

RICE UNIVERSITY

**Robust acquisition of Photoplethysmograms using
a Camera**

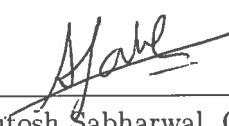
by

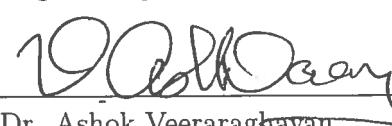
Mayank Kumar

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Science

APPROVED, THESIS COMMITTEE:


Dr. Ashutosh Sabharwal, Chair
Professor of Electrical and Computer
Engineering and Computer Science


Dr. Ashok Veeraraghavan
Assistant Professor of Electrical
Engineering


Behnam Aazhang
J.S. Abercrombie Professor of Electrical
and Computer Engineering

Houston, Texas

May, 2014

ABSTRACT

Robust acquisition of Photoplethysmograms using a Camera

by

Mayank Kumar

Non-contact monitoring of vital signs, such as pulse rate, using a camera is gaining popularity because of its potential for ubiquitous in-situ health tracking. However, current methods of camera-based vital sign monitoring have poor performance for people having darker skin tones and/or in the presence of relative motion between the camera and the subject. In this thesis, we propose *distancePPG*, a new algorithm which addresses aforementioned challenges and can reliably estimate the underlying photoplethysmogram (PPG) waveform for most skin tones in presence of subject motion. We first propose a new method to combine the PPG signal from different regions of the skin, which improves the SNR of the estimated PPG signal by 4.7dB on an average compared to past methods. Second, by tracking different regions of the face independently during motion, our algorithm provides a gain of 5.3 dB in SNR compared to past methods of motion tracking.

Acknowledgments

First and foremost, I would like to acknowledge the help and contribution of Jason Holloway, Adam Samaniego and Vivek Boominathan to this work. They have been instrumental in planning the experiments, developing prototype and volunteered repeatedly for data collection. Moreover, methods developed in this thesis have matured through continuous discussion with Vivek and Jason.

In addition, I would like to acknowledge Evan Everett, Achal Sahai, Ashok Veeraraghavan, and Ashu Sabharval. I have turned to each of them for expert technical guidance related to the work in this thesis.

Contents

Abstract	ii
Acknowledgments	iii
List of Illustrations	v
List of Tables	vi
1 Introduction	1
1.1 Prior art	3
1.2 Motivation	5
1.3 Main contributions	5
1.4 Organization of thesis	6
2 Background	8
2.1 Background of PPG waveform	8
2.2 Cardiovascular system - pressure and flow waveform	11
2.3 PPG - vital signs and waveform features	12
3 Theory of PPG estimation using camera	14
3.1 General camera-based PPG acquisition system	14
3.2 distancePPG algorithm	17
3.2.1 Camera based PPG signal model	17
3.2.2 MRC algorithm for PPG estimation	19
3.2.3 Region based motion tracking algorithm	23
4 Camera based PPG system assessment	26
4.1 Experimental setup	26
4.2 Performance metric	27
4.3 Experiments	29
4.3.1 MRC algorithm evaluation	29
4.3.2 Region-based motion tracking evaluation	30
4.3.3 System parameter evaluation	31
4.4 Results	32
4.4.1 MRC algorithm performance	32
4.4.2 Region-based motion tracking performance	34
4.4.3 System parameter assessment	34

5 Discussion and Conclusion	40
5.1 Goodness metric and SNR	40
5.2 MRC and choice of G_{thres}	41
5.3 SNR threshold for reliable PR	47
5.4 PRV estimation accuracy using camera PPG	49
5.5 Limitation of distancePPG under motion	50
5.6 Conclusion	51

Illustrations

2.1	Extinction coefficient of Hb and HbO ₂	10
2.2	Important features of PPG waveform	13
3.1	A camera-based PPG acquisition system	15
3.2	Effect of skin ROI size and location on PPG signal strength	16
3.3	Block diagram of distancePPG algorithm	18
3.4	Working of MRC algorithm	22
4.1	Performance of MRC algorithm	37
4.2	PR and PRV estimation performance	38
4.3	PPG Shape comparison	38
4.4	Performance of region based motion tracker	39
5.1	Scatter plot between goodness metric (dB) and SNR (dB)	42
5.2	SNR-vs-Threshold - Various skin tones	46
5.3	Pulse rate error Vs SNR	48

Tables

4.1	Performance comparison of distancePPG with current methods under different natural motion scenarios.	35
4.2	Lighting conditions and filters	36

Chapter 1

Introduction

Regular and noninvasive measurement of vital signs such as pulse rate (PR), breathing rate (BR), pulse rate variability (PRV), blood oxygen level (SpO_2) and blood pressure (BP) are important both in hospital and at home due to their fundamental role in the diagnosis of health and well being. Currently, the gold standard technique to measure these vital signs are based on contact sensors such as ECG probes, chest straps, pulse oximeters and blood pressure cuffs. These contact sensors can cause skin damage and infection to patients in hospitals, and are generally inconvenient for continuous use at home. Recently, [1] showed that a normal camera can be used for vital sign (PR,BR) monitoring under ambient light. Being non-contact, camera-based vital sign monitor could have many applications, from monitoring new born babies in the ICU to in-situ continuous monitoring in low-mobility scenarios like sitting in front of a computer.

Current known methods of non-contact camera-based vital sign monitoring [2] do not give good results for people having darker skin tones and under low lighting conditions [3]. Moreover, most of these algorithms require a person to be nearly at rest facing a camera to have reliable measurements. In this thesis, we address these challenges to expand the scope of camera-based vital sign monitors.

Camera-based vital sign monitors work on the principle of Photoplethysmography

(PPG), which is an optical method to measure cardiac-synchronous blood volume change in body extremities such as the face, finger and earlobe. Vital signs such as PR, PRV, SpO₂ and BR can be derived from a well-acquired PPG waveform. Conventionally, the PPG waveform is obtained by using a contact-based pulse oximeter.

The two major challenges in acquiring PPG using a camera are: 1) extremely low signal strength of the recorded PPG signal particularly for darker skin tones and under low lighting conditions, and 2) motion artifacts due to an individual's movement in front of the camera. Our main contribution is a new algorithm, labeled as *distancePPG*, which addresses both these challenges. DistancePPG reliably acquire PPG signal in-situ using a camera from subjects having different skin tones during natural motion.

To improve the signal strength, distancePPG optimally combines PPG signals obtained from different regions (e.g. 20×20 pixel patch) of the face using a weighted average. The weights capture the strength of the PPG waveform obtained from the corresponding skin patches. Our key innovation is a new automatic *online* method of determining the weights based only on the video recording of the subject. When the subject moves in front of camera, we track all the patches independently using the KLT (Kanade-Lucas-Tomasi) feature tracker [4, 5] to extract the PPG waveform under motion.

For different skin tones (pale white to brown), the distancePPG algorithm improves the signal to noise ratio (SNR) of the acquired PPG signal on an average

by 4.7dB compared to current methods. The proposed algorithm is robust against different types of motion exhibited by a person using a laptop or tablet computer to browse the Internet, watch video or talk on Skype and improves the SNR of acquired camera PPG in these scenario by 5.3dB on an average. Further, it improves the SNR of camera based PPG by as much as 8dB under low lighting condition when compared to current methods.

The improvement in SNR helps reduce the error in pulse rate (PR) estimation below the medically acceptable limit of 3 bpm for all skin tones. Further, it pushes the error in pulse rate variability (PRV) estimation below 20 ms for the majority of skin tones. As the shape of the PPG waveform acquired using our algorithm is very similar to the one acquired using a contact-based pulse oximeter, many other applications of PPG which depend on the exact shape of the PPG waveform, such as detecting arterial disease, arterial compliance and aging, venous assessment, endothelial function and vasospastic condition [6], can now become feasible by just using a camera.

1.1 Prior art

Over the last decade, considerable research effort has gone into development of non-contact camera based PPG measurement system [7, 8, 1, 9, 10, 2, 11, 12]. Initially, external arrays of LEDs at different wavelength (red, infrared) were used to illuminate a region of tissue for measuring PR and SpO₂ using a monochrome CMOS camera

[7, 8, 9]. Good correspondence for PR estimation have been established quantitatively, but not much evidence was provided for reliable SpO₂ measurements.

Later, Verkruysse et al. [1] showed that PR and BR can be determined using simply a color camera and ambient illumination. The author reported that the *green channel* of the RGB camera gives the best results for detecting PR and BR. More recently, Poh et al. [10, 2] used a webcam under ambient illumination to detect simultaneous PR, PRV and BR of multiple people in a video by employing Independent Component Analysis (ICA) on three color channels (R,G,B) to extract the PPG signal. This approach was further investigated in Lewandowska et al. [13], where they claimed that principal component analysis (PCA) of the three-color channel can be as effective as ICA.

Most of the past work, however, did not report how the system performs on individuals having different skin tones. It is well-known that the *melanin* present in darker skin tones absorbs a lot of light, and thus attenuates the SNR of the acquired PPG signal, making the system ineffective for extracting vital signs.

To counter the motion artifact challenge, Poh et al. [10] used automatic face detection in consecutive frames to track the face. In Sun et al. [14], the authors computed 2D shifts between consecutive frames using image correlation to model the motion. Both these approaches [14, 10] could only captures translation motion of face, and do not cater to more natural motion like tilting of face, smiling or talking generally found in in-situ scenarios.

1.2 Motivation

Apart from being used as non-contact vital sign monitor, camera based PPG acquisition system also provides a new paradigm of visualizing blood flow in arteries just underneath the skin. In this direction, past researchers [9, 15, 11] have attempted to visualize blood perfusion in an illuminated tissue using a camera. Moreover, PPG waveform in itself provides many other medically relevant information and a well acquired PPG signal can be used for vascular assessment, physiological monitoring, and autonomic function assessment [6] as highlighted earlier.

Most past researchers posed the problem of camera based vital sign (PR,BR,PRV) extraction as a *parameter estimation* problem with RMSE (root mean squared error) as a figure of merit. But, all the above applications could only become possible if the signal quality of the camera acquired PPG waveform is sufficient for further processing. Thus, in this thesis we pose it as a *waveform estimation* problem with SNR (signal-to-noise ratio) as a figure of merit.

In this thesis we focus our efforts in developing algorithms (distancePPG) for PPG waveform estimation for people having different skin tones, under varied lighting condition and natural motion scenarios.

1.3 Main contributions

The main contribution of this thesis is the distancePPG algorithm to improve the SNR of the camera based PPG acquisition. It consists of two main parts - 1) weighted

averaging of PPG waveform obtained from different regions of the face (MRC algorithm), and 2) region based motion tracker to compensate for small natural motion. The key innovation is a new automatic online method of determining the weights in the MRC algorithm based only on the video recording of the subject.

Further, we formally define signal-to-noise (SNR) of camera based PPG waveform, and use it as a metric to characterize and quantify the performance of our proposed waveform estimation algorithm (*distancePPG*) and compare it with prior methods. Next, we present how various optical and camera parameters affect the SNR of the acquired waveform. Few relevant system parameters which we considered are type of illumination , wavelength of light, camera's analog gain, exposure time and frame rate.

1.4 Organization of thesis

Chapter 2 discusses the origin of the PPG waveform, its relationship to cardiovascular system and how various biomarkers can be extracted from the waveform. It also highlights the general imaging PPG system architecture used in past works, as well as prototyped in this work.

