ABSTRACT

Title of Thesis: Cardiovascular Physiological Monitoring Based on Video

Henok Gebeyehu
Master of Science, 2023

Thesis Directed by: Dr. Min Wu
Department of Electrical and Computer Engineering

Regular, continuous monitoring of the heart is advantageous to maintaining one's cardiovascular health as it enables the early detection of potentially life-threatening cardiovascular diseases. Typically, the required devices for continuous monitoring are found in a clinical setting, but recent research developments have advanced remote physiological monitoring capabilities and expanded the options for continuous monitoring from home. This thesis focuses on further extending the monitoring capabilities of consumer electronic devices to motivate the feasibility of reconstructing Electrocardiograms via a smartphone camera. First, the relationship between skin tone and remote physiological sensing is examined as variations in melanin concentrations for people of diverse skin tones can affect remote physiological sensing. In this work, a study is performed to observe the prospect of reducing the performance disparity caused by melanin differences by exploring the sites from which the physiological signal is collected. Second, the physiological signals obtained from the previous part are enhanced to improve the signal-to-noise ratio and utilized to infer ECG as parts of a novel technique that emphasizes interpretability as a guiding principle. The findings in this work have the potential to enable and promote the remote sensing of a physiological signal that is more informative than what is currently possible with remote sensing.

Cardiovascular Physiological Monitoring Based on Video

by

Henok Gebeyehu

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2023

Advisory Committee:
Professor Min Wu, Chair/Advisor
Professor Furong Huang
Professor Gang Qu

Acknowledgments

I want to express my gratitude towards my advisor, Professor Min Wu, who has helped guide my Master's Thesis work. The road to completing my thesis had many obstacles, both technical and interpersonal. But she always firmly expressed her belief in my ability to overcome them, which gave me the confidence to keep pushing when times were tough.

I would like to thank my Master's committee members, Professor Huang and Professor Qu. Their review and feedback are greatly appreciated.

I would also like to thank Anirudh Nakra, Zachary Lazri, and the rest of my fellow students in the MAST research group. Through our countless conversations, they gave valuable insights that helped me to progress my research.

Lastly, I want to say thank you to my parents. Completing my graduate degree has been a challenge unlike any I've encountered before. The time and effort it took to complete my courses and research placed incredible stress on me, and without their advice and support, none of this would be possible.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
AV	Atrioventricular
BSS	Blind-Source-Separation
cPPG	Contact-based Photoplethysmography
CVD	Cardiovascular Disease
ECG	Electrocardiogram
FFT	Fast Fourier Transform
HOG	Histogram of Oriented Gradients
HR	Heart Rate
ICA	Independent Component Analysis
LED	Light Emitting Diode
ML	Machine Learning
MSE	Mean Squared Error
PCA	Principled Component Analysis
POS	Plane-Orthogonal to the skin
PPG	Photoplethysmography
rPPG	Remote Photoplethysmography
SA	Sinoatrial
SVM	Support Vector Machine
XDJDL	Cross-Domain Joint Dictionary Learning

Chapter 1

Introduction

1.1 Motivation

For several decades, cardiovascular disease (CVD) has been the leading cause of death globally, killing tens of millions of people every year [1]. There are many intermediate risk factors that can be measured in primary care facilities before cardiovascular complications become fatal, but unfortunately, not everyone around the world has access to affordable healthcare. And even if they do, it is common for these risk factors to remain undetected for long periods of time, gradually increasing in severity and not being addressed in time until it is too late.

The medical community widely considers electrocardiograms (ECG) the gold standard for assessing a patient's cardiovascular health. ECGs measure the heart's electrical activity, and each segment of an ECG cycle relays information about activity that occurs in the heart during a cardiac cycle. Doctors rely extensively on ECGs to detect CVDs including, but not limited to, Atrial Fibrillation [2], Coronary Artery Disease [3], and Cardiac Arrest Risk [4]. These are all examples of diseases that, once identified, can be reversed with mindful actions. Additionally, studies have shown that regular, continuous monitoring of ECGs has been proven to be effective in the early detection of such threatening CVDs [5] [6].

Unfortunately, limitations in ECG monitoring devices, such as price and restrictiveness, have prevented the widespread accessibility of these signals and have consequently given rise to an active area of research where ECG waveforms are inferred from another physiological signal known as Photoplethysmography (PPG). Unlike ECG, PPG doesn't directly measure the activities of the heart. Instead, PPG measures the volumetric changes of blood in the microvascular bed of tissue located at peripheral sites of the body. Because blood flow is directly influenced by the same processes in the cardiac cycle that influence ECG signals, cardiac information can still be inferred from this signal.

Efforts in the aforementioned research area (referred to as PPG-to-ECG) intend to provide the highest quality of measurements needed for cardiovascular health assessment using more accessible methods of physiological signal acquisition. This area of research has made it possible to acquire ECG signals without requiring the expensive and restrictive equipment usually needed to collect ECG. By inferring ECG from PPG, ECG can be obtained via low-cost & easily accessible contact-based PPG (cPPG) sensors. Although cPPG sensors are far more accessible than ECG equipment, many people are still unfamiliar with the technology, and thus, have not become a commonplace household item. On the other hand, video could become a more ubiquitous form of sensor collection because significantly more people have access to a smartphone camera. Physiological monitoring with video facilitates the process of monitoring vitals outside of a standard clinical setting. With this in mind, this work aims to expand cardiovascular health monitoring in remote scenarios by showcasing the feasibility of reconstructing ECG from video.

1.2 Thesis Organization

The thesis is organized as follows. In this introductory chapter, relevant research that provides foundations of video-based physiological monitoring will be reviewed as it relates to the physiological source of the ECG & PPG signals, why the two signals are intrinsically related (and thus mappable), and established principles that motivate why PPG (and as a result, ECG) can be extracted from video. Chapters 2 and 3 contain the details of this thesis work's main contributions that focus on the steps in the pipeline necessary to reconstruct ECG from video for people of various skin tones. This work explores a well-researched method of extracting PPG remotely from video (rPPG) which is typically extracted from face pixels. In this chapter, we perform a comparison between rPPG cycles obtained from the face pixels and rPPG cycles obtained from palm pixels.

Most specifically, we wanted to understand how well rPPG from either the face or the palm aligned with cPPG. It is well understood that rPPG quality is negatively affected due to varying melanin concentrations across different skin tones. Chapter 2 explores the trade-off between collecting rPPG from the face, as is traditionally done, and from a site that has no melanin, i.e., the palm. Collecting rPPG from the palm has the benefit of protecting the privacy of individuals whose data is being collected and improving the fairness of rPPG for participants with darker skin tones. After collecting the rPPG from multiple participants' faces and palms and comparing each of their alignments with cPPG, we observed that, in our particular setup, rPPG from the face tends to have higher alignment with cPPG than rPPG from the palm. Noting that there was one participant who had their rPPG from the palm have higher alignment with cPPG, we remain optimistic that future studies might be able to identify the suitable recording conditions for rPPG from the palm.

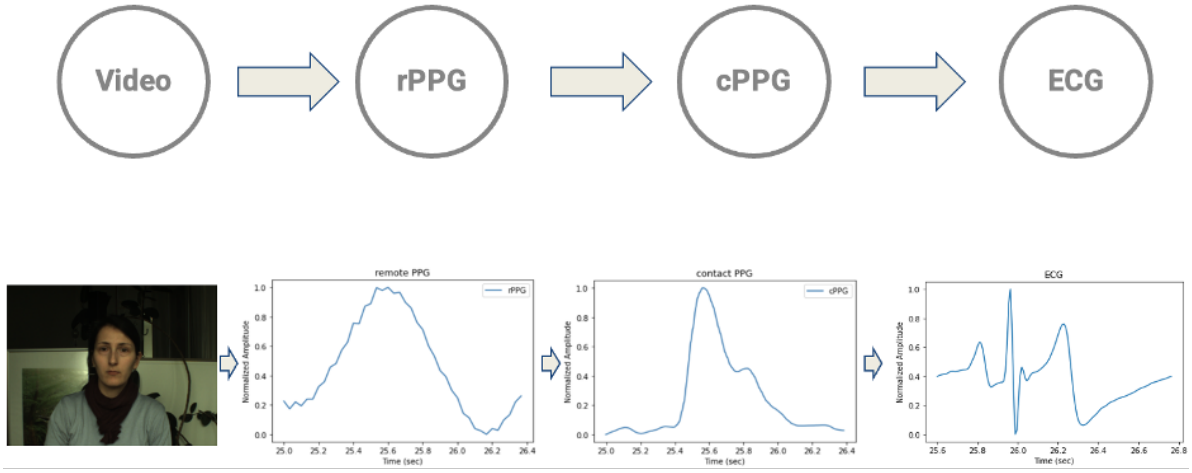


Figure 1.1: High-level outline of the proposed video to ECG technique meant to serve only as an illustrating example to highlight the key characteristics of the pipeline. The actual processing has minor alterations. Video frame was obtained from the PURE dataset [7]. Physiological signals were obtained from our internal dataset

Chapter 3 combines the building blocks of prior research to create a novel and interpretable technique for extracting ECG from video (fig. 1.1), namely: converting video to rPPG signal, mapping the rPPG signal to a cPPG signal, and finally reconstructing ECG from the mapped PPG cycles. In this section, an analysis proves the feasibility of reconstructing ECG from video with reasonably high fidelity for multiple participants of varying skin tones in both subject-specific and group models. Finally, Chapter 4 summarizes the contributions and conclusions of the thesis work and discusses future research opportunities enabled by the contributions of this thesis.

1.3 Background on the Analysis of Physiological Signals

1.3.1 Physiological Source

Typically, ECGs are collected in a hospital or ambulance. The machine used to collect ECG is non-invasive & painless and is collected by attaching up to 10 electrodes to specific parts of the body using conductive adhesives. These electrodes are used to measure the difference in electrical potential generated by the heart. Depending on the electrode placement on the body, the machine obtains a particular electrical ”view” of the heart related to the body location and is denoted by its ECG lead. A 12-lead ECG measures 12 different ”views” and is considered to be the gold standard for evaluating cardiovascular health due to the vast insights that can be obtained from analyzing the ECG waveform.

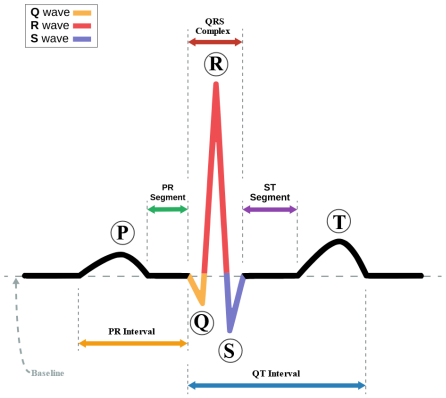


Figure 1.2: Standard Morphology of ECG Cycle. Each cycle consists of a P, Q, R, S, and T wave (from [8]).

The morphology of each ECG cycle (fig. 1.2) is directly influenced by events that occur during each cardiac electrical cycle (fig. 1.3). At the beginning of the cardiac cycle, observation of the P wave occurs when the left and right Atria fill with blood, and the Sinoatrial (SA) node spreads electrical signals throughout the Atria, causing the muscle fibers in the Atria to contract.

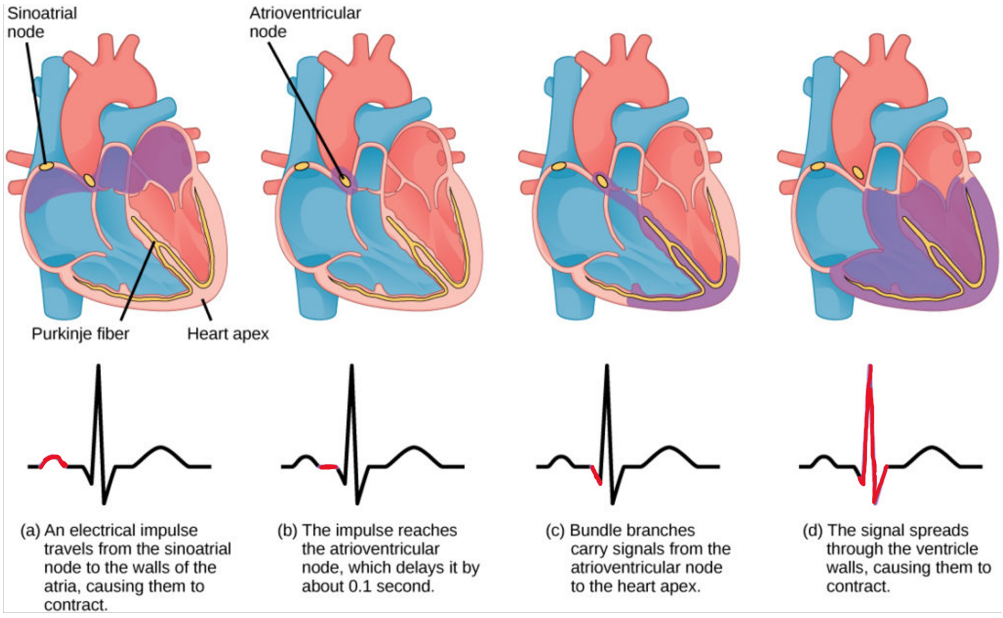


Figure 1.3: Typical Cardiac cycle (best if viewed with color; taken from [9])

The contraction of the heart muscles is referred to as depolarization. Subsequently, the relaxation of the heart muscles is referred to as repolarization. The P-Q segment represents the time it takes for signals to travel from SA node to the Atrioventricular (AV) node. The QRS complex marks the firing of the AV node and represents ventricular depolarization. In the QRS complex, the Q wave corresponds to the depolarization of the interventricular septum, the R wave represents the depolarization of the main mass in the ventricles, and the S wave represents the last phase of ventricular depolarization. Lastly, the T wave represents the relaxation and repolarization of the ventricles.

