 1990.

[78] Nicholas R Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.

[79] Jeffrey D Scargle. Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, 1982.

[80] Pablo Laguna, George B Moody, and Roger G Mark. Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715, 1998.

[81] Georg Lempe, Sebastian Zaunseder, Tom Wirthgen, Stephan Zipser, and Hagen Malberg. Roi selection for remote photoplethysmography. In *Bildverarbeitung für die Medizin*, pages 99–103. Springer, 2013.

[82] Douglas Chai and King N Ngan. Face segmentation using skin-color map in videophone applications. *IEEE Trans. on Circuits and Systems for Video Technology*, 9(4):551–564, 1999.

[83] Jean-Christophe Terrillon, Mahdad N Shirazi, Hideo Fukamachi, and Shigeru Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. In *Proceedings: IEEE International Conference on Automatic Face and Gesture Recognition*, pages 54–61, 2000.

[84] Margaret M Fleck, David A Forsyth, and Chris Bregler. Finding naked people. In *Proceedings: European Conference on Computer Vision*, pages 593–602, 1996.

[85] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva. A survey on pixel-based skin color detection techniques. In *Proceedings: Graphicon*, volume 3, pages 85–92, 2003.

[86] Praveen Kakumanu, Sokratis Makrogiannis, and Nikolaos Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3):1106–1122, 2007.

[87] Oana G Cula, Kristin J Dana, Frank P Murphy, and Babar K Rao. Skin texture modeling. *International Journal of Computer Vision*, 62(1-2):97–119, 2005.

[88] Mehran Fotouhi, Mohammad H Rohban, and Shohreh Kasaei. Skin detection using contourlet-based texture analysis. In *Proceedings: IEEE International Conference on Digital Telecommunications*, pages 59–64, 2009.

[89] Christophe Garcia and Georgios Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Trans. on Multimedia*, 1(3):264–277, 1999.

[90] Alexandru F Drimbarean, Peter M Corcoran, Mihai Cuic, and Vasile Buzuloiu. Image processing techniques to detect and filter objectionable images based on skin tone and shape recognition. In *IEEE International Conference on Consumer Electronics*, pages 278–279, 2001.

[91] Zhiwei Jiang, Min Yao, and Wei Jiang. Skin detection using color, texture and space information. In *Proceedings: IEEE International Conference on Fuzzy Systems and Knowledge Discovery*, volume 3, pages 366–370, 2007.

[92] Kenneth L Campbell. The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *TR News*, 282:30–35, 2012.

[93] Guillaume Charpiat, Matthias Hofmann, and Bernhard Schölkopf. Automatic image colorization via multimodal predictions. In *Proceedings: European Conference on Computer Vision*, pages 126–139, 2008.

[94] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Incremental face alignment in the wild. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, pages 1859–1866, 2014.

[95] Xuehan Xiong and Fernando Torre. Supervised descent method and its applications to face alignment. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition*, pages 532–539, 2013.

[96] Abhijit Sarkar, A Lynn Abbott, Zachary Doerzaph, and Kayla Sykes. Evaluation of video magnification for nonintrusive heart rate measurement. In *IEEE First International Conference on Control, Measurement and Instrumentation (CMI)*, pages 494–498, 2016.

[97] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[98] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[99] Michael J Jones and James M Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.

[100] Benoit B Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman, New York, 1983.

[101] Stan Z Li, Senior RuFeng Chu, ShengCai Liao, and Lun Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):627–639, 2007.

[102] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Proceedings: European Conference on Information Retrieval*, pages 345–359. Springer, 2005.

[103] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.

[104] Wenyi Zhao, Rama Chellappa, and Arvind Krishnaswamy. Discriminant analysis of principal components for face recognition. In *Automatic Face and Gesture Recognition. Proceedings. Third IEEE International Conference on*, pages 336–341. IEEE, 1998.

[105] Peter N. Belhumeur, João P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[106] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. *Computer Vision–ECCV*, pages 504–517, 2010.

[107] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K Jain. Live face detection based on the analysis of fourier spectra. In *Defense and Security*, pages 296–303. International Society for Optics and Photonics, 2004.

[108] Jiamin Bai, Tian-Tsong Ng, Xinting Gao, and Yun-Qing Shi. Is physics-based liveness detection truly possible with a single image? In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 3425–3428. IEEE, 2010.

[109] Di Wen, Hu Han, and Anil K Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

[110] Gang Pan, Zhaohui Wu, and Lin Sun. Liveness detection for face recognition. In *Recent Advances in Face Recognition*. InTech, 2008.

[111] Klaus Kollreider, Hartwig Fronthaler, and Josef Bigun. Non-intrusive liveness detection by face images. *Image and Vision Computing*, 27(3):233–244, 2009.

[112] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *Image Analysis and Signal Processing. International Conference on*, pages 233–236. IEEE, 2009.

[113] Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, (1):1–15, 2014.

[114] Ivana Chingovska, Nesli Erdogmus, André Anjos, and Sébastien Marcel. Face recognition systems under spoofing attacks. In *Face Recognition Across the Imaging Spectrum*, pages 165–194. Springer, 2016.

[115] Benoit B Mandelbrot. A fractals lacunarity, and how it can be tuned and measured. In *Fractals in Biology and Medicine*, pages 8–21. Springer, 1994.

[116] Roy E Plotnick, Robert H Gardner, and Robert V O'Neill. Lacunarity indices as measures of landscape texture. *Landscape Ecology*, 8(3):201–211, 1993.

[117] Roy E Plotnick, Robert H Gardner, William W Hargrove, Karen Prestegaard, and Martin Perlmutter. Lacunarity analysis: a general technique for the analysis of spatial patterns. *Physical Review E*, 53(5):5461, 1996.

[118] James M Keller, Susan Chen, and Richard M Crownover. Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and Image Processing*, 45(2):150–166, 1989.

[119] Soe Win Myint and Nina Lam. A study of lacunarity-based texture analysis approaches to improve urban image classification. *Computers, Environment and Urban Systems*, 29(5):501–523, 2005.

[120] Qiuming Cheng. Multifractality and spatial statistics. *Computers & Geosciences*, 25(9):949–961, 1999.

[121] Samarth Bharadwaj, Tejas I Dhamecha, Mayank Vatsa, and Richa Singh. Computationally efficient face spoofing detection with motion magnification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 105–110, 2013.

[122] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.

[123] Saeid Wahabi, Shahrzad Pouryayevali, Siddarth Hari, and Dimitrios Hatzinakos. On evaluating ecg biometric systems: session-dependence and body posture. *IEEE Transactions on Information Forensics and Security*, 9(11):2002–2013, 2014.

[124] Foteini Agrafioti, Francis M Bui, and Dimitrios Hatzinakos. Medical biometrics in mobile health monitoring. *Security and Communication Networks*, 4(5):525–539, 2011.

[125] Robert Snelick, Umut Uludag, Alan Mink, Mike Indovina, and Anil Jain. Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):450–455, 2005.

[126] Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern recognition*, 38(12):2270–2285, 2005.

[127] Siqi Liu, Pong C Yuen, Shengping Zhang, and Guoying Zhao. 3d mask face anti-spoofing with remote photoplethysmography. In *European Conference on Computer Vision*, pages 85–100. Springer, 2016.