Chapter 3 details the distancePPG algorithm for faithful camera based PPG extraction. We first propose a new camera based PPG acquisition model. Based on this model, we then propose a PPG estimation algorithm labeled as *MRC algorithm*. We then describe a region-based motion tracking algorithm which keeps track of different

regions of the face as the person moves in front of the camera.

Chapter 4 first lays out the evaluation methodology used in this thesis for numerical comparison (SNR metric). Then it details the experiments conducted to characterize and quantify the performance of distancePPG algorithm in comparison to past methods. Also discussed are experiments which are done to study various relevant system parameters. This is followed by a summary of all the results. In addition, we also present results for three main applications - (i) PR, (ii) PRV, (iii) wave shape estimation and how distancePPG algorithm improves the performances of these estimates.

Chapter 5 concludes this thesis with a discussion on the implications of the findings of this work, their application, limitations and scope of future work.

Chapter 2

Background

2.1 Background of PPG waveform

A blood volume waveform (or PPG) comprises a pulsatile ('AC') physiological waveform attributed to cardiac synchronous changes in the blood volume, and is superimposed on a slowly varying ('DC') baseline with various lower frequency components attributed to respiration, sympathetic nervous system activity and thermoregulation [6].

When light is incident on tissue, some part of it is absorbed by the blood in the arteries underneath it. The amount of blood in the arteries changes with each cardiac cycle and thus the level of absorption of light changes as well. Thus, PPG is recorded by measuring the intensity of remaining light transmitted through the tissue (pulse oximeter) or reflected from the tissue (camera based).

The amount of light transmitted through a fluid (blood) is quantified using the Beer-Lambert law which states that the intensity of transmitted light $I(\lambda)$ is given by

$$I(\lambda) = I_0 \cdot e^{-\mu_\lambda l}, \quad (2.1)$$

where λ and I_0 are the wavelength and intensity, respectively, of the incident light

source, μ is the molar coefficient of absorption of the chromophores present in blood such as hemoglobin (Hb) and oxy-hemoglobin (HbO_2).

For a camera-based PPG system, the incident light undergoes multiple scattering events in the tissue and the blood, and is reflected back (back scattering) before being recorded in the camera. Though the Beer-Lambert law is too simplistic to model this complicated light tissue interaction, the basic idea is that the effective optical path length (l) changes as the blood volume changes over time, thus modulating the reflected light intensity with the blood volume waveform.

Since the skin vascular bed contains a very small amount of blood (2 – 5%), and the blood volume itself experiences only a small (5%) change with the cardiovascular pulse wave [16], the pulsatile component amounts for a very small portion of the total reflected light intensity. This weak modulation of the light intensity is the *main* reason for low signal strength of the camera-recorded PPG signal.

The PPG signal can best be recorded in the 520 – 580 nm spectral band as the absorption spectra of hemoglobin (Hb) and oxyhemoglobin (HbO_2), two major chromophores in blood, peaks between 520 – 580nm as highlighted in Figure 2.1. The other peak at 400 – 450nm is not relevant as melanin absorption is very high in that region.

Thus, one can use optical filtering to discard off a large portion of the skin surface reflection which does not exhibit any cardio-synchronous pulsatile changes. Evidently, past researchers have found that a PPG signal is best extracted from the *green channel*

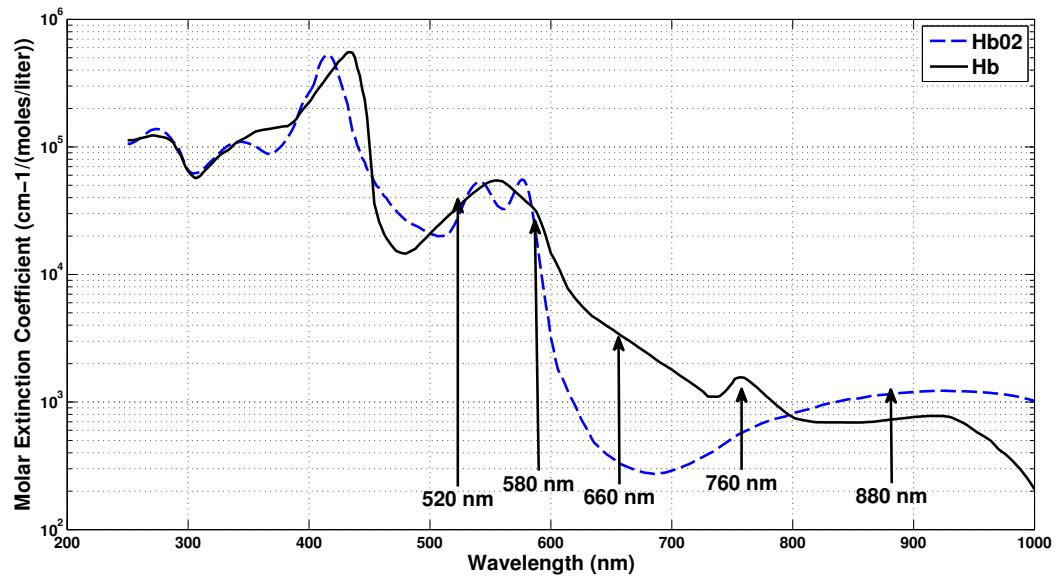


Figure 2.1 : Extinction coefficient of Hb and HbO₂ *

of the RGB camera. In this work, we have used an external green filter in front of a monochrome camera for most of our experiments. This filtering would allow one to improve signal strength of camera-based PPG signal as one can use a larger camera exposure time without saturating the CMOS sensor, and larger analog gain of the camera without saturating the ADC.

*Values compiled by Scott Prahl <http://omlc.ogi.edu/spectra/hemoglobin/summary.html>.

Molar extinction coefficient, e , can be converted to absorbance, A , by multiplying it with molar concentration and path length.

2.2 Cardiovascular system - pressure and flow waveform

The human cardiovascular system consist of the *heart* which pumps the blood into the arteries and provide convective transport of blood between the different organ of the body. Due to the high resistance against flow in the micro-circulation (towards the periphery), the transport of blood require a relatively high perfusion pressure.

When the heart pumps the blood, each *cardiac pulse* is associated with a *pressure wave* which originates from the heart and travels down into the circulatory system towards the periphery, where it terminates. In any arterial cross section, pressure wave has an associated *flow wave* which characterizes the rate of flow of blood through that cross-section. The time integral of the flow wave is the *blood volume waveform* (BVP) and it represent the amount of blood present in the arterial cross-section at any given time. The PPG waveform that we measure is proportional to the BVP at the site of measurement. Thus,

$$PPG(t) \propto \int_0^t Q(\tau) d\tau, \quad (2.2)$$

where $Q(t)$ is the flow wave associated with cardiac pulse.

Several researchers have found that blood circulation in arteries is a linear phenomenon with less than 5 percent energy present in the non-linearities [17]. Thus, to a first-order approximation, a linear relationship may be assumed between harmonic components of pressure and of flow waves, such that the relationship between pressure and flow waveform at some arterial cross-section may be described by vascular impedance [17]. Thus, one can relate PPG waveform to a flow waveform, and

finally to blood pressure waveform. This clearly highlights the feasibility of diagnostic assessment of arterial properties by measuring the PPG waveform, and has been explored by other researchers [18].

2.3 PPG - vital signs and waveform features

PPG waveform is periodic in nature and the first harmonic frequency corresponds to the pulse rate (PR). Thus, the largest peak in the power spectrum density (PSD) of PPG over a time window is the average pulse rate in that window. The PR of a person generally changes over time, hence the time window considered in practice is somewhere between 5 – 60 sec, but not larger. PR is same as heart rate (HR) determined using electrocardiography (ECG) since the mechanical activity of heart (pulse) is coupled to its electrical activity (ECG).

The heart beats irregularly and there is a beat-to-beat difference in pulse interval known as pulse rate variability (PRV). PRV can be estimated by computing the time difference between successive *minima* in the PPG waveform. As with HR, PRV is related to the heart rate variability (HRV) which is traditionally derived from the RR interval time series of the ECG. However, the PPG signal lags behind the ECG signal by the time required for transmission of blood flow wave from the heart to the measurement site. The pulse travel time shows very minor beat-to-beat variations (a few millisecond), such that heart beat interval derived from ECG and PPG can be used as a replacement for HRV derived from ECG. [19]. In this thesis, we compare

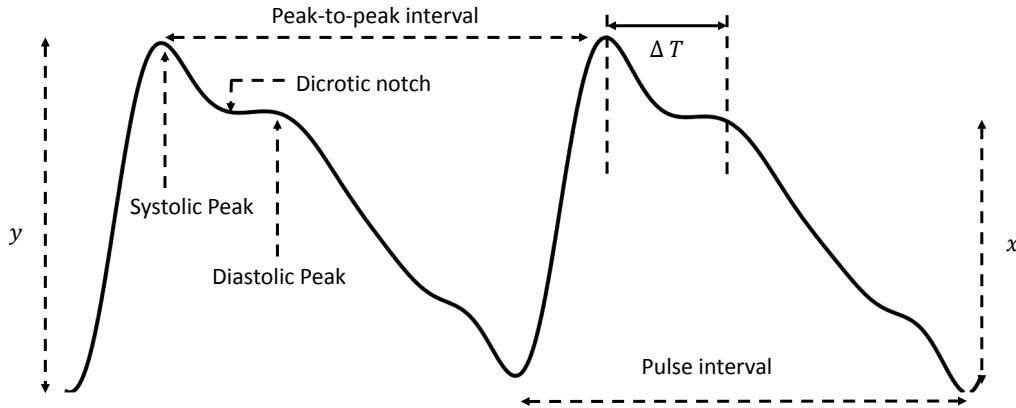


Figure 2.2 : Important features of PPG waveform

and report the accuracy of PRV derived from pulse oximeter-based PPG and our camera-based PPG.

Apart from PR and PRV, many other features of the PPG waveform are medically relevant. In Figure 2.2, we summarized few of them which are discussed in detail in [20]. If camera-based PPG waveform has to replace pulse oximeter-based PPG waveform for these application, the SNR of the camera obtained waveform should be sufficient to extract these features.

Chapter 3

Theory of PPG estimation using camera

3.1 General camera-based PPG acquisition system

In a camera-based PPG system, one records the video of person's face (or other body parts) sitting in front of a camera at a distance d , illuminated by some light source in an uncompressed format (see Figure 3.1). For each frame in the video recording, a region of interest (ROI) is selected from which PPG signal has to be extracted. Then, this ROI is fed into a *PPG estimation algorithm*.

To appreciate the challenges in estimating PPG using a camera, consider the common procedural steps of PPG estimation algorithm followed by most past work. First, average pixel intensity is computed over the selected ROI (spatial averaging) for all the frames in the video. This pixel intensity signal (color change) over time is filtered in the passband of the PPG (0.5Hz - 5Hz) to get an estimate of PPG signal. The recorded PPG signal in this manner is very small in magnitude and is comparable to the quantization step of the 8 – 10 bit ADC present in cameras. Thus, camera-based PPG suffers from large quantization noise.