Despite being rich in cardiovascular information, current methods of collecting ECGs have several limitations that make ECGs inaccessible to the common person. ECG machines are very costly, making it common only to have access when visiting the doctor's office. As mentioned before, the machines used to collect ECGs require attaching nodes to patients, which restricts the patient's activity during collection. Moreover, the adhesive attached to the nodes can irritate the

skin and cause discomfort, especially in scenarios where the skin has taken some damage (e.g., burns). Recently, smartwatches have been able to collect single-lead ECG, but this form of ECG monitoring requires active user participation, which can be infeasible for long-term monitoring because this can be just as restrictive as traditional ECG equipment.

Similar to ECG machines, cPPG sensors are non-invasive and painless, but they differ from their counterpart in that they are far more affordable and measure a different phenomenon. cPPG sensors measure the volumetric changes in blood in the microvascular bed of the skin (typically collected at the fingertip, but can also be collected at the earlobe, forehead, and wrist [10]).

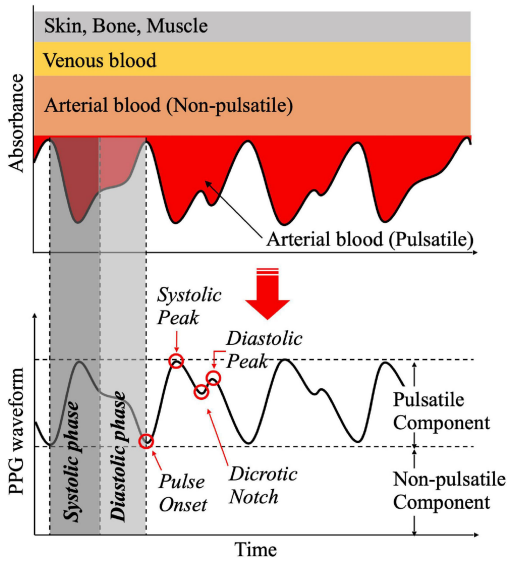


Figure 1.4: Standard Morphology of PPG Cycle. Each cycle consists of a pulse onset, systolic peak, dicrotic notch, and diastolic peak. PPG waveform is obtained by inverting the light intensity recorded by the photodetector in PPG sensor (from [11]).

There are two variants to the contact PPG sensor: a transmissive type and a reflective type. Both types consist of a light-emitting diode (LED) that irradiates light through the skin and a photodetector that captures the light emitted from the LED. The transmissive type is more common. In this mode, the LED and the photodetector are parallel to one another, with skin tissues located in between the two. In the reflective type, the LED and the photodetector are placed on the same side

of the skin, and the photodetector measures the scattered light reflected from the blood vessels. In this mode, the light intensity received by the photodetector is relatively smaller than from the transmissive type, which can cause some degradation of the PPG signal [11].

The amount of light detected by backscattering after irradiating light to the skin changes in synchronization with cardiac activity. In particular, the PPG waveform has a rising curve for an increase in capillary blood volume by cardiac contraction and a descending curve for a decrease in capillary blood volume by cardiac dilation. It is cyclical and initiated by the beating of the heart. The rising curve is denoted by the systolic phase of the PPG waveform and is often analyzed to estimate heart rate, and the descending curve is denoted by the diastolic phase. The systolic and diastolic phases are divided by the waveform region denoted by the dicrotic notch (as can be observed in fig. 1.4). Though ECG and PPG monitoring devices operate differently and measure distinct signals, the source of both processes is the same (i.e., the heart contractions).

1.3.2 Background on PPG-to-ECG

Due to the limitations of ECG monitoring devices, higher accessibility of cPPG sensors, and the relationship between PPG and ECG, recent techniques have been developed to infer ECG from PPG. The depolarization of the ventricles in the heart (captured by ECG) causes the contraction of heart muscles, which in turn causes the blood to be pumped from the heart to the rest of the body (captured by PPG). Because the electrical and mechanical activities of the heart are coupled with the dynamics of blood flow in the rest of the body, PPG and ECG represent the same cardiac process measured in different sensing modalities. Due to the intrinsic correlation between the two signals, Zhu et al. [12], have identified that ECG can be recovered from PPG. This can be viewed

as an inverse filtering process where both the low-frequency and high-frequency components of ECG can be inferred from the PPG.

A few techniques have been developed in recent years for inferring ECG from PPG. Zhu et al. [12], learn the Discrete Cosine Transform (DCT) coefficients of both PPG and ECG along with an affine transformation matrix that maps the coefficients of the PPG to the coefficients of the ECG for reconstructing the ECG signal. This approach has been shown to be effective in the subject-specific setting where there is one set of coefficients for each participant, but not very effective in learning the coefficients for a group of people. In another approach, Tian et al. [13], investigated the usage of dictionary learning to learn a mapping between PPG and ECG. They jointly learn two dictionaries and sparse representations of the two types of signals, one for the PPG and one for the ECG, and similarly to the DCT method mentioned earlier, they learn an affine transformation to map the sparse representation of a PPG to that of an ECG and use the learned ECG dictionary to reconstruct the ECG. Tang et al. [14], tackle this problem in a different direction and propose a neural network solution that can reconstruct the ECG signals without the need to segment the cycles.

1.3.3 Background on rPPG

Despite the rising popularity of cPPG sensors for health monitoring purposes like PPG-to-ECG, video can still be a more ubiquitous form of sensor collection due to the significant presence of smartphones in today's society. Strong, motivating principles enforced by optical and physiological properties show it is possible to recover rPPG signals from video, as can be observed from the dichromatic reflection model (shown in fig. 1.5).

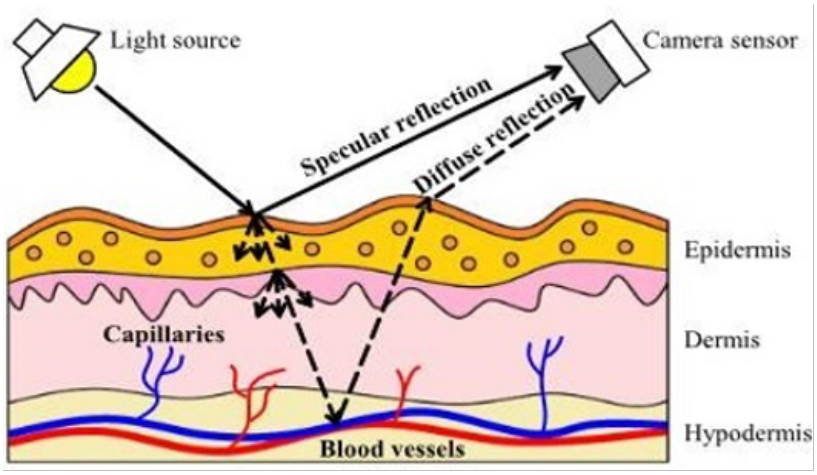


Figure 1.5: Figure capturing key characteristics in dichromatic reflection model (from [15])

In this model, as light is reflected off the surface of the skin, two types of reflections can be observed from a camera. The first type: specular reflection is simply the light that reflects directly off the skin's surface. The second type: diffuse reflection, is observed when light permeates the surface of the skin, travels through the epidermis, dermis, and hypodermis, makes contact with the blood vessels, and reflects to the camera sensor. This diffuse reflection captures information about the volumetric blood changes in the region the camera observed. As such, the signal recovered by these techniques carries meaningful pulsatile information that can be used to analyze cardiovascular health without needing any sensors other than a camera.

As blood travels to and from the skin, the camera can observe small fluctuations of color unnoticeable to the human eye. These color fluctuations are in accordance to the absorption of light in the blood vessels, so rPPG can be viewed as an extension of the reflective type of cPPG sensors mentioned before. But now, neither the light source nor the photodetector is in direct contact with the skin. Instead, the light source becomes the general lighting in the atmosphere, and the photodetector becomes the recording device located some distance away from the skin (e.g., the camera). Despite now being entirely non-contact, rPPG still measures the same physical

observations as cPPG. Because rPPG and cPPG both measure the absorption/reflection of light in the skin due to volumetric changes of the blood during the same cardiac cycle, the two are extremely correlated, but because the reflected light has to travel further for rPPG, this method of acquisition may result in higher attenuation of the signal.

The technique of extracting rPPG from video has been studied extensively over the past several years. Early works began by formulating what is captured by an RGB channel as a function of the dichromatic reflection model and treating the task of extracting rPPG as a signal demixing problem. They successfully show that using Blind-Source-Separation (BSS) techniques like Principled Component Analysis (PCA) [16] & Independent Component Analysis (ICA) [17] can extract rPPG signals that can be used to elicit heart rate information from video. But these techniques are blind, meaning they do not exploit the unique properties of the dichromatic reflection model that can be used to extract rPPG as robustly as other methods. Following these developments, researchers developed several principled approaches like Plane-Orthogonal-to-Skin (POS) [15], CHROM [18], and 2SR [19]. These approaches make fewer assumptions about the input and leverage meaningful properties about the respective channels and optical & physiological properties to extract rPPG. In recent years, several Neural Network-based approaches have been developed [20] [21] [22]. These approaches have effectively developed an end-to-end pipeline that removes the need for a lot of intermediate processing while also extracting the rPPG with high accuracy.

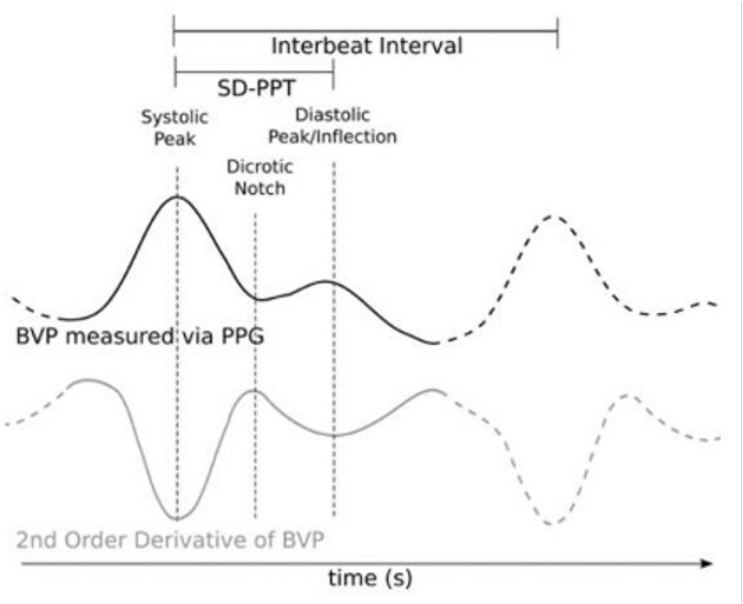


Figure 1.6: The signal on top is a Blood Volume Pulse (BVP), which is a highpass filtered PPG signal. The bottom shows the second-order derivative. Local minimas and maximas in the second order derivative reveal the location of systolic and diastolic peaks and dicrotic notch (from [23])

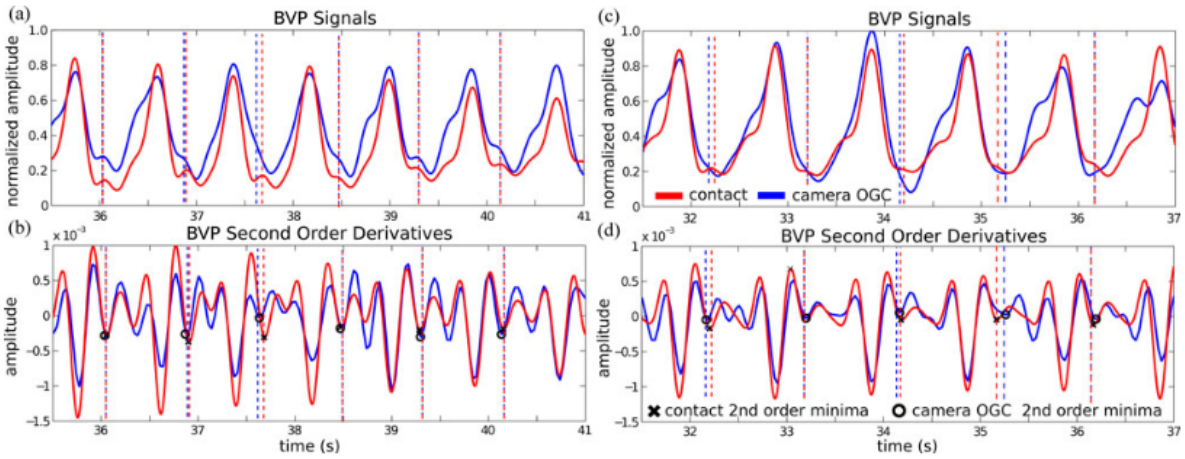


Figure 1.7: The figures on top are two segments of cPPG (red) and rPPG (blue). Cycles of rPPG on the top left have more pronounced dicrotic notch and diastolic peaks than the cycles of rPPG on the top right. Second-order derivatives of both cPPG and rPPG (bottom left and bottom right) reveal that the locations of systolic peak, dicrotic notch, and diastolic peaks are very aligned between the rPPG and cPPG cycles (from [23])

Another area of research that has recently been gaining popularity is the relationship between rPPG and cPPG. McDuff et al. [23], have shown that, after performing standard post-processing of the rPPG signal, the second order derivative of a PPG signal (fig. 1.6) can reveal the location of diastolic inflections even when the diastolic peak is not observable to the human eye in the waveform. The authors applied the second-order derivative to both rPPG and cPPG, and they found that there was high alignment between the systolic and diastolic peak locations of the two signals (fig. 1.7), which validates that morphological information is indeed present in rPPG. Since then, Kim et al. [24], took it a step further and showed that it is possible to reconstruct cPPG from rPPG. As revealed by the previously mentioned work, morphological information is present in the rPPG signal even when it might not appear to be, and with just a simple neural network, it is possible to map the processed rPPG to cPPG.

To summarize all of the background information. Despite being the gold standard for assessing cardiovascular health, limitations in ECG monitoring devices have prevented the widespread accessibility of ECG for continuous monitoring and given rise to an active area of research where ECG waveforms are inferred from cPPG. Despite the rising popularity of cPPG sensors, due to recent advancements in smartphone camera technology, video has become a more ubiquitous form of sensor collection as there are optical and physiological properties that allow rPPG to be extracted from video. Because rPPG contains morphological information and can be mapped from rPPG to cPPG, there is a strong motivation for inferring ECG from rPPG as opposed to cPPG, as done in prior works. This extension enables the reconstruction of ECG from video, as will be shown in this work.

Chapter 2

Remote PPG for Varying Skin Tones

2.1 Motivation

Neural network approaches are incredibly powerful at learning from the hidden features embedded in facial videos to extract rPPG. However, real-world lighting conditions tend to be inconsistent, which can reduce the robustness of the features extracted in a neural network. Additionally, neural networks only perform well for the particular configuration it was trained on. Any other considerations will require retraining the neural network, which can be profligately time-consuming. For these reasons, this work avoids using neural network approaches and instead utilizes a principled approach that offers more flexibility in terms of recording configurations. As mentioned previously, the reason that rPPG works is that it measures the volumetric changes of pulsatile information at the peripheral sites of the body distant from the heart. As such, there is no restriction to only collect rPPG from the face. Fundamentally, rPPG can be collected from any skin patch in which blood vessels are present and conducive for blood flow, including the palms.

It is well-understood that melanin concentrations in the skin absorb light [18] [25] [26]. Darker skin tones have higher concentrations of melanin, causing a degradation in the performance of rPPG. This work explores the feasibility of extracting rPPG from palms because the skin tone

of the palms for participants of all skin tones is more distributionally centered than the rest of the skin on the body. By processing the rPPG of the palm instead of the face, we motivate developing an rPPG technique that is fairer for all skin tones as well as respects the privacy of the participant in the video by no longer needing to process sensitive attributes such as the pixels that correspond to their face. Because principled approaches provide more flexibility than neural approaches and better heart rate estimation than BSS methods, this work uses the Plane-Orthogonal-To-The-Skin (POS) algorithm for extracting rPPG from video.

2.2 Review of Plane-Orthogonal-to-Skin

POS [15] begins by modeling the reflection of each skin pixel recorded by a camera. Based on the dichromatic reflection model, this can be defined as a time-varying function for each of the RGB channels $\mathbf{C_k}(t) \in R^3$ where C_k(t) denotes the channel information of the kth skin pixel at frame t of the video.

$$\mathbf{C_k}(t) = I(t) \cdot (\mathbf{v_s}(t) + \mathbf{v_d}(t)) + \mathbf{v_n}(t) \tag{2.1}$$

In equation 2.1, I(t) denotes the luminance or luminous intensity which quantifies the amount of light reflected from the skin ROI. I(t) is modulated by specular reflection $\mathbf{v_s}(t)$ and diffuse reflection $\mathbf{v_d}(t)$ which are both observed in the dicrhomatic reflection model. These three terms are dependent on time due to both body motion and blood volume changes that occur during the video. The final component, $\mathbf{v_n}(t)$, represents the quantization noise of the camera sensor, but this gets negated by taking a spatial average of nearby, similar skin pixels.

The authors of POS model specular reflection with equation 2.2 and diffuse reflection with equation 2.3 below.

$$\mathbf{v_s}(t) = \mathbf{u_s} \cdot (s_0 + s(t)) \tag{2.2}$$
$$\mathbf{v_d}(t) = \mathbf{u_d} \cdot d_0 + \mathbf{u_p} \cdot p(t) \tag{2.3}$$

Because specular reflection reflects off the surface of the skin in a mirror-like fashion, the authors claim that its spectral composition matches that of the light source from which it originates. Additionally, any movement (even subtle movements like breathing) will cause geometric differences between the light source and the skin surface being illuminated which causes specular reflection to be a time-varying function of the orientation that a person in a video is in. As such, $\mathbf{u_s}$, s_0, and $s(t)$ from equation 2.2 denotes the unit color vector of the light source as well as the stationary and time-varying parts of the specular reflections respectively.

Diffuse reflection carries with it information about the person's skin color - which is constant - and the pulsatile information under the skin - which is time-varying and depends on blood flow. Because blood is always flowing to and from the skin in sync with the cardiac cycle of the heart, this causes minor fluctuations in the perceived color of the diffuse reflection that is difficult for a human to perceive but easily tractable with a digital camera.

In equation 2.3, $\mathbf{u_d}$ denotes the unit color vector of the skin tissue, d_0 denotes the stationary reflection strength, $\mathbf{u_p}$ denotes the relative pulsatile strengths in RGB channels, and $p(t)$ denotes the pulse signal.

The stationary parts of the specular and diffuse formulas can be captured by:

$$\mathbf{u_c} \cdot c_0 = \mathbf{u_s} \cdot s_0 + \mathbf{u_d} \cdot d_0 \tag{2.4}$$

where $\mathbf{u_c}$ denotes the unit color vector of the skin reflection and c_0 denotes the reflection strength. Furthermore, the luminance intensity $\mathbf{I(t)}$ and the light intensity can be distributed into a stationary, I_0, and time-varying, $i(t)$, component.

$$I(t) = \mathbf{I_0} \cdot (1 + i(t)) \tag{2.5}$$

After substituting equations 2.2 - 2.5 into equation 2.1, expanding and simplifying terms, 2.1 can be approximated by equation 2.6. (Full derivation can be found in [15])

$$\mathbf{C(t)} \approx \quad \mathbf{u_c} \cdot I_0 \cdot c_0 + \mathbf{u_c} \cdot I_0 \cdot c_0 \cdot i(t) + \mathbf{u_s} \cdot I_0 \cdot s(t) + \mathbf{u_p} \cdot I_0 \cdot p(t) \tag{2.6}$$

Because C(t) represents spatially averaged pixels of similar skin, the average skin reflection color is the most dominant signal. To remove the effect of the average skin reflection color, or the dc level, the authors perform temporal normalization by solving for a diagonal normalization matrix N that can be used to temporally normalize C(t). After temporal normalization, equation 2.6 can be reduced to:

$$\mathbf{C(t)} = \overbrace{\mathbf{1} \cdot (1 + i(t))}^{\text{Light Intensity}} + \overbrace{\mathbf{N} \cdot \mathbf{u_s} \cdot I_0 \cdot s(t)}^{\text{Specular}} + \overbrace{\mathbf{N} \cdot \mathbf{u_p} \cdot I_0 \cdot p(t)}^{\text{Pulse}} \tag{2.7}$$

Extracting the pulse component from a video frame is now a signal demixing problem as stated before. Because each channel in an RGB frame contributes different amounts of information regarding blood pulsation, the authors experimentally find two projection vectors that can project the time series into a direction that maximizes pulsatility and minimizes the effect of specular reflections while also giving more weight to the color channels that capture more pulsatile

information. They also ensured that both are orthogonal to each other and the plane denoted by the average skin tone (**1**) to minimize redundancy of the projected signals.

$$\mathbf{p_p} = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix}$$

When multiplying $\mathbf{P_p} \in R^{2x3}$ to $\mathbf{C_n}(t) \in R^{3xl}$, this results in an $\mathbf{S}(t) \in R^{2xl}$. The two rows of the projected signal (S_1 and S_2) are combined together using alpha tuning $h = S_1 + \alpha \cdot S_2$. The alpha tuning has the benefit of correcting the projection depending on which of specular or pulsatile variations dominate S(t). Finally, because h(t) is estimated from short video intervals in a sliding window, the long-term pulse signal for the entire video is stitched together using overlap-adding.

By modeling the optical and physiological properties of the channel information and generating projection vectors that capitalize on the varying pulsatile strengths of different channels, POS has been shown to be more interpretable than neural-based approaches. Also, since it requires no data-driven learning, the POS algorithm can be applied to any patch of skin, unlike a neural network that would require additional model retraining, showcasing the flexibility of the POS algorithm.

2.3 Extraction of the rPPG from Video

To extract rPPG signals from the videos using POS, the pixels corresponding to either the face or hand must first be identified. For this work, face detection is performed on a per-frame basis (an example frame can be observed in 2.1) using a Histogram of Oriented Gradients (HOG) cascade face detector because it is faster at processing frames compared to other more advanced techniques. In HOG, the magnitude and angle for each pixel are computed and used to determine that pixel's gradient. Once the gradient for each pixel is available, the image is divided into blocks of pixels, and the histogram of each block's gradient is computed. The contours in a person's face don't vary significantly from person to person as opposed to other objects that may appear in a video, making it possible to feed the HOGs into a Linear Support Vector Machine (SVM) to classify if the image contains a frontal face. The human head has a lot of variances between the eyes, nose, hair, and skin, which motivates the need for a face detection algorithm.

On the other hand, the hand is relatively uniform, so when extracting rPPG from the hand, we don't use an automatic hand detector. Instead, we predefined the location where the hand must be for rPPG to be extracted. Then for both hand and face rPPG, an unsupervised skin classifier learning algorithm is trained to detect the pixels that correspond to the skin as opposed to the background (see fig. 2.2 and 2.3). These pixels are then fed into the POS algorithm for rPPG extraction.

Figure 2.1: Example frame collected from video used for recovering rPPG. The participant is sitting about .7 meters away from a smartphone camera located on a stand. This figure showcases that in this configuration, both the palm and the face are present in the same video frame making it possible to select which patch of skin pixels to pass as input to rPPG algorithm.

Figure 2.2: Example of Face Segmentation performed by HOG cascade face detector. Participant is from a video obtained from public PURE dataset

Figure 2.3: Examples of Palm Segmentation performed by custom skin learning algorithm. Participants are from internal data collection. Showcases how participants' privacy is protected because they cannot be identified from these images.

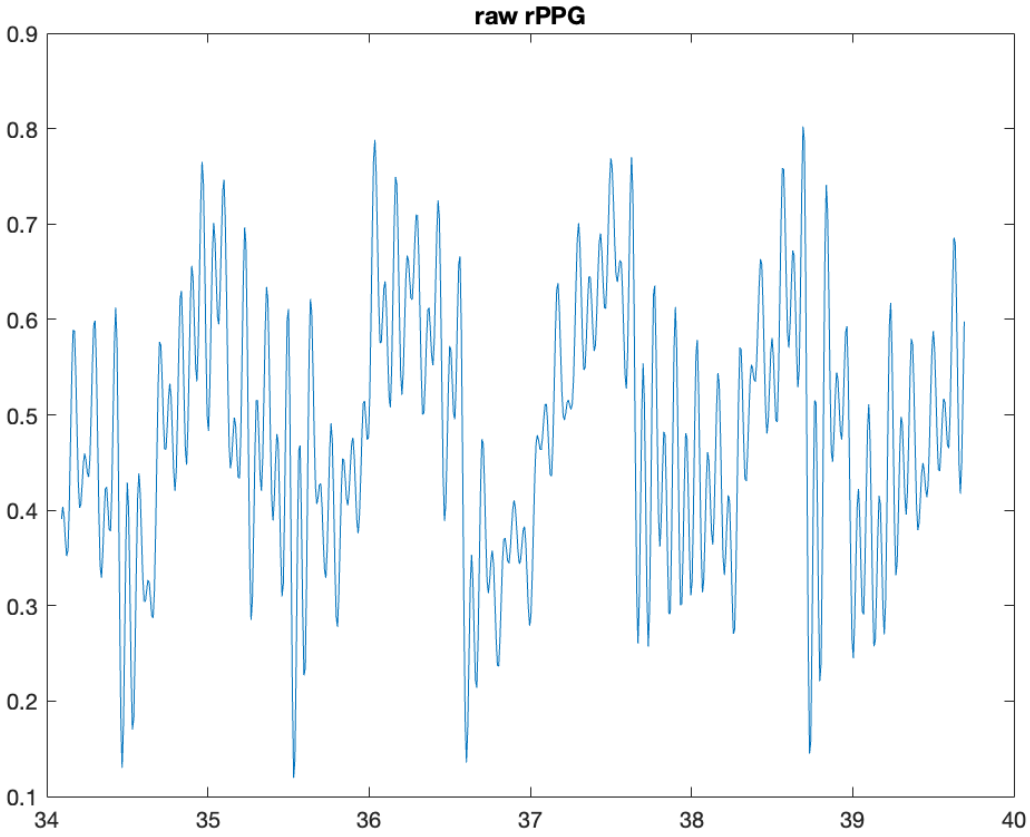


Figure 2.4: Example segment of unprocessed rPPG signal.