To overcome the quantization noise challenge, generally a larger ROI on skin surface is used for averaging. In most past work, researchers have used either the

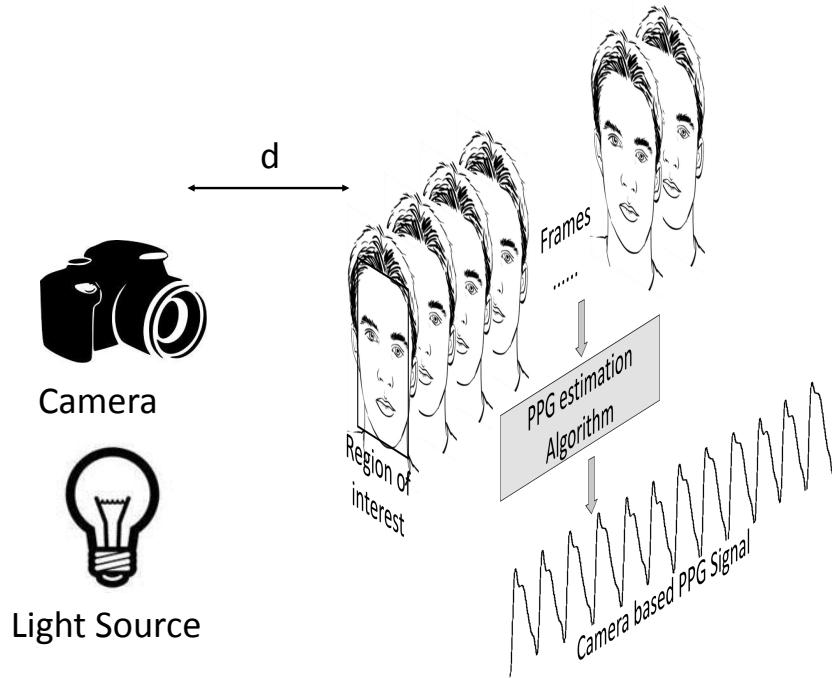


Figure 3.1 : A camera-based PPG acquisition system

whole face region or the forehead region for extracting PPG. However, using very large skin region can actually *reduce* the signal strength of the PPG. This is because the strength of the PPG signal extracted from different regions of the skin surface depends on the blood perfusion in that region. The blood perfusion depends on the arterial concentration underneath the skin, and it varies from subject to subject. When the pixel intensity is averaged over a large skin ROI, we end up including regions which do not show any blood perfusion and hence do not contain the PPG signal. See Figure 3.2 which clearly shows that the signal strength of PPG acquired from the whole face is much weaker than the one acquired from a small region in the forehead.

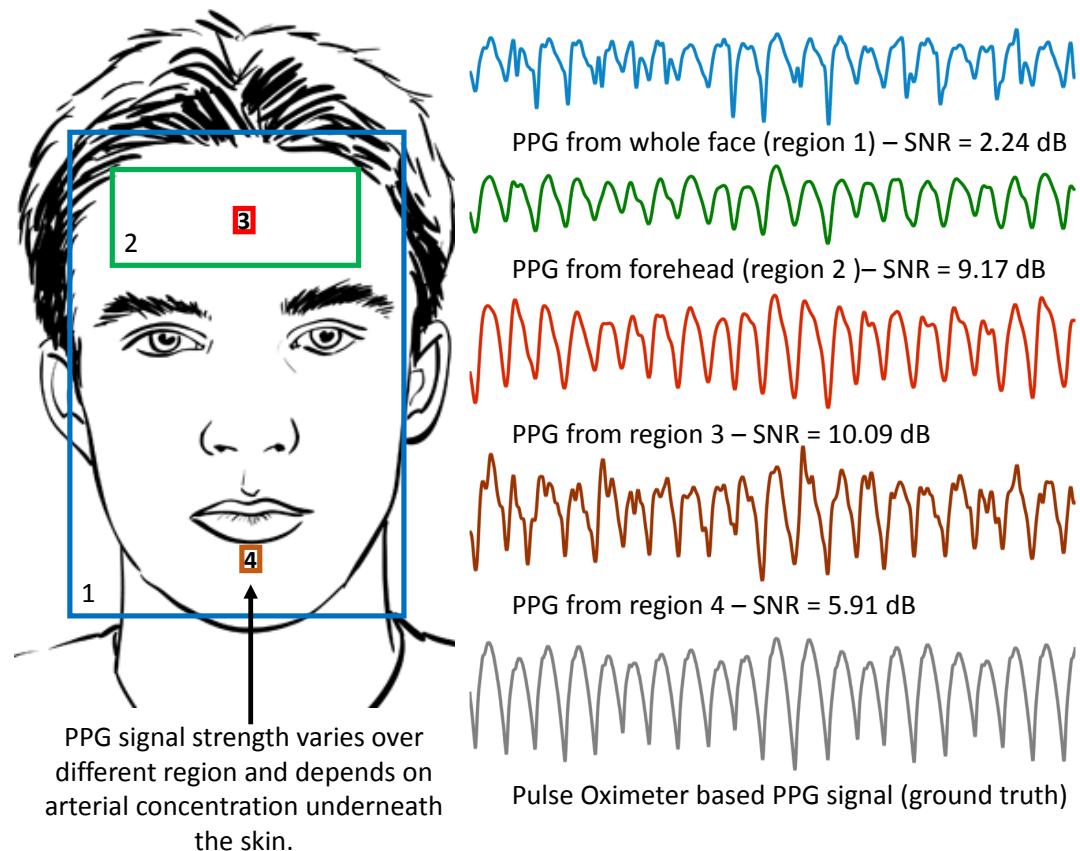


Figure 3.2 : Effect of skin ROI size and location on PPG signal strength

Further, the direction of incident and reflecting light also impacts the PPG signal strength, and regions of specular (surface) reflection add significant noise. Our proposed PPG estimation algorithm, distancePPG, models these differences and automatically identifies regions having high blood perfusion and weigh signals obtained from such regions of face highly, and automatically gives less weights to regions which add more noise.

3.2 distancePPG algorithm

We first propose a camera based-PPG signal acquisition model. It captures the effect of incident light intensity, spatial differences in blood perfusion and camera noise on the quality of acquired PPG. Based on this model, we then propose an PPG estimation algorithm (labeled as *MRC algorithm*) which builds on the ideas of maximal ratio combining (MRC) to combine the PPG signals extracted from different regions. We then describe a region-based motion tracker algorithm which keeps track of different regions of the face as the person moves in front of the camera. The overall block diagram of the distancePPG algorithm is shown in Figure 3.3.

3.2.1 Camera based PPG signal model

Let us divide the imaged skin surface (e.g. face) into many small region of interests (ROI) denoted by the set \mathcal{R} , such that the blood perfusion inside each ROI can be assumed to be constant. Let us assume that $p(t)$ is the true PPG signal that we want

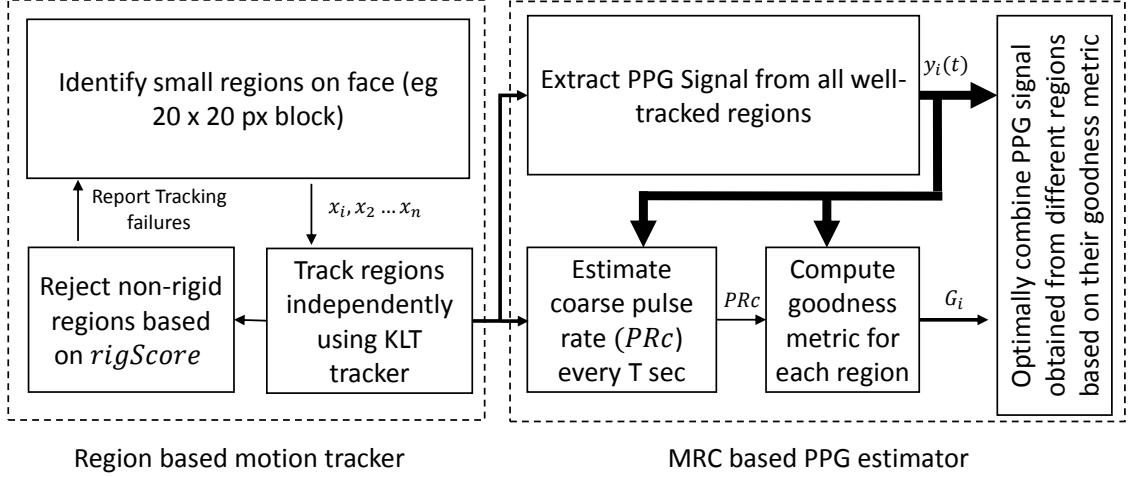


Figure 3.3 : Block diagram of distancePPG algorithm

to estimate.

Let $y_i(t)$ be the average pixel intensity from some ROI, $x_i \in \mathcal{R}, i \in \{1, 2 \dots, n\}$, on the skin at time t . Then, $y_i(t)$ can be modeled as

$$y_i(t) = \overbrace{I(x_i, t)}^{\text{Illumination}} \underbrace{\left(\underbrace{\alpha(x_i) \cdot p(t)}_{\text{PPG reflection}} + \underbrace{b(x_i)}_{\text{Surface reflection}} \right)}_{\text{Reflectance}} + \overbrace{w_i(t)}^{\text{Camera shot noise}} + \overbrace{q_i(t)}^{\text{Quantization noise}}, \quad (3.1)$$

where $I(x_i, t)$ is the incident light intensity on region x_i at time t , $\alpha(x_i)$ is the strength of the PPG in region x_i , $b(x_i)$ is the non-pulsatile surface reflection from the skin in region x_i , and $w_i(t)$ and $q_i(t)$ are camera shot noise and quantization noise, respectively.

The main aim of this work is to estimate $p(t)$ in the presence of surface reflection, $b(x_i)$ which constitutes a dominant portion of the signal captured by the camera

sensor. Quantization noise is also significant as our signal of interest, $p(t)$, varies by a small amount relative to the surface reflection, $b(x_i)$. Camera shot noise, $w_i(t)$, is dominant in darker parts of the image, and is typically caused by variations in the number of photons sensed at a given exposure level. Generally, the frequency band of interest for PPG signal, $p(t)$, is between 0.5 Hz and 5 Hz. If the person moves in front of the camera, the ROI x_i need to be tracked.

Here, $\alpha(x_i)$ represent the strength of the PPG signal obtained from selected ROI and varies over different parts of the skin and is different for different individuals as well. To improve the signal strength, we must combine the PPG obtained from different ROI in a manner which gives higher weight to those regions where the $\alpha(x_i)$ is large and noise is small, and lower otherwise. Moreover, the skin surface reflection, $b(x_i)$, which constitute a major portion of the averaged pixel intensity $y_i(t)$ at time t need to be filtered out.

3.2.2 MRC algorithm for PPG estimation

In the MRC algorithm, we divide the whole face into a set of small regions, \mathcal{R} , which are imaged by a pixel block in the camera (e.g. by a 20×20 pixel rectangular blocks). For all these regions, we extract the mean pixel intensity signal $y_i(t)$ by averaging the corresponding camera pixel region. According to our model, these could be written

as

$$\begin{aligned}
 y_1(t) &= I(x_1)[\alpha(x_1)p(t) + b(x_1)] + n_1(t), \\
 y_2(t) &= I(x_2)[\alpha(x_2)p(t) + b(x_2)] + n_2(t), \\
 &\vdots \\
 y_n(t) &= I(x_n)[\alpha(x_n)p(t) + b(x_n)] + n_n(t),
 \end{aligned} \tag{3.2}$$

where $i \in \{1, 2, \dots, n\}$ represents different ROI. Our aim is to faithfully extract $p(t)$ by combining camera recorded signals, $y_i(t)$, obtained from all these ROIs. For simplicity, we have assumed that the illumination is constant over time and only varies over regions x_i . Henceforth, we will refer to these $y_i(t)$ as *channels*.