2.4 Post-processing Steps

After using POS to extract rPPG from video, additional processing is necessary to improve the quality of the signal such that it can be used in further downstream tasks. The initial rPPG signal, at first glance, looks quite noisy (as can be observed in fig. 2.4). This is due to a combination of motion artifacts, noise, specular reflections, and slight inconsistencies in the room's illumination at the time of recording and the face detection step that identifies the necessary pixels to pass into the POS algorithm. The post-processing steps are as follows: first, a 5-tap moving average is applied to the raw signal to smoothen it. The 5-pt window size was selected because the rPPG cycles have a sampling rate of 30 frames per second, so this minimizes information lost during smoothing. The next step is to bandpass the signal using a Butterworth bandpass filter. The literature typically bandpasses the signal between 42 - 240 bpm, as this is where the PPG signal's crucial frequency band resides. This can be verified by performing a time-frequency analysis of the filtered signal. From a time-frequency analysis, one should be able to observe a dominant trace around the heart rate frequency (as can be observed in the spectrograms in table 2.2), which is typically between 60-100 bpm in resting scenarios.

Typically at this point, the rPPG signal is in a state in which heart rate information can be processed and analyzed. This is because the systolic phase of the PPG is the most prominent segment in each cycle (as can be observed in fig. 2.5). But upon further inspection, one can see that the rPPG's morphology doesn't always match that of its contact PPG due to the presence of noise and artifacts. In many rPPG cycles, the presence of diastolic peak and dicrotic notch are noticeable but can often be challenging to read, as was observed in [23]. The reduced quality of the rPPG guided research in improving the SNR of the signal in [24]. To remove the effects of breathing, we

detrend the signal using a technique based on a smoothness priors approach [27]. Following the detrending, the Physionet toolbox [28] is used to detect onsets of the PPG signals. For each cPPG onset, it's subsequent onset is used to determine the start and end of a PPG cycle. These start and end times are then used to segment the rPPG and cPPG signals to extract aligned cycle pairs. Since the rPPG and cPPG cycle pairs have sampling rates of 30 and 1000 Hz, respectively, a cubic spline interpolation function is used to resample both signals to 50 samples per cycle. This new sampling rate is determined so that there is sufficient resolution for all physiological signals used in this work. We understand that PPG amplitude information can shed light on potential cardiovascular health concerns, but for this work, only waveform shape is being investigated. With this being the case, we standardize each cycle pair between 0 and 1 to simplify comparisons.

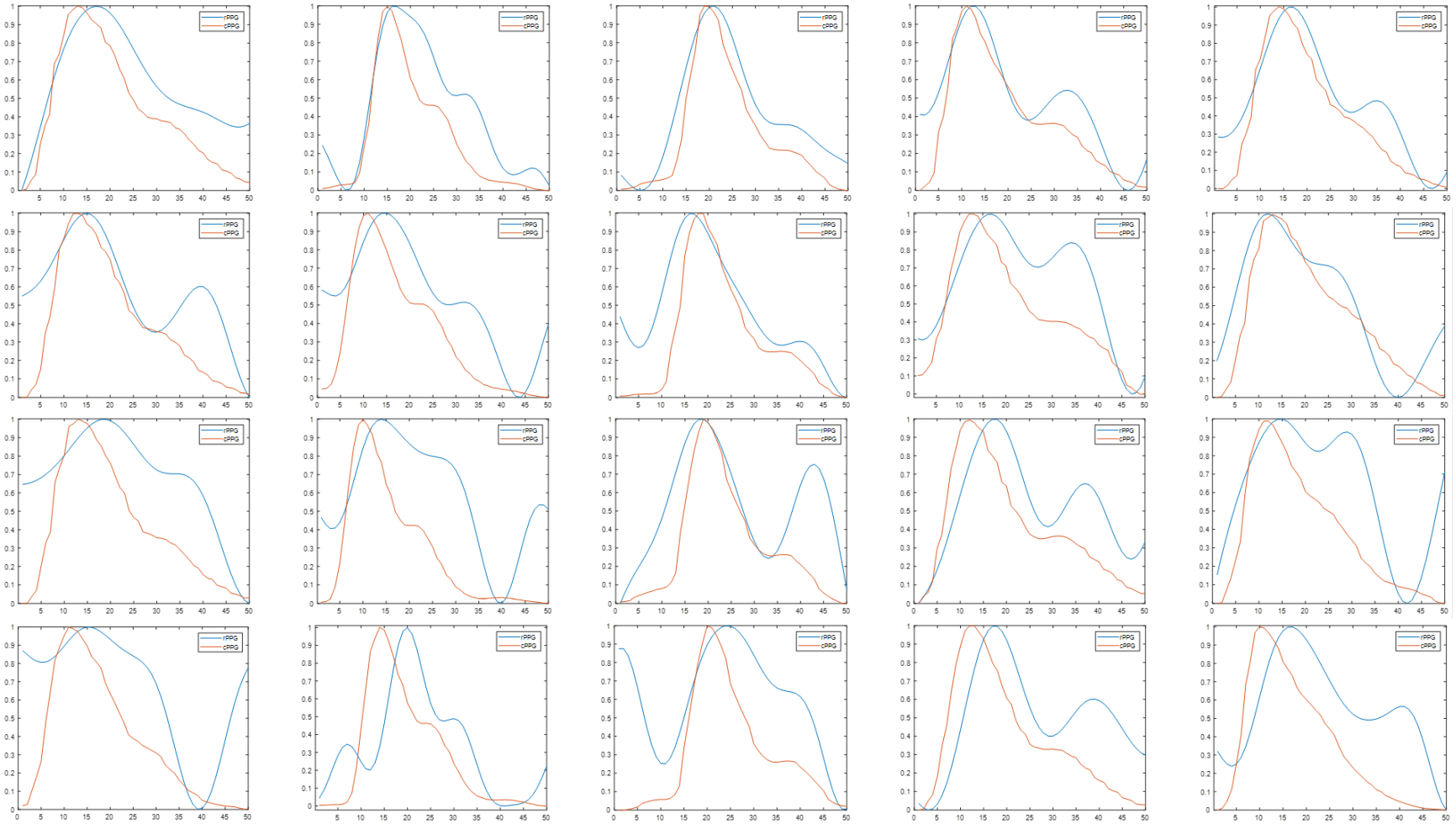


Figure 2.5: Single Cycle of rPPG and cPPG after postprocessing. Cycle pairs are extracted by detecting cPPG onsets using the Physionet toolbox and extracting both signals between pairs of onsets. All cycle pairs displayed here have Pearson correlations higher than .6. Each column corresponds to participants 1-5, from left to right. Each row contains a cycle pair with decreasing alignment

2.5 Experiment

2.5.1 Palm vs Face rPPG

As mentioned previously, the pigmentation of the skin in the palms has fewer variations in color between participants of various skin tones due to a lack of melanin in the palms. This property of the palm makes it an interesting site to investigate for the purposes of collecting rPPG. The first rigorous analysis of rPPG collected from the hands was performed recently by [29], and they investigated the differences in rPPG collected from glabrous and non-glabrous skin. They have shown that rPPG collected from glabrous skin has a higher amplitude than non-glabrous skin. Here, glabrous skin is defined as the skin on the body that lacks melanin and dermal filaments like hair. It composes approximately 10 percent of the skin on the body, including the palms of the hands and soles of the feet. In contrast, non-glabrous skin composes the other 90 percent of the skin on the body and includes the skin on the face and the back of the hands. Additionally, glabrous and non-glabrous skin have structural differences in skin blood vessel architecture that contribute to variations in blood flow between the two skin types [30]. Although glabrous and non-glabrous skin has structural and functional differences, the fact that glabrous skin contains little to no melanin suggests that it might be a more suitable location to collect rPPG from for people whose rPPG quality is low due to having a darker skin tone.

With these insights in mind, we performed a small-scale, controlled study to investigate the differences between the rPPG collected from glabrous and non-glabrous skin. But there are two critical distinctions between the study conducted in [29] and here. First, in their experiments, the non-glabrous skin was the back of the hand, whereas, in this study, the non-glabrous skin was the

face. Second, instead of comparing the amplitudes and temporal lags, this study seeks to compare the overall morphologies of the rPPG generated from skin collected at the face and the palm to identify which has a higher alignment with cPPG.

2.5.2 Dataset

At the time of writing, no public remote physiological sensing datasets contained video with both face and palms, so we collected a new dataset to perform our study. Video recordings were captured using the iPhone 11's Dual 12MP Wide & UltraWide cameras at 30 fps with a resolution of 1080p. Simultaneously, Biocapture Radio was used to extract cPPG signals that will be used as ground truth for comparison with the rPPG from face and palm. The next chapter incorporates the data collected for this study, so lead II ECG signals are also recorded but will not be used for this study. The PPG and ECG signals are recorded at a frequency of 1000 Hz.

Figure 2.6: Fitzpatrick Scale for Skin Tones (best if viewed with color; taken from [31])

Before the recordings, participants were asked to provide the Fitzpatrick Scale Skin Tone that they believe best represents the skin color of their face. Typical public remote physiological sensing datasets record videos between 1 to 5 minutes long. But one of the goals of this work is to contribute to long-term physiological monitoring, so for each participant, a minimum of ten-minute videos were recorded. Participants were asked to raise their hands and remain as still as possible during recordings, but understandably, recordings are quite long, so some slight motion

may be introduced into the recordings. Every recorded video has the participant's face and hand raised, as in fig. 2.1. Using a raised hand configuration is beneficial because it ensures that both the face and the palm are within the same frame for the entire video to enable a direct comparison. To this end, the rPPG from five participants (4 male and 1 female) of skin tone between III and V according to the Fitzpatrick Skin Tone Scale were collected for representative diversity in skin tones.

2.5.3 Metrics

Next, POS is applied to each subject's palm and face. Then following the preprocessing steps mentioned before, a cPPG cycle and its corresponding rPPG cycle can be extracted and paired together for further analysis. The goal is to identify which region between the face and the palm generates rPPG cycles that have a higher alignment with cPPG cycles. To this end, Mean absolute error (MAE), Root Mean Squared Error (RMSE), Cosine Similarity, and Pearson Correlation (ρ) are employed.

This dataset contains several thousands of rPPG & cPPG cycle pairs. MAE and RMSE are useful for evaluating the error between the two signals, and Cos Sim & ρ quantify the similarity and correlation between the signals, respectively.

Mean Absolute Error

$$MAE = \frac{\Sigma_{i=1}^{N}|x_i - y_i|}{N}$$

Cosine Similarity

$$CosSim = \frac{\Sigma_{i=1}^{N} x_i y_i}{\sqrt{\Sigma_{i=1}^{n} x_i^2}\sqrt{\Sigma_{i=1}^{n} y_i^2}}$$

Root Mean Squared Error

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{N}(x_i - y_i)^2}{N}}$$

Pearson Correlation

$$\rho = \frac{\Sigma_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i - \overline{y})^2}}$$

2.5.4 Results

Table 2.1: Average of all cycle-wise alignment of cPPG with rPPG obtained from face and palm (from multiple recordings). A higher ρ and Cos Sim indicate a higher alignment of rPPG with cPPG as does a lower RMSE. Skin tone is according to the Fitzpatrick Scale (see fig. 2.6).

Participant	**Skin**	**Num**	**rPPG from Face**			**rPPG from Palm**		
ID	Tone	Runs	ρ ($\uparrow$)	Cos Sim ($\uparrow$)	RMSE ($\downarrow$)	ρ ($\uparrow$)	Cos Sim ($\uparrow$)	RMSE ($\downarrow$)
1	3	1	0.51	0.83	0.34	0.29	0.75	0.38
2	3	4	**0.64**	**0.84**	**0.31**	0.25	0.71	0.41
3	4	2	0.26	0.70	0.42	**0.46**	**0.77**	**0.37**
4	4	1	0.38	0.78	0.36	0.27	0.75	0.39
5	5	4	0.36	0.75	0.38	0.27	0.72	0.40

As one can see from the table 2.1, the rPPG obtained from the face of participants with skin tone 3 (Participants 1 and 2) had higher alignment with cPPG than rPPG from the face of participants with skin tones 4 and 5. This confirms that darker skin tones on the face negatively affect remote physiological sensing. Furthermore, the rPPG from the palm (except for Participant 3) is generally consistent across all skin tones. This matches our expectations because there is not much variation in the skin color of the palms of the participants. Also, the rPPG from the palm of Participant 3 had significantly higher alignment with cPPG than rPPG from that participant's face, but they were the only participant to experience this.