The first step is to filter each channel using a bandpass filter [0.5Hz, 5Hz] to reject the constant skin surface reflection, $b(x_i)$, and other noise outside the band of interest. The second step is to combine these channels using the idea of *diversity combining* from wireless communications [21] which states that we weigh each channel by the SNR of the PPG signal acquired from that region and compute a weighted average of these channels.

As SNR of the PPG signals cannot be directly measured, so we devise a *goodness metric* as a substitute of SNR. We define the *goodness metric* based on the understanding that the power of the PPG signal is concentrated in a small frequency band around the pulse rate, while the system noise power can be assumed to be uniformly distributed in the passband of the filter. Thus, a ratio of power of recorded PPG signal around the PR to the power of the noise in the passband can be defined as

goodness metric, and used as a substitute for SNR.

The goodness metric is based solely on the recorded video of a person's face, and thus adapts to changes in blood perfusion for different people or changes in lighting conditions. Let $Y_i(f)$ be the *power spectral density* (PSD) of $y_i(t)$ over time duration $[0, T]$. Then goodness metric G_i is

$$G_i = \frac{\int_{C_1}^{C_2} Y_i(f) df}{\int_{B_1}^{B_2} Y_i(f) df - \int_{C_1}^{C_2} Y_i(f) df}, \quad (3.3)$$

where $[C_1, C_2]$ is a small region around the pulse Rate of the person, and $[B_1, B_2]$ is the passband of the bandpass filter, $[0.5\text{Hz}, 5\text{Hz}]$ for our case. Thus, the goodness metric, G_i , for region i will be proportional to the strength of region i 's PPG signal relative to its noise.

To define the region $[C_1, C_2]$, we would need to know the pulse rate of the person. Under ideal conditions, the largest peak in the PSD of each channel, $Y_i(f)$, would correspond to the pulse rate (PR) of the person. But due to low strength of PPG signal extracted from many regions and large noise, the largest peak could be spurious. Therefore, we detect the three largest peaks in the PSD of the PPG extracted from each channel. The frequency location of all these peaks for all channels are recorded. An estimate of pulse rate (PR_C) is obtained by computing the *mode* of these frequency locations. The basic idea is that spurious peaks would vary in their frequency location across different channels, and peak due to pulse rate would be consistent across all the channels. Thus, the mode of these frequency locations correspond to the pulse rate PR_C . For defining mode, we first estimate a histogram of the frequency locations

with suitable number of bins ensuring a resolution of 0.1 Hz in frequency.

Using the goodness metric, G_i , as the weight for the corresponding channel, we can combine the PPG signal acquired from different regions using weighted average

$$y(t) = \frac{\sum_{i=1}^n G_i \hat{y}_i(t)}{\sum_{i=1}^n G_i}, \quad (3.4)$$

where $\hat{y}_i(t)$ is the filtered version of $y_i(t)$ ([0.5Hz, 5Hz]).

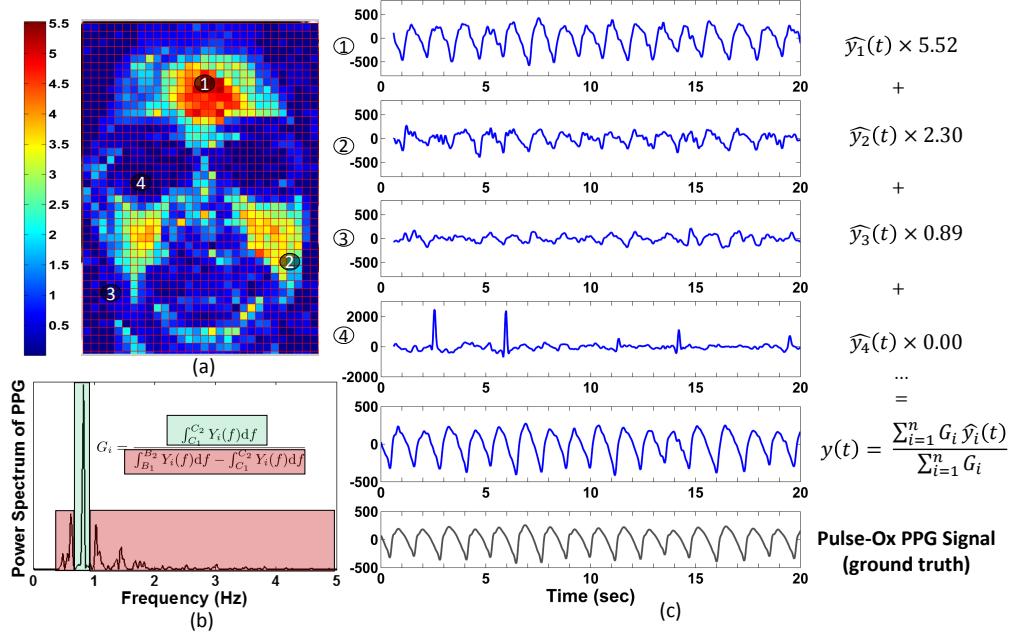


Figure 3.4 : Working of MRC algorithm: (a) shows a face with the goodness metric overlay, *Red* regions have higher goodness metric, *blue* regions have lower goodness metric, (b) illustrates the goodness metric definition based on the area under the PSD, (c) shows the PPG signal extracted from four different regions marked on the face and shows how the weighted average results in camera PPG being very similar in shape to pulse-ox PPG signal

Figure 3.4(a) shows a face with the goodness metric overlay. Regions shown as more red have higher goodness value. Figure 3.4(b) shows the PSD of acquired PPG signal and highlights how the goodness metric is defined based on the area under the PSD. Figure 3.4(c) shows the PPG signal extracted from four different regions marked on the face. When we compare these with the pulse-ox based ground truth, it is evident that the goodness metric correctly predicts the quality of the PPG signal extracted from these regions. Thus, forehead (ROI 1, Goodness = 5.52) and cheek (ROI 2, Goodness = 2.30) regions give higher strength PPG signals, whereas the region around mouth (ROI 3, Goodness = 0.89) and eyes (ROI 4, Goodness = 0.00) do not yield good signal.

As PPG signal amplitude is generally small, we also reject regions which give unusually large signal. Large variations are mostly due to motion artifacts, e.g. ROI 4 around eyes gives spiky signal due to eye movements. We reject all such regions having amplitude greater than a threshold τ_{amp} . Thus, the weight given to ROI 4 is 0.00.

3.2.3 Region based motion tracking algorithm

In camera-based acquisition of PPG, the camera and subject are completely decoupled so even a slight motion can change the relative position of selected camera's ROI and imaged skin portion. Our MRC algorithm requires PPG extraction from all the different ROIs separately, so we would need to keep track of each selected ROI. As

the subject’s face is non-planar and can also exhibit non-rigid motion (due to talking, smiling etc), we independently keep track of each $K \times K$ px region.

For this we identify $M \simeq 10 - 20$ feature points inside each ROIs which are considered *good features to track* [22]. We then use the Kanade Lucas Tomasi (KLT) feature tracker [4, 5] to track these features across the video. We use an affine transform to model the motion of each ROI between consecutive frames based on the location of the feature points inside it.

We use more feature points than is minimally required for the affine fit as feature tracking is generally error prone due to feature points changing appearance or disappearing from camera view due to occlusion. We used the method proposed in [23] to automatically detect tracking failures based on the forward-backward error and do not consider erroneous points for affine transform estimation. From the remaining well-tracked feature points, we use the random sample consensus (RANSAC) [24] algorithm to compute a robust estimate of affine fit by considering only the inlier points.

Various regions of the face can go through non-rigid shape changes (e.g region around the cheeks, mouth and forehead during smiling or talking). These regions are generally not well tracked as the affine transform model assumed would be invalid. We identify these modeling failures using a rigidity score (*rigScore*) based on the ratio of inlier count to the total number of points used for the affine fit. Here, inlier points are those which are within b pixels of the estimated affine surface. For each

frame, we do not consider those regions whose rigidity score is below a threshold τ_{rig} , and extract the PPG from the remaining well-tracked rigid regions. We combine the PPG signal from all the these well-tracked ROI using the MRC algorithm.

Chapter 4

Camera based PPG system assessment

The first goal of the experiments reported here is to characterize and quantify the performance of the two key components of distancePPG algorithm: : (i) MRC algorithm for PPG estimation (ii) region-based motion tracker, and compare them with prior methods. The first set of experiments are done by varying skin tone and light intensity while keeping the motion nearly zero and would be used to evaluate the MRC algorithm. The second experiment involved different natural motion scenario and is used to evaluate the performance of region based motion tracker.

The second goal is to study how different lighting condition and camera parameters affects the performance of camera based PPG estimation process.

4.1 Experimental setup

For all data collection in this study, we have used Flea 3[®] USB 3.0 FL-U3-13E4M-C monochrome camera operated at 30 frames per second, with a resolution of 1280 × 1024. We used TI's AFE4490SPO2EVM pulse oximeter evaluation module to record contact-based PPG signal for comparison. It operates at a sampling rate of 500 Hz. The distance between camera and subject, d , is 0.5m.

Both the camera system and the pulse-oximeter are started simultaneously, and

the data is recorded in a PC workstation. All processing is done using MATLAB[®] software with custom written codes. All experiment (except the lighting and filter experiment) is conducted under ambient (fluorescent) lighting at 600 lux brightness with a 40 nm full-width half-max (FWHM) Green filter (FB550-40 from Thor labs[®]) put on the monochrome camera.

4.2 Performance metric

To quantify the performance of our camera-based PPG estimation algorithm, we used standard PPG signal obtained from pulse oximeter connected to a person's earlobe. We chose earlobe instead of finger probe because of its proximity to the face region.

Using pulse oximeter as the benchmark will be slightly pessimistic as it also suffers from motion artifacts [25]. Moreover, the underlying PPG waveform recorded from these different modalities might not have direct correspondence as they are generally acquired from different regions of the skin (e.g., face vs. ear in our study).

The amplitude of the PPG signal recorded by pulse oximeter is unrelated to the amplitude of the camera-based PPG signal as both systems have completely independent sensor architecture and analog gain stage. So, we have defined the performance metric in a manner that it does not depend on the amplitude of the PPG waveform, but rather only depends on its temporal shape.

Let $z(t)$ denote the PPG acquired using pulse oximeter. Let $y(t)$ denote the PPG

estimated from a camera based system.

$$y(t) = A_1 p(t) + \overbrace{w(t) + q(t) + \text{motion artifact}}^{n_y(t)} \quad (4.1)$$

$$z(t) = A_2 p(t) + n_z(t) \quad (4.2)$$

Here, $n_y(t)$ is the noise in the PPG signal acquired from camera. Apart from the quantization and sensor noise, $n_y(t)$ also includes uncompensated motion artifacts. The noise present in the pulse oximeter system is denoted as $n_z(t)$. Let us assume that all signals are defined in the time interval $[0, \tau]$ and we use \mathcal{L}^2 inner product and norms for all definitions.

The noise $n_y(t)$ would be uncorrelated to $n_z(t)$ as both acquisition systems are unrelated, and both noise is uncorrelated to the underlying *zero mean* PPG signal $p(t)$. If we assume that all signals follow *ergodicity* and consider integration over large time window (τ is large), then

$$\langle n_y(t), n_z(t) \rangle \approx 0 \quad (4.3)$$

$$\langle n_y(t) \text{ or } n_z(t), p(t) \rangle \approx 0. \quad (4.4)$$

and further, we can write

$$\frac{\langle y(t), z(t) \rangle}{\langle z(t), z(t) \rangle} \approx \frac{A_1 A_2 \cdot \|p(t)\|^2}{A_2^2 \cdot \|p(t)\|^2 + \|n_z(t)\|^2} \quad (4.5)$$

Since the signal quality of PPG derived from pulse oximeter is very good, we can assume that noise power $\|n_z(t)\|^2$ is much smaller than signal power $\|p(t)\|^2$ in the denominator. Thus, $\frac{A_1}{A_2}$ can be approximated as

$$\frac{A_1}{A_2} \approx \frac{\langle y(t), z(t) \rangle}{\langle z(t), z(t) \rangle}. \quad (4.6)$$

Thus, we can estimate the noise present in the $y(t)$ as

$$n_y(t) \equiv y(t) - \frac{\langle y(t), z(t) \rangle}{\langle z(t), z(t) \rangle} \cdot z(t) \quad (4.7)$$

and, underlying signal $s_y(t)$ can be approximated to

$$s_y(t) \equiv \frac{\langle y(t), z(t) \rangle}{\langle z(t), z(t) \rangle} \cdot z(t), \quad (4.8)$$

which leads us to define signal to noise ration or SNR simply as

$$SNR \equiv \frac{\|s_y(t)\|}{\|n_y(t)\|}. \quad (4.9)$$

This SNR measure would be used to compare performance of camera based PPG estimation algorithms under various system parameters.

4.3 Experiments

4.3.1 MRC algorithm evaluation

To quantify the performance of proposed MRC algorithm in comparison to the conventional averaging method, we collected video data from 10 subjects (6 Male, 4 Female) with different skin tones (from light, pale white to dark brown/black)*. Subjects were asked to face the camera and be stationary as possible for a duration of 40 sec (involuntary motions were not restricted).

*The experiments were approved by the RICE University institute review board (IRB). Protocol number: 14 – 145E, Approval Date: 3/04/2014

For the conventional method, a rectangular box is selected encompassing the person's face (Voila Jones face detector [26] is used) and PPG signal is computed by averaging the pixel intensity inside the whole face region. For MRC algorithm, we extract PPG signal from each 20×20 px region inside the face box and then combine them.

For defining the *Goodness metric* of the PPG signal extracted from each 20×20 px ROI (eq (3.3)), we chose $C_1 = PR_C - 0.2Hz$ and $C_2 = PR_C + 0.2Hz$. Here, PR_C is the coarse estimate of the pulse rate (PR).

To quantify the performance of MRC algorithm under low lighting condition, we varied the illumination level from 25 lux upto 400 lux (ambient light is around 500lux).

4.3.2 Region-based motion tracking evaluation

We tested our region based motion tracking algorithm under three practical scenarios - (i) *Browsing Internet*, (ii) *Watching video*, (iii) *Talking on Skype*. These scenarios are reflective of a large class of motion exhibit by users of tablet, phone, or laptops while facing the screen.

For our region based motion compensation algorithm we identify the face region using Voila Jones face detector. We chose 40×40 px square block ($K = 40$) all over the face as our ROI, and identify $M = 10$ to 20 *good features* to track within each of these ROI. Larger ROI is chosen compared to stationary case as we need more feature

points to track robustly. We reject non-rigid regions of the face by choosing $b = 2.5\text{px}$ and $T_{rig} = 0.3$. These choices are made to optimize the tracking performance and to easily detect non-rigid regions in face.

We combine the PPG signal from all the remaining tracked ROI using the MRC algorithm discussed in Section 3.2.2. All computations are done over a time window of 5 sec and the system is re-initiated again after that. This involves re-detecting the face region and initiating all the ROI.

For comparison to past method, we computed 2D shifts between consecutive frames using correlation [14] and extract the PPG by averaging pixel intensity in the tracked face region.

4.3.3 System parameter evaluation

We collected video for a total of 4 lighting scenario -(i) Ambient (fluorescent) Light at 550-600 lux brightness, (ii) Custom build green LED illumination (560nm) at 250 lux, (iii) Incandescent light at 1100 lux, (iv) Near IR Light (Clover Electronics IR045 Night vision IR light) at $0.3W/m^2$. All light intensity is measured at the person's face. For first three lighting scenario, we recorded video at three different optical filter configuration - (i) 40nm FWHM Green filter (FB550-40), (ii) 10nm FWHM Green filter (FB530-10), (iii) No filter. For IR illumination, we used two configurations (i) A 680nm IR longpass filter and (ii) 40nm FWHM IR filter (FB880-40). All filters are from Thor Labs. Thus, for a total of 11 scenarios, we collected face video and pulse

oximeter recording.

4.4 Results

4.4.1 MRC algorithm performance

Figure 4.1(a) compares the SNR of the camera-based PPG signal acquired using MRC algorithm (green) and that using the conventional averaging method (red) for people having different skin tones (pale white to brown). MRC algorithm provides on an average 4.7 dB of improvement in SNR compared to current methods. This improvement is critical in darker skin tones where current methods produces very low SNR of camera PPG signal which would lead to unreliable PR estimation. A minimum SNR of -0.5 dB is needed for estimating reliable PR with less than 3 bpm error. This threshold is determined by simulation (see Section 5.3). In Figure 4.1(c), we can see that the PPG signal estimated using MRC algorithm (green) is very similar in shape to the PPG signal extracted from pulse oximeter (black). The improvement over current method (red) is significant in darker skin tones.

Figure 4.1(b) compares the SNR of camera PPG signal acquired using MRC algorithm (green) and using conventional averaging method (red) for varying illumination (25 lux to 400 lux). The gain due to MRC algorithm is around 8 dB in very low lighting conditions (50 – 100 lux) and is around 5 dB for low-medium illumination (100 – 400 lux).

This improvement in SNR helps reduce the error in pulse rate (PR) estimation

below the medically acceptable limit of 3bpm for all skin tones. We determine the PR every 5 sec by detecting the peak in the PSD of the estimated camera PPG and compared it with pulse oximeter derived PR. Figure 4.2(a) shows the RMSE of PR estimation from camera PPG using both MRC algorithm (green) and current method (red) for different skin tones.

Further, the SNR improvement makes camera-based PPG signal very similar in shape to the pulse oximeter-based PPG signal. This improves the performance of pulse rate variability (PRV) or peak-to-peak interval measurement. We determine the PRV from camera based PPG by a basic peak detection algorithm which also rejects spurious close-by peaks and compare the derived PRV with the PRV obtained from pulse oximeter. Figure 4.2(b) shows the RMSE of PRV estimate from camera PPG using both MRC algorithm (green) and current method (red) for different skin tones. Using MRC algorithm we can achieve below 20 ms accuracy in PRV estimation for majority of skin tones.

But, the SNR improvement, though significant, is not sufficient to extract other waveform parameters which depends on clearly detecting the systolic and diastolic peaks. Parameters like x , y , ΔT (refer Figure 2.2) would show significant error as is clearly evident from Figure 4.3. The systolic peak obtained from camera-based PPG (using MRC algorithm) shows occasional spikes, and many a time, the diastolic peak is completely missed by the camera-based PPG. The main reason for these errors are limited bit-width of camera ADC having only 10 ENoB (effective number of bits) for

our camera as opposed to 20 ENoB for pulse oximeter system.

4.4.2 Region-based motion tracking performance

Table 4.1 summarizes how the motion compensation algorithm performs in different scenarios (browsing, watching TV, and talking) for three different individuals (U_1, U_2, U_3). We get on an average 5.33 dB improvement in SNR by using distancePPG (region based motion tracker and MRC algorithm) as opposed to current methods (2D tracking and conventional averaging). Also the performance of distancePPG deteriorates when compared to the stationary case, particularly for the talking scenario. This is because a large portion of the face goes through non-rigid deformation during talking and are not considered for extracting PPG. Figure 4.4 shows the estimated camera PPG waveform obtained by distancePPG (green) and current method (red) for three very different cases: (i) U_1 while browsing, (ii) U_2 while watching video, and (iii) U_3 while talking. When compared to ground truth pulse oximeter-based PPG (*black*), it is clear that distancePPG significantly improves the quality of acquired camera PPG signal.

4.4.3 System parameter assessment

Table 4.2 summarizes the SNR of the imaging PPG under various lighting and filter scenario. It is evident that 40nm FWHM green filter works better than having no filter or having 10nm FWHM green filter. This is because 40nm green filtered centered around 550nm passes only light where absorption of hemoglobin is maximum,

Motion Scenario	SNR^{U1}	SNR^{U2}	SNR^{U3}
Stationary	10.3dB	15.2dB	10.73dB
Browsing Internet	4.14dB	1.16dB	2.13dB
Watching Video	2.96dB	5.31dB	-1.84dB
Taking on Skype	-2.35dB	3.50dB	-2.87dB

(a) Region based motion tracker+MRC algorithm

Motion Scenario	SNR^{U1}	SNR^{U2}	SNR^{U3}
Stationary	7.3dB	13.6dB	5.55dB
Browsing Internet	-6.21dB	0.87dB	-2.64dB
Watching Video	-5.14dB	4.07dB	-4.99dB
Taking on Skype	-4.91dB	0.52dB	-6.32dB

(b) 2D tracking method + simple averaging

Table 4.1 : Performance comparison of distancePPG with current methods under different natural motion scenarios.

Light	Filter	SNR dB	Camera Parameters
Ambient Fluorescent Light Illuminance = 550 lux	No Filter	5.99	Exp=2.68ms, Gain=0.4dB
	Green Bandpass (FWHM=40nm)	10.05	Exp=14ms, Gain=0.4dB
	Green Bandpass (FWHM=10nm)	-4.95	Exp=33.3ms, Gain=18dB
Incandescent Lamp Illuminance = 1100 lux)	No Filter	0.67	Exp=0.42ms, Gain=0.4dB
	Green Bandpass (FWHM=40nm)	9.91	Exp=16ms, Gain=0.4dB
	Green Bandpass (FWHM=10nm)	8.14	Exp=33.3ms, Gain=9.54dB
Green LED Light Illuminance = 250 lux	No Filter	0.61	Exp=9ms, Gain=0.4dB
	Green Bandpass (FWHM=40nm)	10.37	Exp=33.3ms, Gain=1.8dB
	Green Bandpass (FWHM=10nm)	5.13	Exp=33.3ms, Gain=9.8dB
Near IR LED Light IR Power = $0.3W/m^2$	Bandpass (FWHM=40nm)	1.72	Exp=10.6ms, Gain=0.4dB
	IR Longpass (680nm onwards)	-0.18	Exp=4.0ms, Gain=0.4dB

Table 4.2 : Lighting conditions and filters

and rejects irrelevant light, allowing more camera exposure and analog gain without saturating the sensors. The 10 nm FWHM green filter gives worse performance as the amount of light entering the camera sensor is very low in this case, resulting in sensor photon noise. Further, we found that Green LED light (530nm) works much better than IR LED light for acquiring imaging PPG. This is fundamentally different from a pulse oximeter operation where IR and RED light (660nm, 880nm) has conventionally been preferred.