So from this analysis, we conclude that rPPG from the face is more reliable than rPPG from the palm. However, it is essential to note that the lighting in the room where the recordings took place was not controlled for. As such, both specular reflections and undesired shadings may be introduced into the videos, which might have influenced the results of this study. Notwithstanding, the results from Participant 3 give optimism that when lighting is controlled for, and the study is expanded, it might be possible to extract rPPG from the palm more reliably.

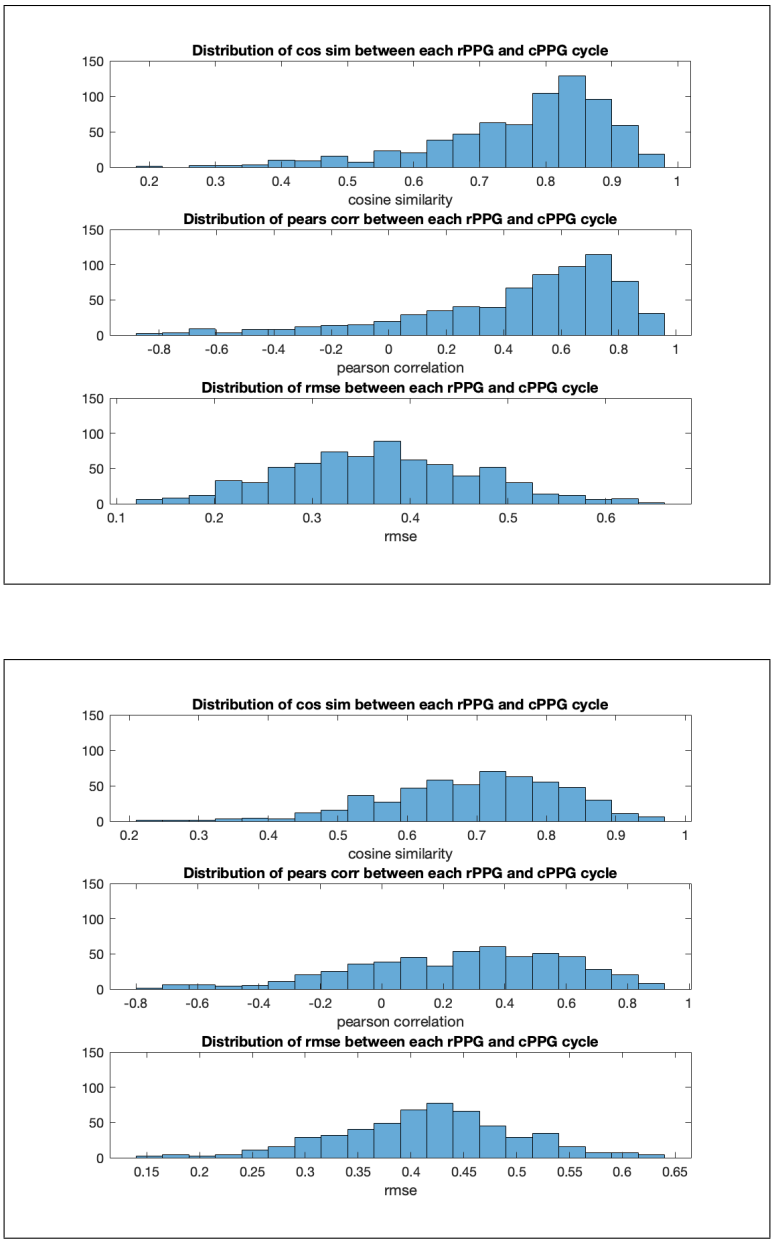


Figure 2.7: Distribution of Different Metrics for Subj II of skin tone type III. Clearly the rPPG from the face (top) has higher alignment with cPPG than does the rPPG from the hand (bottom)

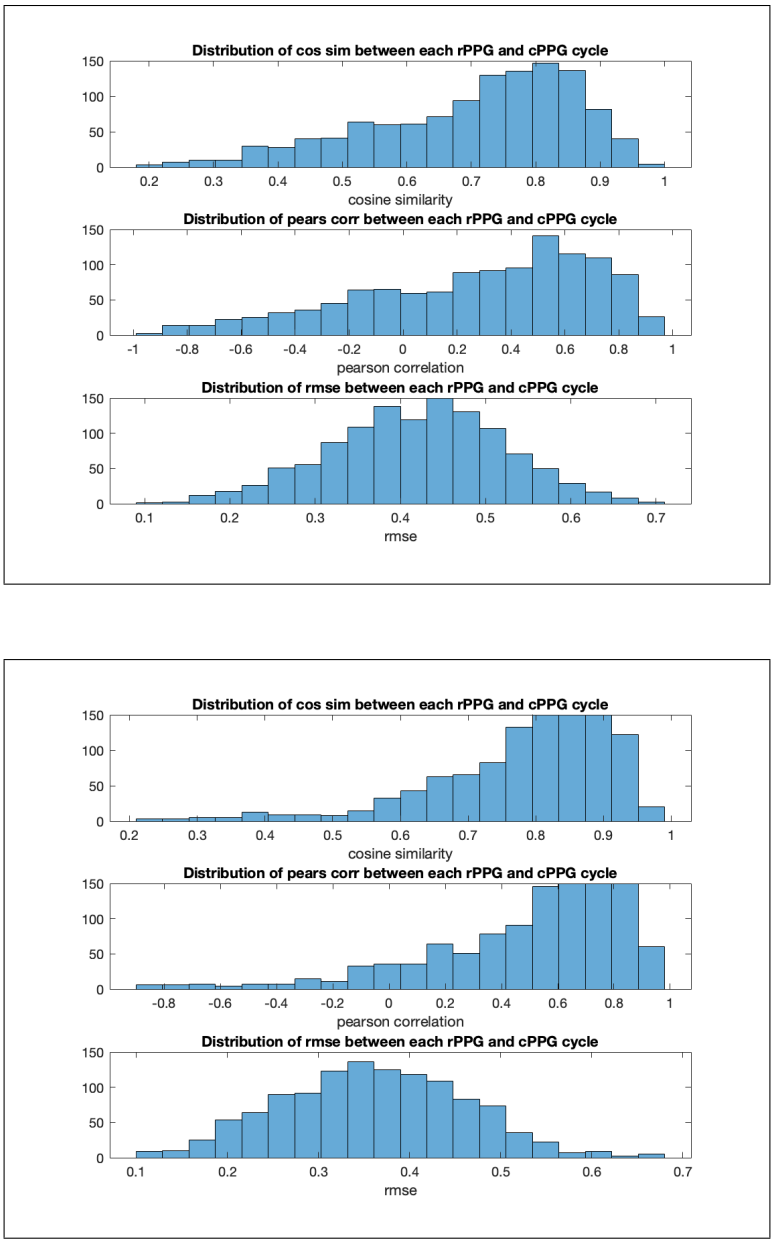


Figure 2.8: Distribution of Different Metrics for Subj III of skin tone type IV. The rPPG from the hand (bottom) has higher alignment with cPPG than does the rPPG from the face (top).

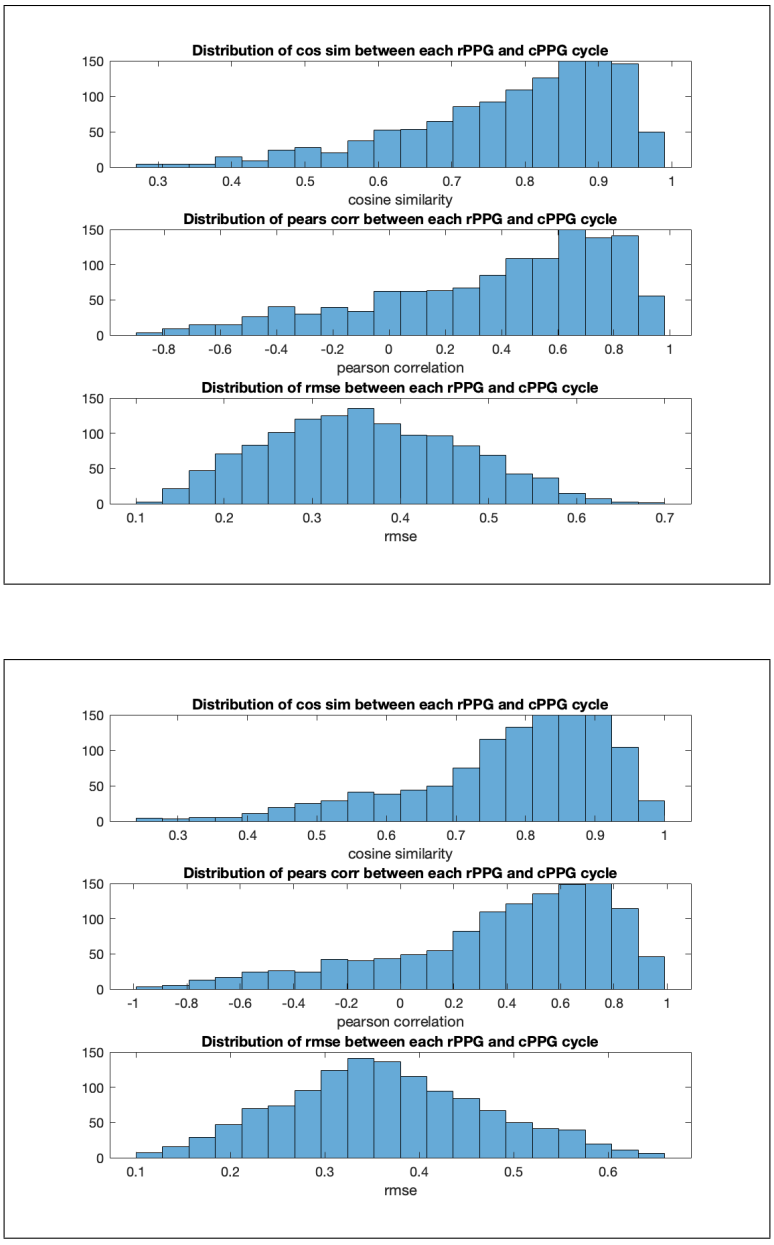


Figure 2.9: Distribution of Different Metrics for Subj V of skin tone type V. The rPPG from the face (top) has a comprable distribution to the rPPG from the hand (bottom).

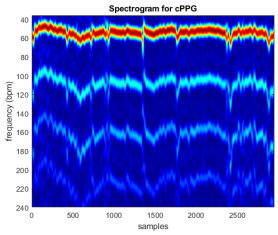
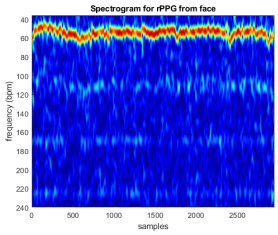
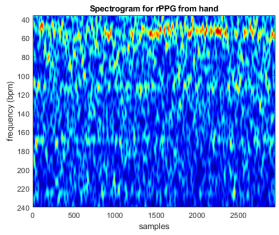
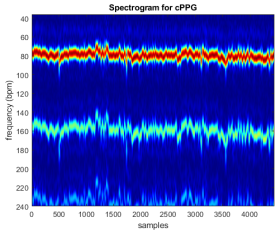
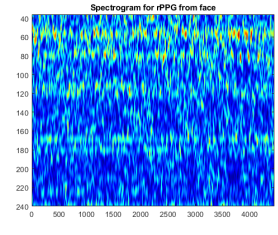
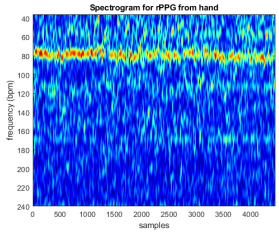
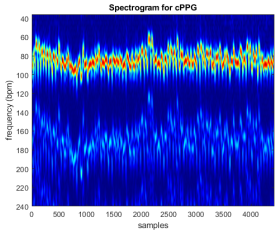
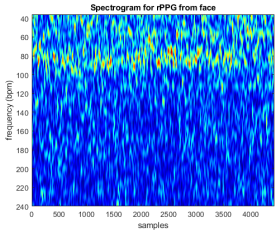
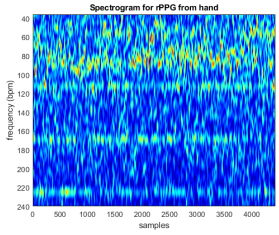
Participant	cPPG	rPPG from Face	rPPG from Hand
2			
3			
5			

Table 2.2: Spectrograms obtained for cPPG and rPPG of participants with the highest alignment for rPPG from face (Participant 2), highest alignment for rPPG from hand (Participant 3) and participant with comparable rPPG quality between hand and face (Participant 5). Frequency tracing was performed using AMTC [32] The dominant frequency trace in the cPPG spectrogram denotes the ground truth heart rate frequency.