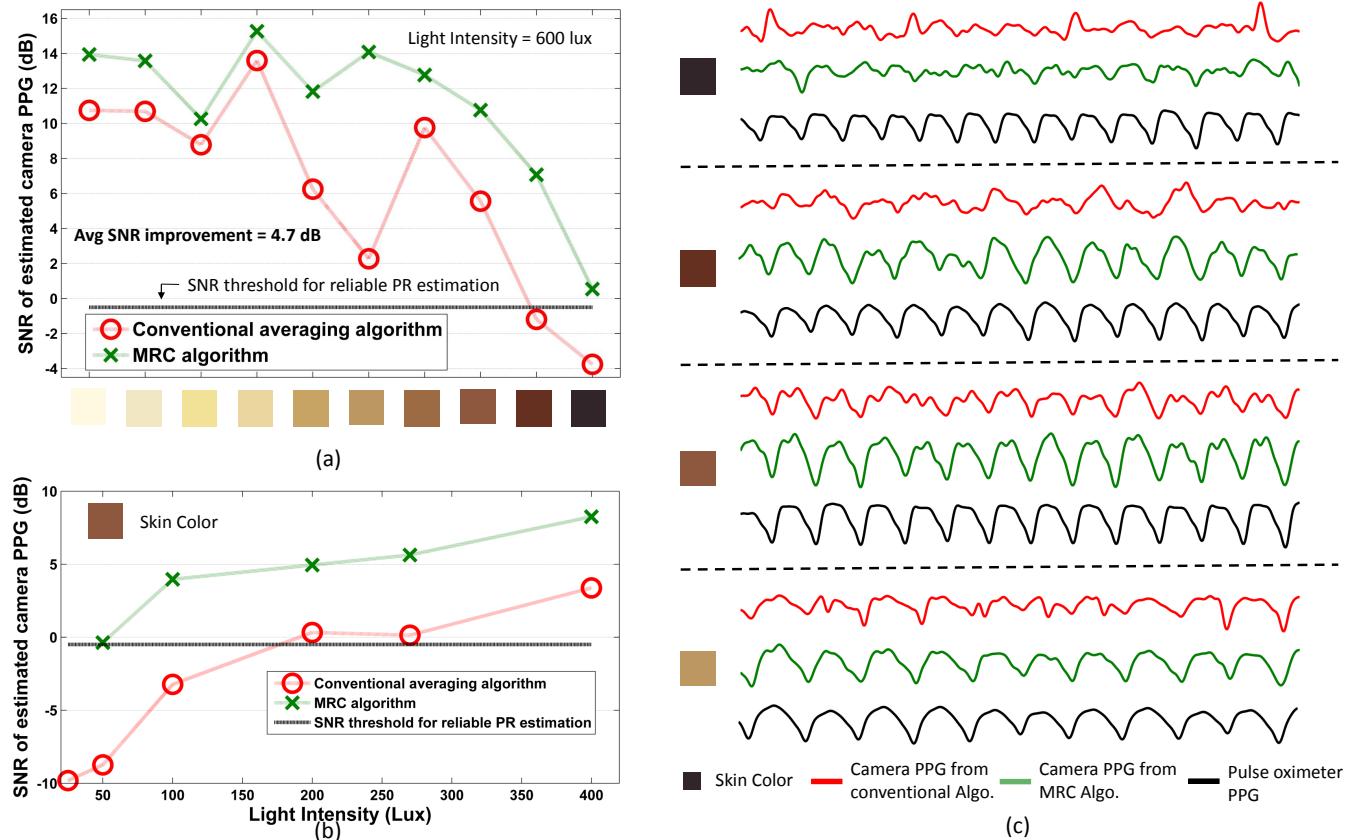


Figure 4.1 : Performance of MRC algorithm in comparison to conventional averaging method: (a) shows improvement in SNR of camera PPG for different Skin tones (pale white to brown), (b) shows improvement in SNR of camera PPG under different illumination level, (c) shows the actual PPG waveform obtained using MRC algorithm and conventional averaging, and compares it with pulse oximeterbased PPG signal.

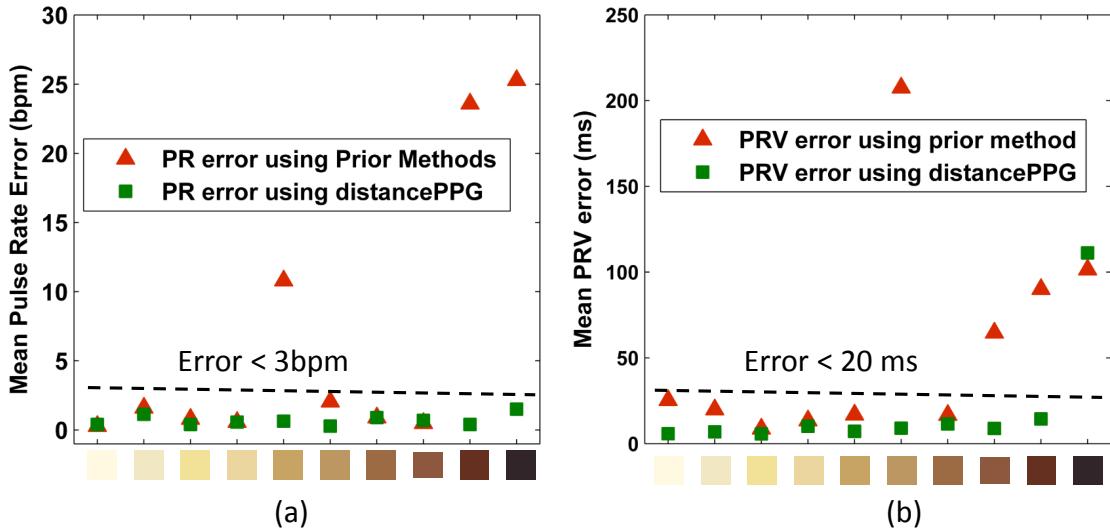


Figure 4.2 : PR and PRV estimation performance (a)Using distancePPG reduces error in PR estimation below 3bpm for all skin tones. (b) distancePPG reduces error in PRV below 20 ms for majority of skin tones.

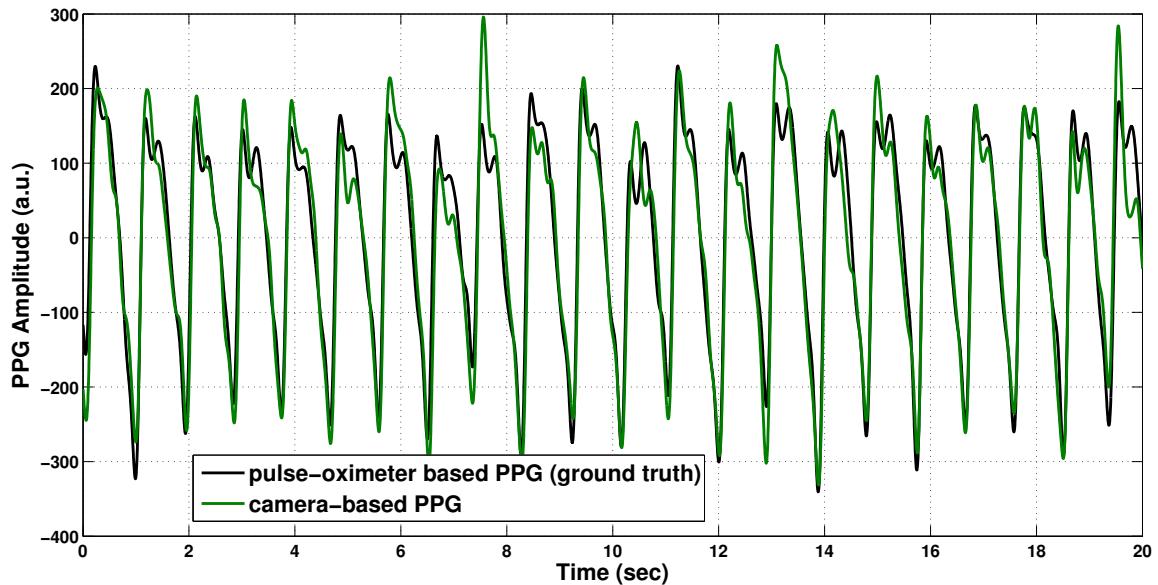


Figure 4.3 : Comparing PPG shape for detecting waveform feature

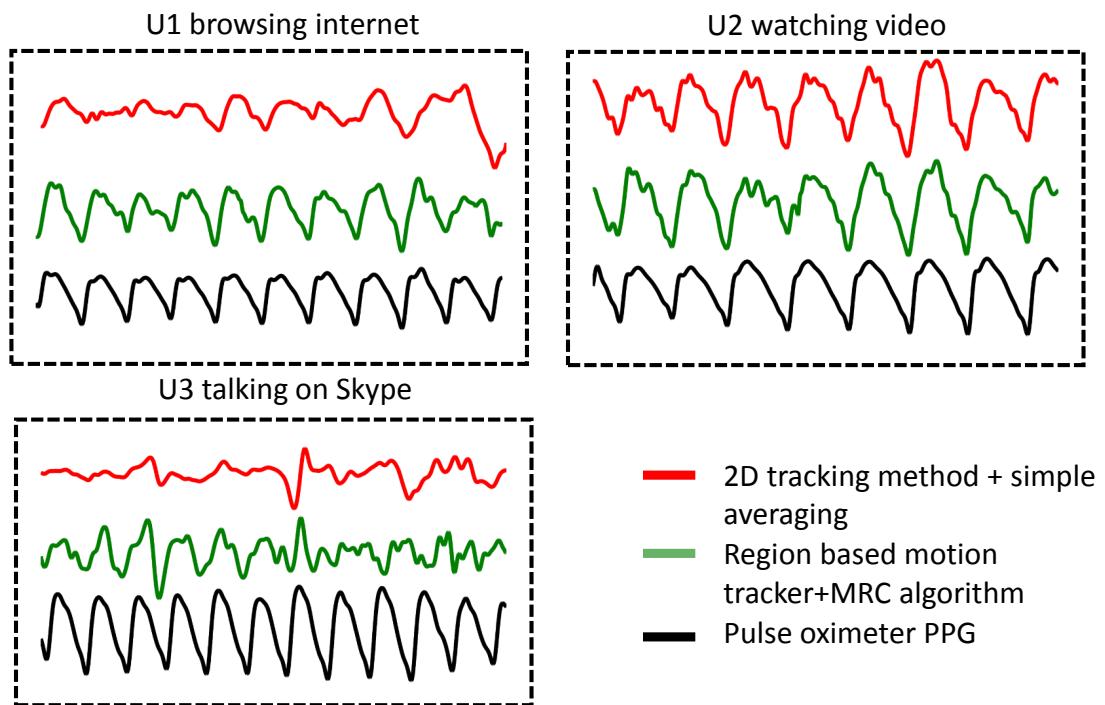


Figure 4.4 : Comparison of camera PPG waveform using 2D tracking method (*red*) and distancePPG (*green*) with the pulse oximeter-based PPG (*black*) under three scenario - (i) U1 browsing internet, (ii) U2 watching video, (iii) U3 talking on skype

Chapter 5

Discussion and Conclusion

5.1 Goodness metric and SNR

The goodness metric, G_i , defined for PPG signal obtained from some region in the face was based on the ratio of areas under the power spectrum density (PSD), $Y_i(f)$, i.e.

$$G_i = \frac{\int_{C_1}^{C_2} Y_i(f) df}{\int_{B_1}^{B_2} Y_i(f) df - \int_{C_1}^{C_2} Y_i(f) df}, \quad (5.1)$$

where $C_1 = PR_C - 0.2$ Hz and $C_2 = PR_C + 0.2$ Hz, PR_C is the coarse estimate of the pulse rate, $B_1 = 0.5$ Hz, $B_2 = 5$ Hz as discussed in Chapter 3.