2.6 Case Study

Table 2.3: Metrics for HR Estimation for the duration of individual recordings of each participant

Participant	**rPPG from Face**			**rPPG from Palm**		
	MAE (↓)	RMSE (↓)	ρ (↑)	MAE (↓)	RMSE (↓)	ρ (↑)
2	**0.61**	**0.74**	**0.97**	1.90	2.93	0.84
3	22.42	22.79	-0.10	**0.95**	**1.69**	**0.79**
5	6.76	11.73	0.46	5.75	9.98	0.51

The remainder of the analysis in this chapter reduces the scope from all data recordings to just a single recording for a few participants, as this allows us to highlight an interesting observation. Individual recordings contain anywhere between 540 and 1230 cycle pairs depending on the length of the video and the participant's heart rate, so a diverse range of cycles are present to perform analysis. Also, because the recordings are between 10 and 15 minutes, heart rate estimation throughout the recordings is sufficiently comprehensive.

Average values for each metric give a representative value for all rPPG cycles, but generating a distribution of the metrics gives a more holistic understanding of the rPPG alignment with cPPG. Using rPPG from both the face and the palm, each histogram is generated by dividing the range of scores into 20 bins. From the distributions of the metrics for each cycle pair, we can make a few observations. Firstly, for subject II of skin tone type III, it is evident from a glance at the histograms (fig. 2.7) that the rPPG collected from the face has a higher alignment with cPPG than rPPG collected from the palm. The opposite is true for subject III of skin tone type IV (fig. 2.8). For the rPPG from the face, there is a higher concentration of scores in the lower alignment ranges than the distribution of rPPG from the palm. The distributions for subject V of skin tone V (fig. 2.9) are much more comparable. The difference between the average value of each metric for rPPG

from the face and palm is within 1/100th of a fraction, making it difficult to assess which is more reliable.

This motivates using a time-frequency analysis to further validate the reliability of rPPG from the face vs. the palm. Spectrograms can provide additional insight into the reliability of the rPPG signals because they visually display the signal's frequency content at a particular point in the time-series signal. The frequency of this signal corresponds to the estimated heart rate. Considering the spectrogram of the rPPG can be highly corrupted with noise, we used Adaptive Multi-Trace Carving (AMTC) to identify the dominant (heart rate) frequency throughout the rPPG signal. Because AMTC can robustly track the heart rate frequency through the noise, we can extract and compare heart rate estimates from the (ground truth) cPPG, rPPG from the face, and rPPG from the palm. These heart rate estimates can then be used to compute MAE, RMSE, and Pearson Correlation as shown in table 2.3. The results support our time-domain analysis for subjects II and III in the previous paragraph. But it provides some additional insights for Participant 5. The MAE, RMSE, and Pearson Correlation of HR estimates for rPPG from the palm have a higher alignment with ground truth than rPPG from the face. This case study shows that a time-frequency analysis can validate our experimental observations from the previous experiment. Also, for instances with some ambiguity in the time-domain analysis, the time-frequency analysis can provide more confidence regarding which of rPPG from face or palm is more reliable.

2.7 Chapter Summary

Prior works have shown that there are optical and physiological motivations based on the dichromatic reflection model that suggests that it is theoretically possible to extract PPG signals from video. Unfortunately, rPPG has worse quality for people with darker skin tones because the higher melanin concentrations in their skin absorb more light and decrease the pulsatile information contained in the light reflected to the camera. Typically, rPPG is extracted from pixels that correspond to a person's face, but because PPG measures volumetric changes of blood in the microvascular bed of tissue, the technique should theoretically work for any region of skin that contains blood vessels, including the palms. There are fewer variations in the color of a people's palm regardless of skin tone type. Changing the recording from a participant's face to their hand protects the individual's privacy. Despite being well-motivated, this work showed that using the POS algorithm to extract rPPG from videos of a person's face has more alignment with cPPG than rPPG from videos of a person's palm. Experiments show this is a common trend, but the reverse can be true for some participants. This motivates identifying an ideal experimental setup to more consistently identify when rPPG from the palm could substitute rPPG from the face.

Chapter 3

Remote Physiological Sensing of ECG

3.1 Motivation

Chapter 2 explored the different areas of skin that could be used to process and extract an rPPG signal from video. Chapter 3 will build upon the results of the analysis performed in Chapter 2 to investigate the feasibility of the reconstruction of ECG from video. As mentioned earlier, prior research has shown that it is both possible and well-motivated to reconstruct an ECG signal given the corresponding PPG signal. Because cycles of remote and contact PPG represent the same signal observed from a different sensor, theoretically, it should be possible to infer ECG directly from rPPG. And since rPPG is extracted from video, this would entail a viable solution for the problem of inferring ECG from video. However, even after the post-processing steps detailed in the previous chapter, several rPPG cycles remain noisy and low in signal-to-noise ratio (SNR). Moreover, cPPG consistently has higher SNR than rPPG, which motivates introducing an intermediate mapping between remote and contact PPG before using the resulting signal to infer ECG. The full sequence of converting video to ECG can be observed in (fig. 1.1).

Alternatively, a fully end-to-end neural network-based approach could learn to reconstruct ECG by learning from the physiological signals embedded in the video; however, a benefit of our

approach is that each stage can be designed with interpretability as a guiding principle. It is well understood in the field of machine learning that neural networks behave like ”black boxes,” and it can be challenging to understand how a model came to its prediction. The modularity of this approach makes it simpler to understand the expected behavior at each stage, the final outcome of the model, and, if any issues were to arise, where the model is failing. Additionally, the techniques selected to implement each stage have components of interpretability that are absent in an entirely neural approach, as will be made clearer in the following sections. Lastly, because each stage in the video-to-ECG pipeline is its own independent and active area of research, as performance in each research area improves, the enhanced techniques can be substituted for the techniques selected here to improve the overall quality of the reconstructed ECG signals and the robustness of the overall approach.

3.2 Remote-to-Contact PPG

As stated before, other works have shown that it is possible to learn a mapping between remote PPG obtained from video and contact PPG obtained from a contact sensor [24]. In this work, we use just the simple feed-forward neural network as we have observed it to be efficacious. Prior to fitting the model, the rPPG and cPPG cycle are resampled to 50 samples per cycle to ensure a high resolution of the physiological signals for this experiment. The input and output of the model contain 50 nodes where each node corresponds to a sample in the cycle of the remote and mapped PPG, respectively. The architecture has two hidden layers, and the number of nodes in each layer was selected to be expressive enough for both a subject-specific and a group-based model. ReLU and MSE loss are used as the activation and loss functions, respectively.

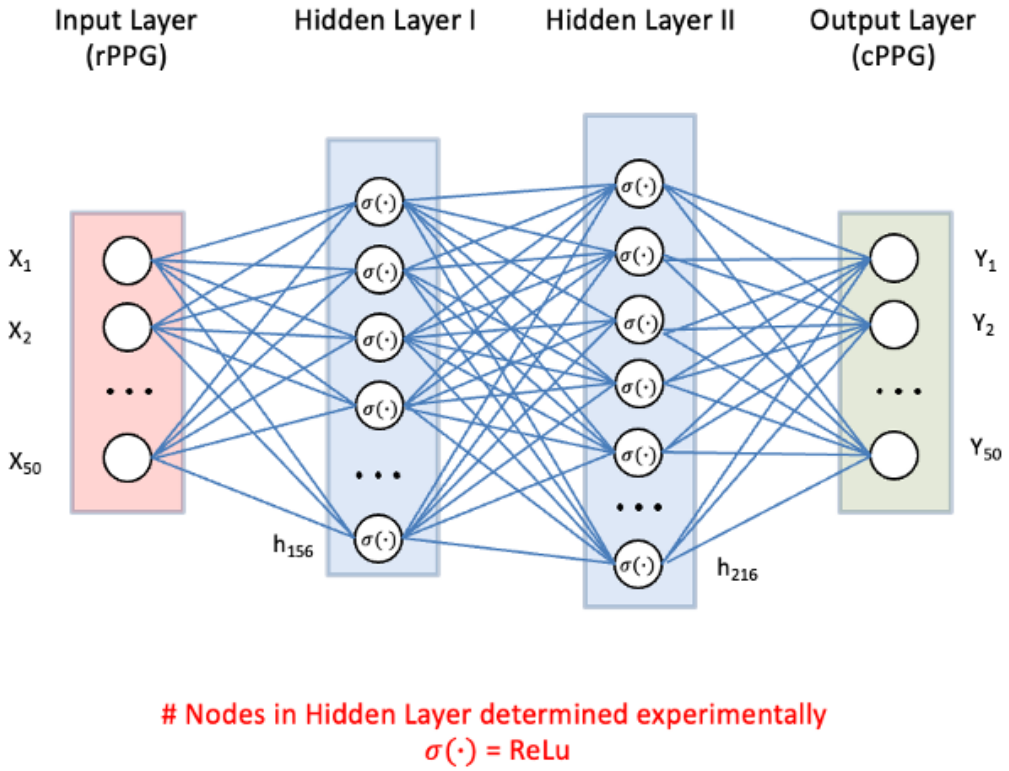


Figure 3.1: Remote-to-Contact NN Architecture

Each model is trained for up to 1500 epochs in the subject-specific case, but the precise number is determined via early stopping to prevent overfitting. The method in which we segment the rPPG and cPPG, as described in Chapter 2, keeps the cycle pair centered within its 50 samples. This has the added benefit of ensuring that the parameters learned during the mapping are only with respect to the nonstationarity of the amplitudes of both the dicrotic notch and diastolic peaks. We observed an improvement in the learned mapping when cycle pairs were consistently centered instead of located arbitrarily within the 50 samples.

Although the architecture described here is still a neural network, this stage still promotes interpretability. This is because the network employed here is not used to extract rPPG from physiological parameters embedded within a video like other approaches. Instead, the neural network

is simply used to improve the SNR of rPPG extracted using a mathematical model based on optical and physiological considerations (see POS; Chapter 2.2).

3.3 Cross-Domain Joint Dictionary Learning

The trend for neural network-based approaches has been to increase the number of parameters to increase the model's expressivity. Due to the various nonlinearities and other operations that occur, it is incredibly complicated to understand the effect of each coefficient on the model's output. On the other hand, dictionary learning techniques are inherently more interpretable due to the nature in which signals are reconstructed. In dictionary learning algorithms, a signal is reconstructed by taking a sparse linear combination from a set of atoms learned to represent the signal. The sparsity constraint imposes an upper limit on the number of atoms that can be used to reconstruct the signal, so if the sparsity constraint is small, it is easy for humans to simulate model behavior without ruining the expressivity of the algorithm. The interpretability factor further motivates us to select Cross-Domain Joint Dictionary Learning (XDJDL) for performing PPG-to-ECG. The mathematical formulations for XDJDL [13] are provided below:

$$\begin{aligned} &\min_{D_e,A_e,D_p,A_p,W} \|X_e - D_e A_e\|_F^2 \quad + \quad \alpha\|X_p - D_p A_p\|_F^2 \quad + \quad \beta\|A_e - W A_p\|_F^2 \\ &\text{subject to} \quad \|a_{p,j}\|_0 \leq t_p, \quad \|a_{e,j}\|_0 \leq t_e, \quad j = 1, ..., n \end{aligned} \tag{3.1}$$

The ECG and PPG dictionaries are denoted by D_e and D_p respectively. Sparse coding matrices for ECG and PPG are denoted by A_e and A_p respectively, and $a_{e,j}$ and $a_{p,j}$ are the jth columns of the sparse coding matrices. In this formulation, the first two norms are data fidelity terms necessary to ensure that the reconstructed ECG/PPG cycle is close in Frobenius Norm sense to the ground-

truth ECG/PPG cycle. The third norm ensures that the linear transformation matrix W learns the mapping between the sparse coding matrices for ECG and PPG. The final two norms ensure that sparsity conditions of the sparse coding matrices are maintained, i.e. the number of nonzero elements in all columns of both matrices remain less than some constant t_e and t_p. In this work, both t_e and t_p are 10.

D_e and D_p are initialized with a randomly selected subset of columns from the ECG and PPG cycles in the training split. Then a ridge regression model is used to initialize W which has the closed-form solution below:

$$W = A_e A_p^T (A_p A_p^T + \lambda I)^{-1} \tag{3.2}$$

After initializing the necessary variables, XDJDL works by iteratively updating each term in the optimization problem above (equation 3.1). First, the sparse coding matrices, A_e and A_p, are computed using Orthogonal Matching Pursuit (OMP). This is then followed with an update to each of the dictionaries. Dictionaries, D_e, D_p, and W (despite being an affine transformation matrix, W is treated as a dictionary), are updated by applying KSVD to each of the norms in the optimization problem.

3.4 Experiments

3.4.1 Video to ECG

By combining the three techniques of extracting rPPG from video, mapping rPPG to cPPG, and inferring ECG from cPPG, we are able to formulate a new technique that can reconstruct ECG from video. In this experiment, we show that rPPG from either the hand or the face of participants in the previous study can be used to reconstruct high-quality ECG signals. In this work, we train both a subject-specific and a group-based model. The subject-specific model is trained for individual participants as this allows for more personalized monitoring and potentially enables individualized treatments. The group model is beneficial because it includes data from all participants and enables the model to learn from common features present in multiple people.

3.4.2 Dataset

There exist many public remote physiological sensing datasets, but many are limited by the need for more variety in physiological signals collected. To perform the experiments in this chapter, we needed access to PPG and ECG cycle pairs. Of the many publicly available datasets, only a few record simultaneously record facial video, PPG, and ECG [33] [34] [35]. Unfortunately, when requested, the authors of the papers that developed these datasets politely denied us access to their datasets. So the dataset described in Chapter 2 is used to perform the experiments in this chapter. As mentioned earlier, the BioRadio was used to record the ECG & contact-based PPG signals. This device supports up to 12 lead node configurations, but we only utilize three nodes for the experiments conducted in this work. A different view of the ECG is obtained depending on

the electrode placements. In this dataset, the Lead II ECG is collected as this gives a good view of the P wave in the ECG cycle, which is typically the lowest amplitude segment of the signal. After the physiological signals were collected, the dataset was formed by segmenting each signal and creating tuples of ECG cycles with their corresponding rPPG and cPPG cycles. 70% of tuples are used for training, 15% are used for testing, and the final 15% are used for validation.

3.4.3 Subject-Specific Model

For each of the 5 participants for whom we recorded data, only a single recording was used for this experiment. Recordings are a minimum of 10 minutes long, so there should be sufficient diversity of rPPG and cPPG cycles. Because we have the option to use rPPG from either the face or the palm in the remote-to-contact mapping, we determine which rPPG to use by performing a time-frequency analysis as described in Chapter 2. After the time-frequency analysis, we select the rPPG with higher alignment to ground-truth cPPG as this is the more reliable source of rPPG.

Other studies typically remove bad cycle pairs when training a neural network to avoid the ”garbage in, garbage out” dilemma common in machine learning. However, in our work, after aligning rPPG and cPPG, the rPPG cycles are segmented by taking the information in between cPPG onsets. Contact PPG onsets are incredibly robust due to the low-noise nature of the signal, and thus, even though some rPPG cycles may look very noisy or contain artifacts, they are the signal generated by POS according to the video input. We hypothesize that even cycle pairs with low alignment contain micro-signals that are embedded in the POS signal that a neural network can learn from.

After performing the mapping between remote and contact PPG, the resulting mapped PPG

Table 3.1: Cycle-wise alignment of cPPG with rPPG before and after reconstruction (from single recording). rPPG before mapping is rPPG with post-processing outlined in Chapter 2. rPPG after mapping is output of neural network after remote-to-contact mapping. Participants are the same from the study in Ch. 2. Skin tone is according to the Fitzpatrick Scale (see fig. 2.6)

Participant	**Skin**	**rPPG before mapping and cPPG**			**rPPG after mapping and cPPG**		
ID	Tone	ρ (↑)	Cos Sim (↑)	RMSE (↓)	ρ (↑)	Cos Sim (↑)	RMSE (↓)
1	3	0.509	0.826	0.340	0.988	0.994	0.046
2	3	0.777	0.879	0.259	0.955	0.976	0.083
3	4	0.501	0.784	0.359	0.951	0.976	0.073
4	4	0.378	0.781	0.365	0.897	0.960	0.119
5	5	0.377	0.783	0.364	0.983	0.991	0.059

Table 3.2: Alignment of ground-truth and reconstructed ECG for subject-specific models

Participant	**Skin**	**GT and Reconstructed ECG**		
ID	Tone	ρ (↑)	Cos Sim (↑)	RMSE (↓)
1	3	0.95	0.99	0.16
2	3	0.87	0.98	0.21
3	4	0.97	0.98	0.20
4	4	0.98	0.99	0.15
5	5	0.90	0.96	0.28

has higher alignment in every metric for every participant. As can be observed from table 3.1, the Pearson correlation between the mapped PPG for all participants is consistently above .9. This is an indication that the mapped PPG signals have a significant correlation with the contact-based PPG. The same can be observed for cosine similarity in all participants. The average root mean squared error is also significantly lower for all participants after performing the mapping using the simple neural network. These results suggest that this intermediate mapping significantly improved the SNR of the PPG signal obtained from video, making it very suitable to use in the video to ECG pipeline. Also, the spread of the alignment with contact PPG is much smaller after performing the mapping, suggesting that there is a lot less variance in the quality of PPG. This should also improve the performance when using PPG from video in the PPG-to-ECG experiment in the next section.

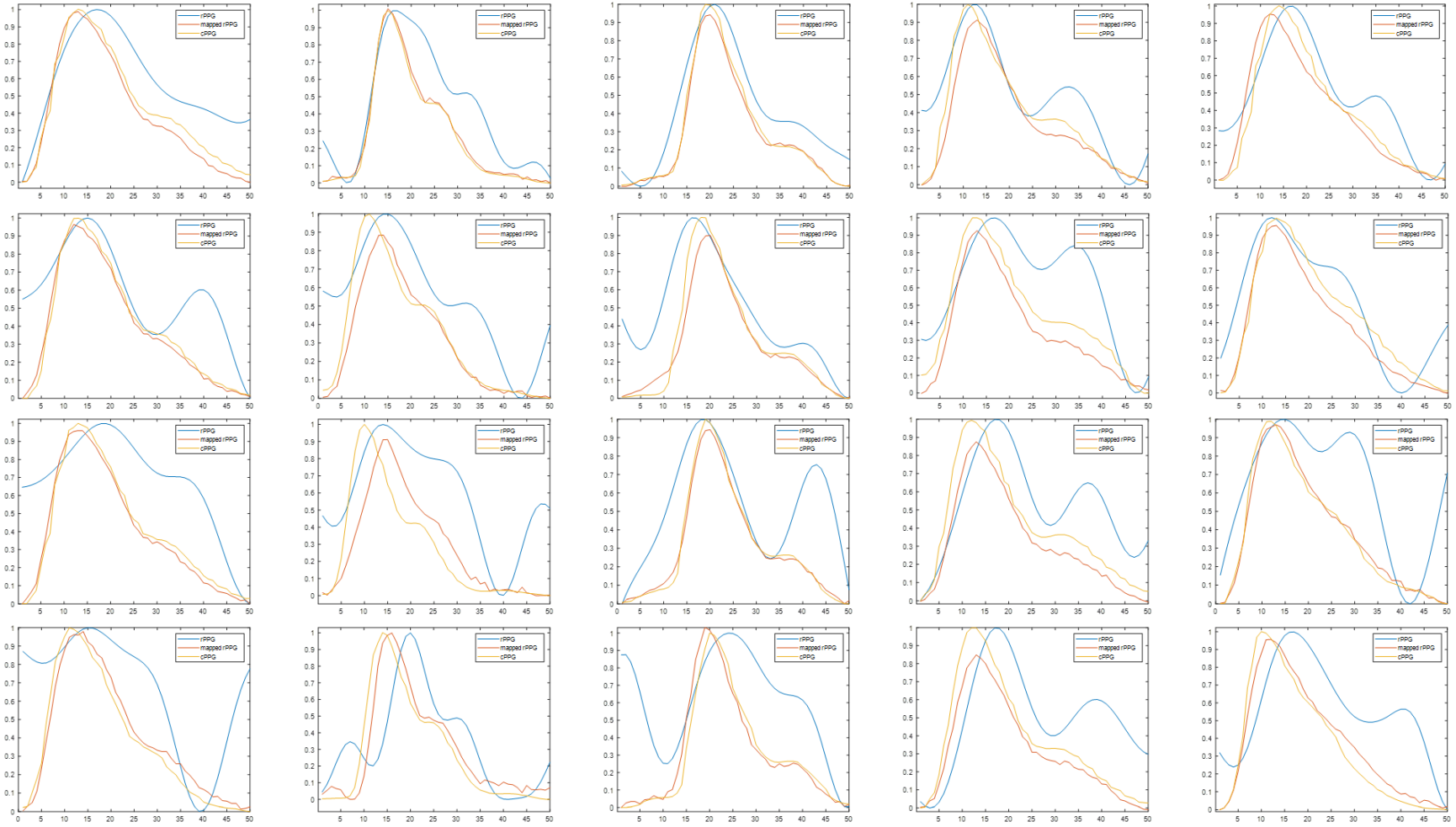


Figure 3.2: Remote-to-Contact PPG Qualitative Results. Each column corresponds data from participants 1-5, from left to right.

It is evident when looking at the visual results in figure 3.2, the mapped rPPG that was obtained from the video has a higher alignment with the ground truth contact PPG than the rPPG after simple processing. The most apparent difference between the rPPG and the output of the neural network is that the mapped rPPG no longer has any of the artifacts that were present in the original rPPG signal. Upon further examination, one can also see that the diastolic peak location is now more consistent with that of the cPPG. The mapped rPPG contains much less noise than the original rPPG making it far more suitable as input to the XDJDL, as the PPG-to-ECG technique will be more likely from the more robust features in the rPPG rather than the noise.

When looking at table 3.2, we can observe that the reconstructed ECG cycles have very high alignment with the ground-truth ECG cycles. Subject II has the lowest alignment with a Pearson correlation of 0.87. On the other hand, subject III has the highest alignment with a Pearson correlation of 0.97. This is interesting to note because subject II had the highest alignment between rPPG and cPPG in the results from Chapter 2's study but now has the lowest alignment for the reconstructed ECG. One possible explanation could be that this subject had the highest diversity in their ground-truth ECG cycles, making it difficult to reconstruct ECG cycles compared to the other subjects.

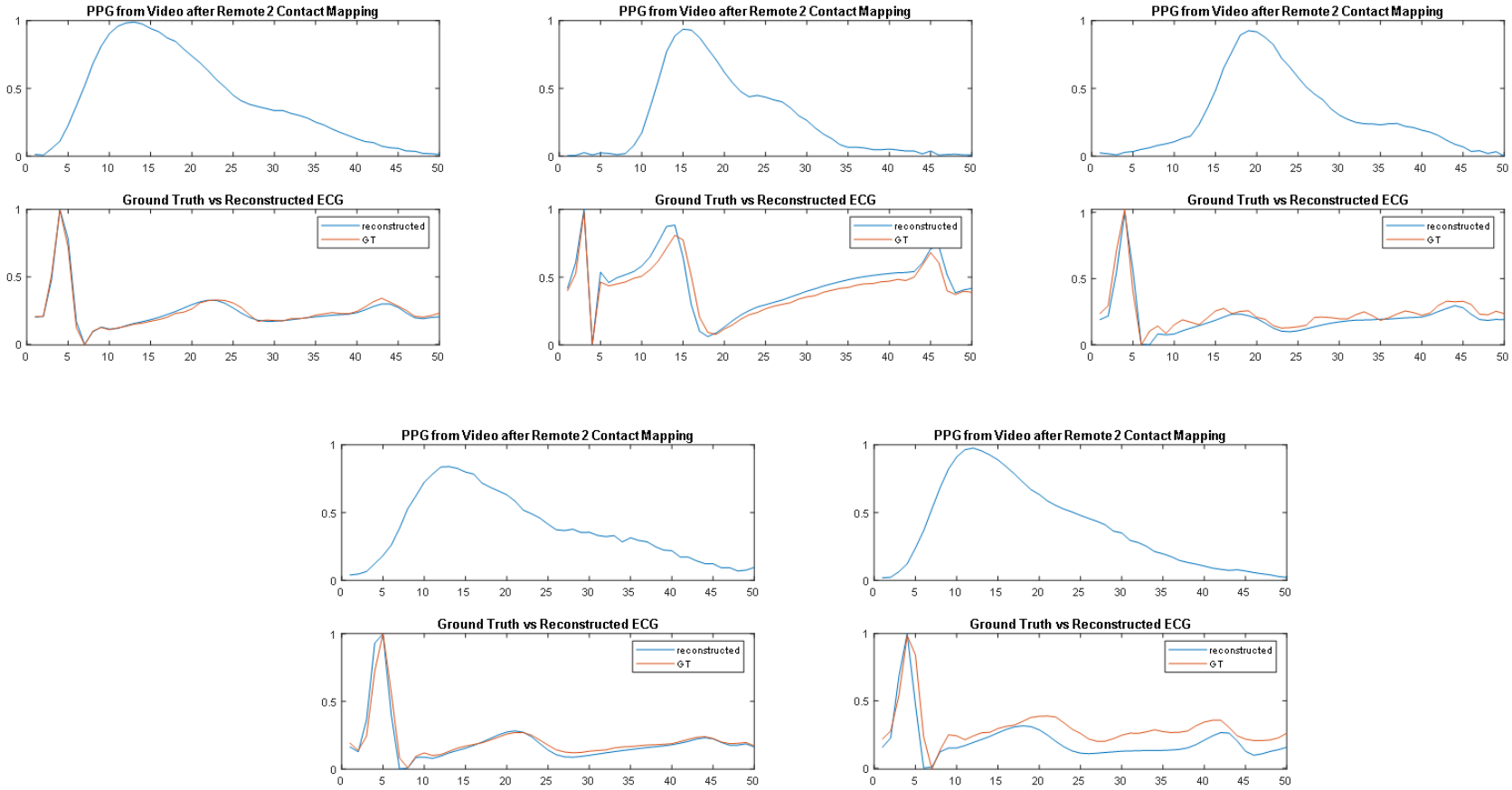


Figure 3.3: Video-to-ECG Qualitatitve Results. Each figure corresponds to participants 1-5, from left to right, top to bottom. ECG segmentation is from Q onset to the subsequent Q onset. For every participant, there is a high alignment between the reconstructed ECG cycle and the ground-truth ECG cycle.