The goodness metric has been used as a substitute for the SNR of the camera based PPG signal obtained from a region. Ideally, it should be equal to the ratio of the power of underlying PPG signal upon the power of noise present in that region. As the power of PPG signal is unknown, our definition of goodness metric is only an approximation, and it would be worthwhile to evaluate experimentally how good that approximation is.

For six subjects of varying skin tones (pale white to brown/black), Figure 5.1 shows the scatter plot between the goodness metric, G_i , computed for the camera PPG signal, $\hat{y}_i(t)$, obtained from different regions of the face ($x_i \in \mathcal{R}$) and the corresponding

SNR obtained by comparing the regional camera PPG signal ($\hat{y}_i(t)$) and the pulse oximeter-based PPG signal (ground truth).

The scatter plot clearly shows that goodness metric is a good substitute for SNR in regions having SNR greater than 0 dB (PPG signal power greater than noise power) for all skin tones. For regions having lower SNR, goodness metric overestimates the SNR significantly, particularly for regions having SNR lower than -3 dB. Thus, it would be desirable to not include regions having very low goodness metric in the weighted average equation (3.4) because their weights could not be faithfully determined. Thus, one needs to set up a threshold, G_{thres} , below which the corresponding regions would not be included in the weighted average. In the next section, we will experimentally determine the choice of the threshold.

5.2 MRC and choice of G_{thres}

To understand how the choice of G_{thres} affects the resultant SNR of the MRC PPG, we sorted PPG obtained from all the different regions of the face ($\hat{y}_i(t)$) according to their goodness score (G_i) and computed a cumulative weighted averaged PPG (y_{cum_n}) starting from regions having highest goodness metric and progressively including regions having lower goodness metric, i.e.

$$y_{cum_n}(t) = \frac{\sum_{\substack{i_{sort}=1^{st} \\ n^{th}}}^{n^{th}} G_i \hat{y}_i(t)}{\sum_{\substack{i_{sort}=1^{st}}}^{n^{th}} G_i} \quad (5.2)$$

Figure 5.2 shows the plot of SNR of the cumulative MRC PPG (top) and the

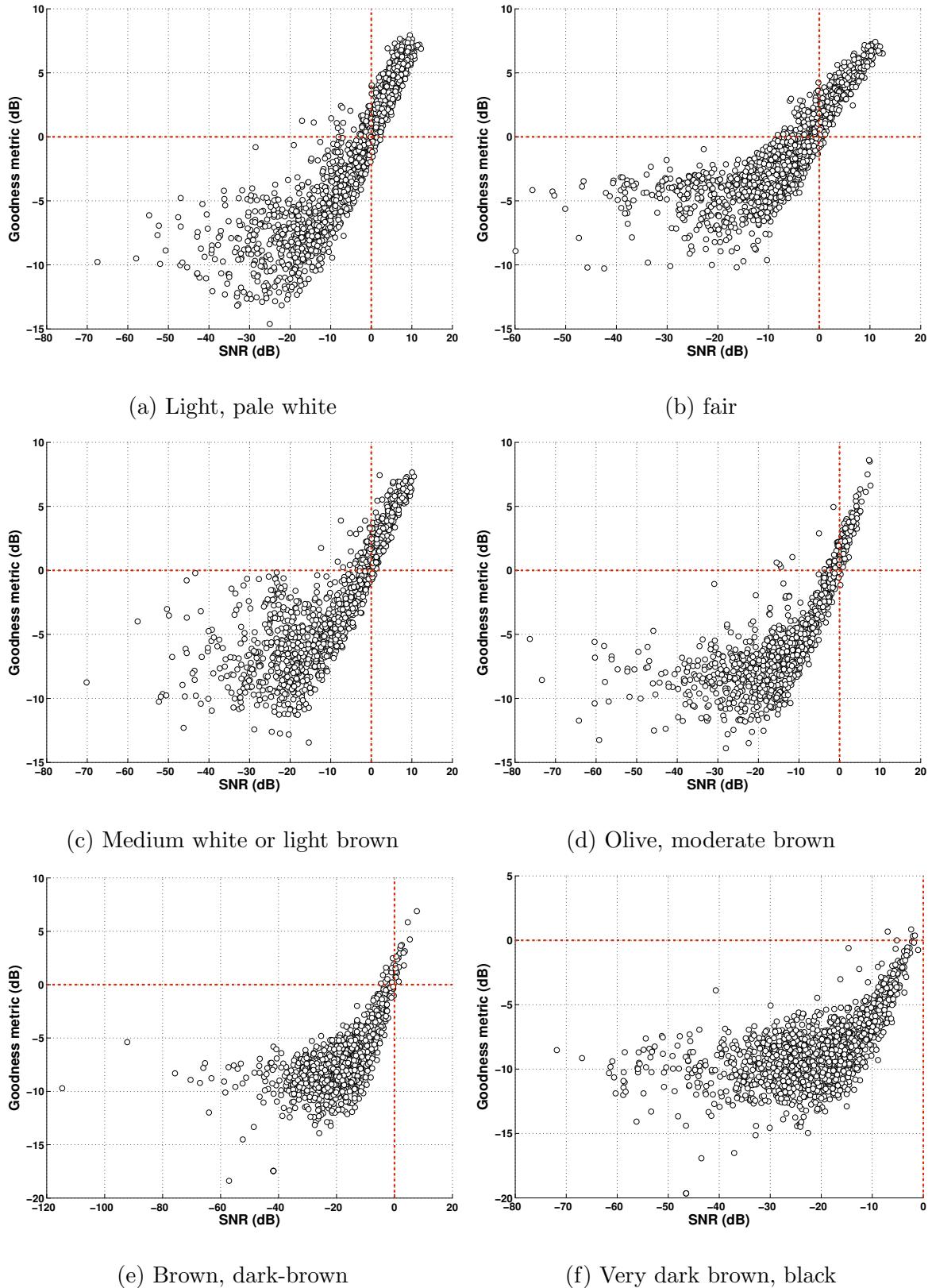


Figure 5.1 : Scatter plot between goodness metric (dB) and SNR (dB)

corresponding G_{thres} (bottom) for people having varying skin tones (pale white to brown/black).

In Figure 5.2, we can see a general trend visible in the SNR of the cumulative MRC PPG - the SNR first increases as we add good regions, and then it starts reducing as bad regions are added.

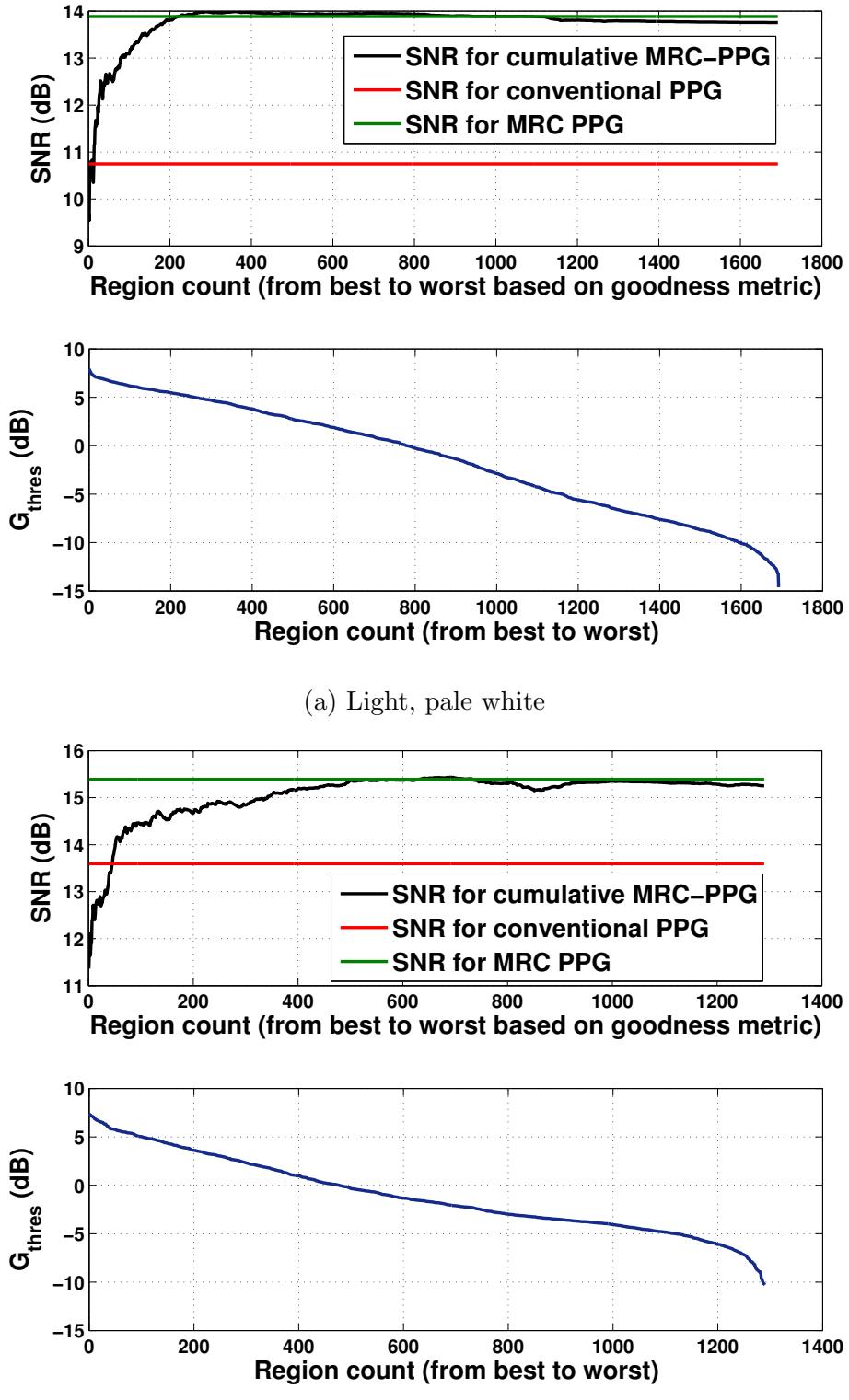
The general trend visible in Figure 5.2 clearly highlights two underlying aspects of the MRC algorithm:

1. Adding PPG obtained from good regions overcomes the inherent quantization noise due to camera ADC. Thus, the SNR increases initially when we add additional good regions.
2. As the goodness metric is an overestimation of the SNR, adding not-so-good regions starts reducing the overall SNR.