3.4.4 Group Model

In the previous subsection, a subject-specific model was trained to reconstruct the rPPG to its cPPG form and then fed to the XDJDL model to reconstruct the ECG. This was done for each of the participant. In this section, the data for each of the participants (except participant 4) that was used to train individual models were concatenated together to form a group-based dataset. This group-based dataset was then used to train a group-based model using the same architecture for both the remote-to-contact PPG and XDJDL mapping. The table below contains the results of the model trained on the group-based dataset.

	Mapped rPPG & GT cPPG	Reconstructed ECG & GT ECG
$\rho \uparrow$	0.84	0.71
$RMSE \downarrow$	0.17	0.42

To accommodate the larger dataset, we had to increase the size of the dictionaries D_e and D_p linearly according to the number of participants incorporated into the group model. All other training requirements were left the same as when training a subject-specific model. When the dataset is combined, the reconstructed PPG has lower alignment with the ground truth cPPG than any of the individual participants' subject-specific models. This has a huge implication on the resulting mapping using XDJDL as there is a significant drop off in the ability of the model to reconstruct the ECG for all participants in the study. This is indicative that the simple neural network has trouble learning a mapping for the entire group. It is also important to mention here that the limitations of the survey in Chapter 2 also apply here. Since the videos are recorded using the ambient light in the lab, the rPPG quality could be higher using an ideal lighting configuration. Also, a variety of skin tones are used in this group model, which influences the overall shape of the PPG signal.

3.5 Chapter Summary

ECG could have been inferred from video directly, but this work uses principles from prior research to motivate each stage of the process. In this chapter, the rPPG is obtained from the inherently interpretable POS algorithm (as described in the previous chapter) and is used as the first stage in the novel video to ECG pipeline. This chapter shows that the quality of the processed rPPG signal can be improved significantly by using a simple feed-forward neural network to learn a mapping between each rPPG and cPPG cycle. Prior works have shown that when the second-order derivative is applied to both signals, one can observe high alignment between the two signals, even when it may not appear to be, which further motivates a mapping between the two domains. After each cycle of rPPG has been mapped to its corresponding cPPG cycle, the signal has a high enough SNR to be used to infer its corresponding ECG cycle. Using Cross-Domain Joint Dictionary Learning, this work shows that ECG cycles can be reconstructed from the processed PPG cycles that originated from video, thus completing the video to ECG pipeline. Additionally, due to the design and implementation of the pipeline, interpretability was enforced as a guiding principle in each stage.

Chapter 4

Conclusion and Future Perspectives

4.1 Remote PPG for Varying Skin Tones

In this chapter, we have performed a study that aimed to increase fairness and promote the privacy of individuals when extracting their rPPG. Because the palm has fewer variations in color between people of various skin tones and, overall, less melanin than the rest of the skin, the palm is an interesting option for collecting rPPG. However, rPPG from the palm has yet to be studied extensively, so in this work, we compared the morphology between rPPG from face pixels and palm pixels and their alignment with cPPG. The results showed that although for most participants, rPPG from the face had higher alignment than rPPG from the palm, for one individual, the opposite was true. For this individual, rPPG from the palm had higher alignment with cPPG than rPPG from the face. Although this was for a single individual, this finding is encouraging that there might be suitable conditions for when a person's palm might be able to generate rPPG that has a higher quality than the rPPG from the face.

4.2 Remote Physiological Sensing of ECG

By combining the research areas of video to rPPG, remote-to-contact PPG, and PPG-to-ECG, we developed a novel and interpretable technique for reconstructing ECG from video. Our experiments show that a subject-specific model can reconstruct ECG for participants of various skin tones. Additionally, we showed that, although there was a drop in performance when compared to the subject-specific model, a group-based model could still be trained to reconstruct the ECG cycles of its individuals. We speculate that the group model would have a higher performance if a more complex model was used for the intermediate mapping from remote to contact PPG.

4.3 Opportunities for Future Work

There are mutiple pathways for improving remote physiological sensing for those with darker skin tones. First, in this work, remote PPG was collected from recordings of participants; however, this was not in a controlled environment. The light source used during the recording remained constant for all participants, but neither the lux of the light nor the presence of specular reflections was controlled for. The main goal of this work was to explore the feasibility of converting video to ECG for participants of various skin tones. As only a few works have experimented with recovering rPPG from the hand, future works can investigate the ideal lighting configurations necessary to maximize pulsatile information in rPPG from the palm. Secondly, this work has shown that varying skin tones has some effect on the quality of the remote PPG signal observed. This could be further expanded to identify the relationship between the amount of lux needed from a light source and particular skin tone for recovering the best rPPG signal. In this work, the only rPPG technique

used was POS. Performance could potentially be improved using data-driven rPPG techniques trained for particular lighting conditions, and skin tones instead of a general-purpose, one size fits all algorithm like POS.

Regarding video to ECG, there are several directions for future work. Firstly, because video-to-ECG is a multi-stage pipeline, where each stage is its own independent research area, improvements in each research area will lead to increased robustness in the video-to-ECG pipeline. Since a group model has been shown to work effectively, this could further lead to the development of a pre-trained model that other future collaborators could fine-tune in a transfer learning setting to enable digital twin capacities. Additionally, more work can go into further validating the results observed here. Recently explainable AI (XAI) techniques have been used to explore what are the salient parts of the input that lead to a model's prediction, which gives additional insights into how the model is making a decision. This could be applied to the current works to get a better understanding of the reconstructed signal. This work laid a foundation for why it should be feasible to reconstruct ECG from video, but in the process, there was significant pre and post-processing of various signals. Now that the foundation has been laid out, it should be possible to construct an end-to-end model for video to ECG.

Bibliography

[1] Gregory A Roth, George A Mensah, Catherine O Johnson, Giovanni Addolorato, Enrico Ammirati, Larry M Baddour, Noël C Barengo, Andrea Z Beaton, Emelia J Benjamin, Catherine P Benziger, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update from the GBD 2019 Study. *Journal of the American College of Cardiology*, 76(25):2982–3021, 2020.

[2] Nuzhat Ahmed and Yong Zhu. Early Detection of Atrial Fibrillation Based on ECG signals. *Bioengineering*, 7(1):16, 2020.

[3] Solmaz Mahmoodzadeh, Mansour Moazenzadeh, Hamidreza Rashidinejad, and Mehrdad Sheikhvatan. Diagnostic performance of electrocardiography in the assessment of significant coronary artery disease and its anatomical size in comparison with coronary angiography. *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, 16(6):750, 2011.

[4] Samy A Abdelghani, Todd M Rosenthal, and Daniel P Morin. Surface Electrocardiogram Predictors of Sudden Cardiac Arrest. *Ochsner Journal*, 16(3):280–289, 2016.

[5] Anna Rosiek and Krzysztof Leksowski. The risk factors and prevention of cardiovascular disease: the importance of electrocardiogram in the diagnosis and treatment of acute coronary syndrome. *Therapeutics and Clinical Risk Management*, pages 1223–1229, 2016.

[6] Steven R Steinhubl, Jill Waalen, Alison M Edwards, Lauren M Ariniello, Rajesh R Mehta, Gail S Ebner, Chureen Carter, Katie Baca-Motes, Elise Felicione, Troy Sarich, et al. Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: the mSToPS Randomized Clinical Trial. *JAMA*, 320(2):146–155, 2018.

[7] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062, 2014.

[8] Wikipedia. Electrocardiography, May 2023. URL: `https://en.wikipedia.org/wiki/Electrocardiography`.

[9] Lumen Learning. Biology for majors ii, May 2023. URL: https://courses.lumenlearning.com/wm-biology2/chapter/the-cardiac-cycle/.

[10] Denisse Castaneda, Aibhlin Esparza, Mohammad Ghamari, Cinna Soltanpur, and Homer Nazeran. A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, 4(4):195, 2018.

[11] Junyung Park, Hyeon Seok Seok, Sang-Su Kim, and Hangsik Shin. Photoplethysmogram Analysis and Applications: An Integrative Review. *Frontiers in Physiology*, 12:2511, 2022.

[12] Qiang Zhu, Xin Tian, Chau-Wai Wong, and Min Wu. Learning Your Heart Actions From Pulse: ECG Waveform Reconstruction From PPG. *IEEE Internet of Things Journal*, 8(23):16734–16748, 2021.

[13] Xin Tian, Qiang Zhu, Yuenan Li, and Min Wu. Cross-Domain Joint Dictionary Learning for ECG Inference From PPG. *IEEE Internet of Things Journal*, 10(9):8140–8154, 2023.

[14] Qunfeng Tang, Zhencheng Chen, Rabab Ward, Carlo Menon, and Mohamed Elgendi. PPG2ECGps: An End-to-End Subject-Specific Deep Neural Network Model for Electrocardiogram Reconstruction from Photoplethysmography Signals without Pulse Arrival Time Adjustments. *Bioengineering*, 10:630, 05 2023.

[15] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[16] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011.

[17] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011.

[18] Gerard de Haan and Vincent Jeanne. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[19] Wenjin Wang, Sander Stuijk, and Gerard de Haan. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Transactions on Biomedical Engineering*, 63(9):1974–1984, 2016.

[20] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks. In *British Machine Vision Conference*, 2019.

[21] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018.

[22] Akash Kumar Maity, Jian Wang, Ashutosh Sabharwal, and Shree K. Nayar. RobustPPG: camera-based robust heart rate estimation using motion cancellation. *Biomedical Optics Express*, 13(10):5447–5467, Oct 2022.

[23] Daniel McDuff, Sarah Gontarek, and Rosalind W. Picard. Remote Detection of Photoplethysmographic Systolic and Diastolic Peaks Using a Digital Camera. *IEEE Transactions on Biomedical Engineering*, 61(12):2948–2954, 2014.

[24] So-Eui Kim, Su-Gyeong Yu, Na Hye Kim, Kun Ha Suh, and Eui Chul Lee. Restoration of Remote PPG Signal through Correspondence with Contact Sensor Signal. *Sensors*, 21(17), 2021.

[25] Wenjin Wang, Sander Stuijk, and Gerard de Haan. Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG. *IEEE Transactions on Biomedical Engineering*, 62(2):415–425, 2015.

[26] Wenjin Wang and Caifeng Shan. Impact of makeup on remote-PPG monitoring. *Biomedical Physics Engineering Express*, 6, 10 2019.

[27] Mika P Tarvainen, Perttu O Ranta-Aho, and Pasi A Karjalainen. An advanced detrending method with application to HRV analysis. *IEEE Transactions on Biomedical Engineering*, 49(2):172–175, 2002.

[28] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. 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(23):e215–e220, 2000 (June 13).

[29] Meiyun Cao, Timothy Burton, Gennadi Saiko, and Alexandre Douplik. Remote Photoplethysmography with a High-Speed Camera Reveals Temporal and Amplitude Differences between Glabrous and Non-Glabrous Skin. *Sensors*, 23(2):615, 2023.

[30] Christopher J Abularrage, Anton N Sidawy, Gilbert Aidinian, Niten Singh, Jonathan M Weiswasser, and Subodh Arora. Evaluation of the microcirculation in vascular disease. *Journal of Vascular Surgery*, 42(3):574–581, 2005.

[31] Emerge Medical. Fitzpatrick scale, Mar 2023. URL: `https://emergetulsa.com/fitzpatrick/`.

[32] Qiang Zhu, Mingliang Chen, Chau-Wai Wong, and Min Wu. Adaptive Multi-Trace Carving for Robust Frequency Tracking in Forensic Applications. *IEEE Transactions on Information Forensics and Security*, 16:1174–1189, 2020.

[33] Xiaobai Li, Iman Alikhani, Jingang Shi, Tapio Seppanen, Juhani Junttila, Kirsi Majamaa-Voltti, Mikko Tulppo, and Guoying Zhao. The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 242–249, 2018.

[34] Justin R. Estepp, Ethan B. Blackford, and Christopher M. Meier. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1462–1469, 2014.

[35] Amogh Gudi, Marian Bittner, and Jan van Gemert. Real-Time Webcam Heart-Rate and Variability Estimation with Clean Ground Truth for Evaluation. *Applied Sciences*, 10(23), 2020.

[36] Xin Tian. *Digital Smart Health Via Physiological Signal Sensing and Learning*. PhD thesis, University of Maryland, College Park, 2022.

[37] Carl Frederick Steinhauser. Compression and multi-spectral sensing for video based physiological monitoring, 2022.