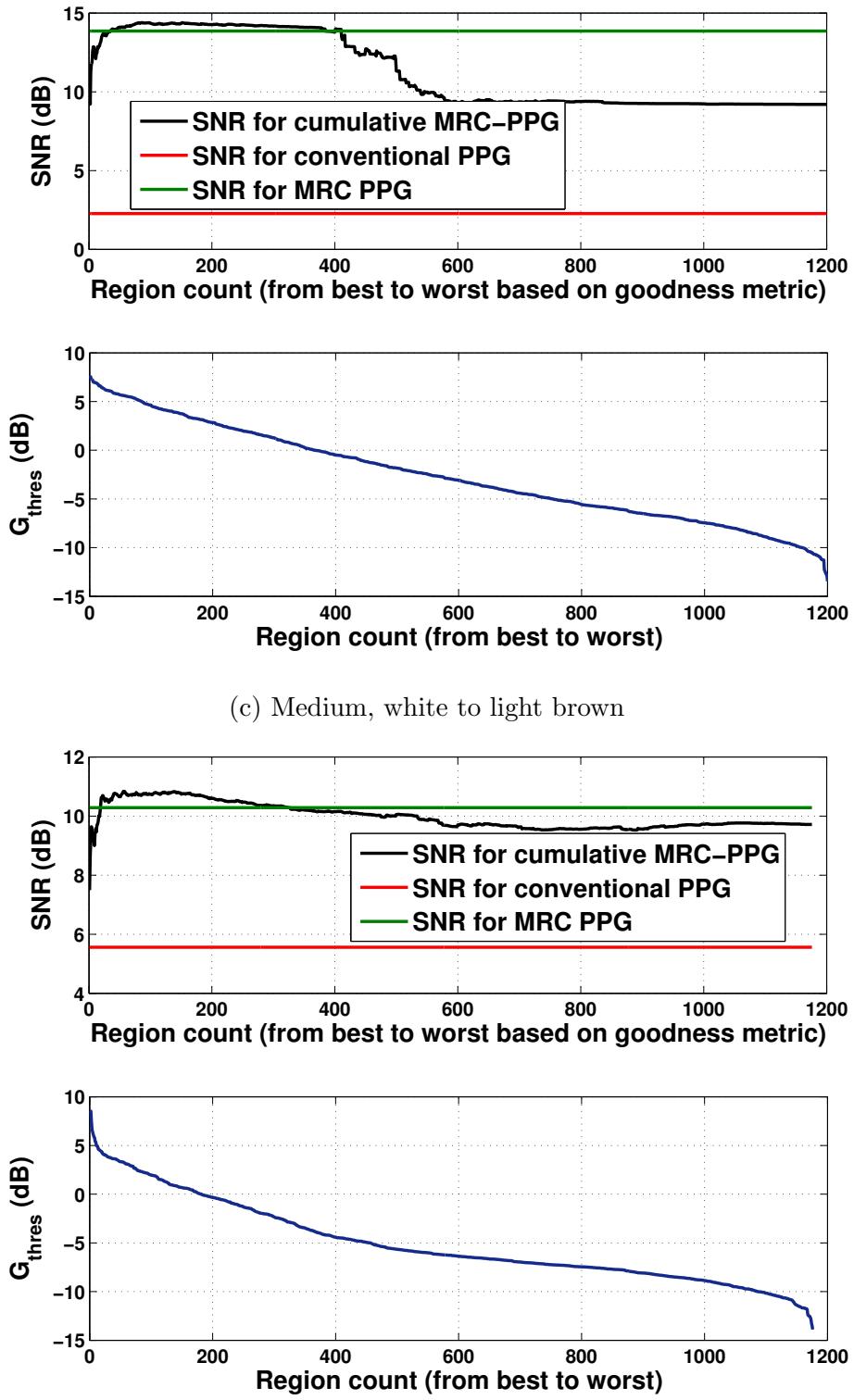
Looking at plots in Figure 5.2 for different skin tones, it is evident that for fairer skin tones most of the regions are good and we do not face the SNR-overestimation problem. This is evident from the fact that SNR keeps on increasing for fairer skin tone as we keep on adding additional regions.

For moderate brown skin tones, the SNR initially increases due to reduction in quantization error, but then starts reducing midway as new regions add more noise.

For darker skin tones, it is evident that just picking the best region is a good approach as the SNR starts reducing from the very beginning. The reduction is due to the lack of many good regions in darker skin tones.



(b) Fair



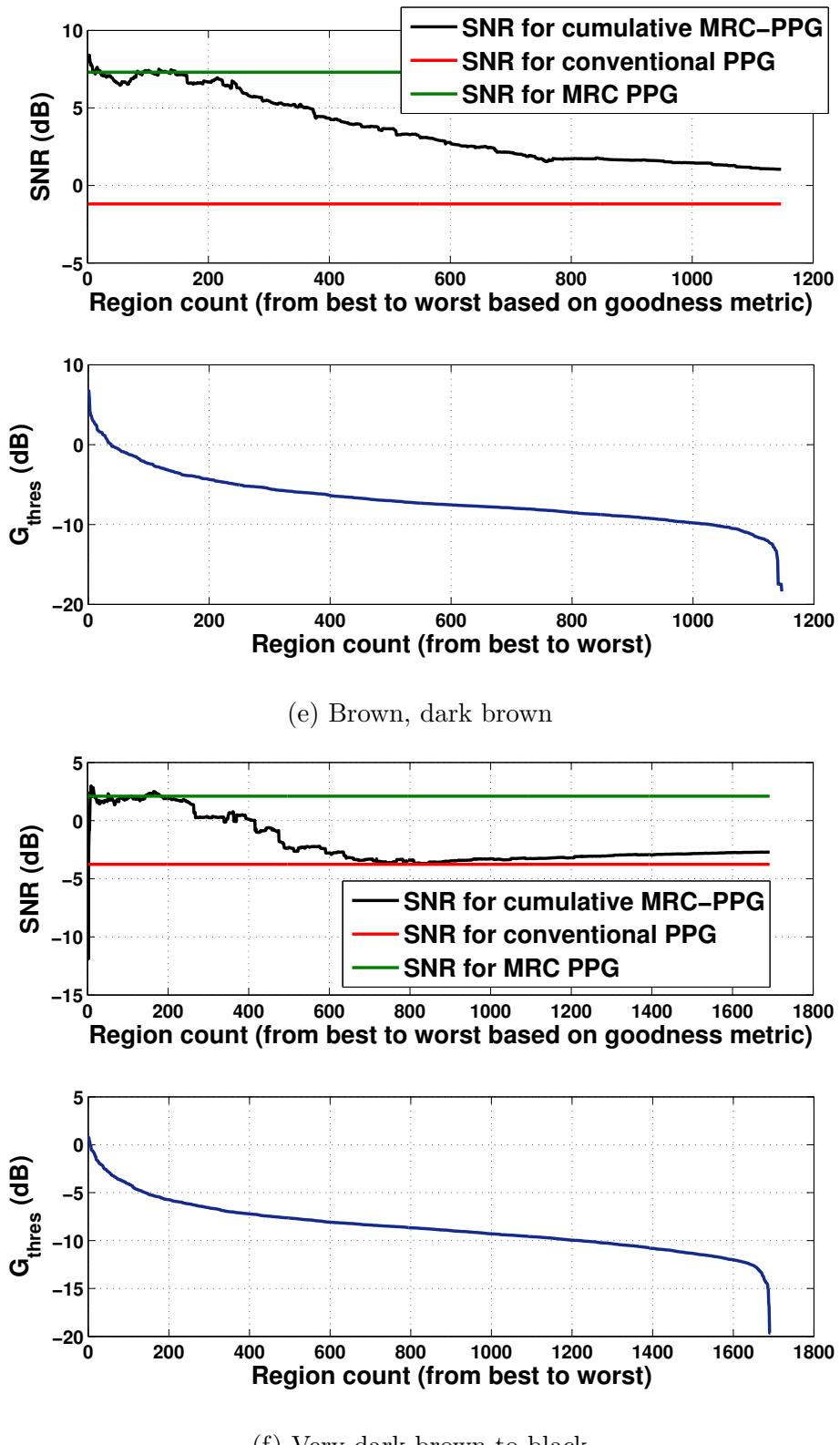


Figure 5.2 : SNR-vs-Threshold - Various skin tones

Barring experimental variations, another trend evident from the results shown in Figure 5.2 is that the SNR starts reducing when we add regions having goodness metric lower than -3dB irrespective of skin tones. Thus, for all our previous experiments, we selected the threshold for goodness metric (G_{thres}) to be 0.5 or -3 dB , and did not consider regions having lower goodness metric for extracting PPG.

5.3 SNR threshold for reliable PR

For being medically relevant, any camera-based PPG system should provide accuracy better than 3bpm in PR estimation. To translate this requirement to SNR of the estimated camera PPG signal, we simulated camera PPG signal at various SNR level by adding bandlimited [0.5Hz, 5Hz] white Gaussian noise to the pulse oximeter derived PPG signal (ground truth). We then computed the error in PR estimation for 10000 different simulation of the noise (all pulse rate are estimated over 5 sec window). Figure 5.3 shows the dependence of error in pulse rate of camera PPG on the SNR.

It is evident from simulations that one would require SNR better than -0.5 dB to get an accuracy below 3 bpm. This clearly highlights the critical importance of the SNR gain of 4.7 dB by using MRC algorithm for different skin tone people. These simulations are consistent with the PR estimated for 10 different individuals having diverse skin tones as discussed in the result section (4.2(a)). As shown in 4.1(b), the SNR of the camera PPG obtained using MRC algorithm is above -0.5 dB for illumination as low as 50 lux. Thus, it makes a strong case for using distancePPG

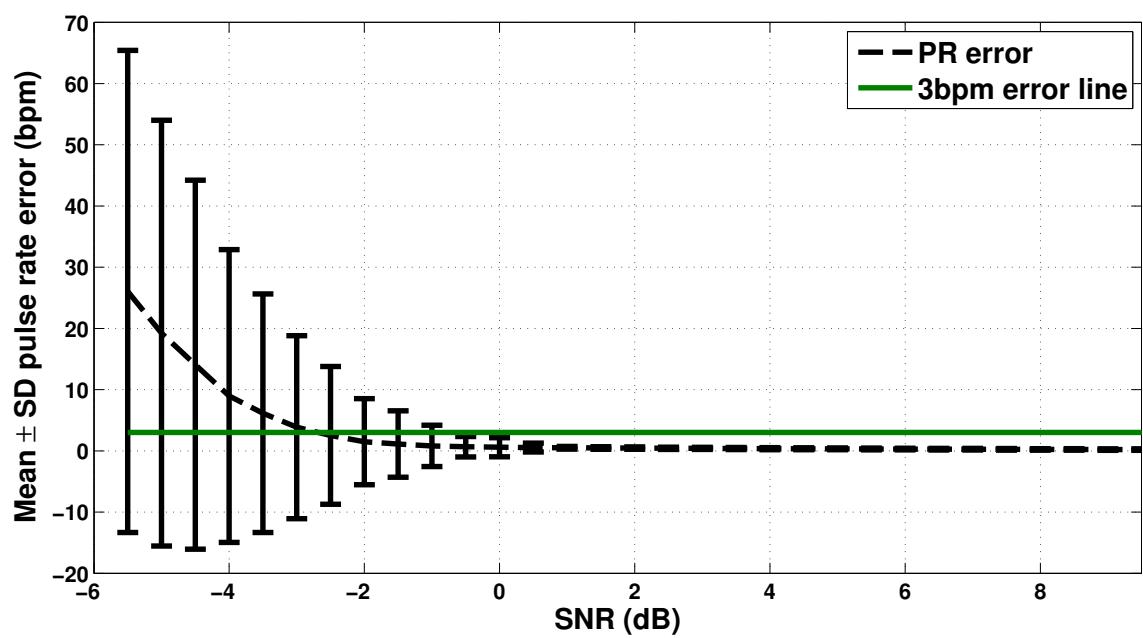


Figure 5.3 : Dependence of pulse rate error on SNR of camera PPG determined by simulations

under low lighting scenarios like neonatal ICU for detecting PR of new born babies [3].

5.4 PRV estimation accuracy using camera PPG

Generally speaking, PPG signal obtained from different regions of the skin would exhibit varying delays as the blood reaches these regions at different times. For example, there is a time lag of 80 ms in PPG recorded between finger and toe [27]. However, the time lag between the PPG signal obtained from close-by regions are small, e.g. regions inside the face or hand. Our own measurements show that the delay is less than 10 ms between farthest point in face. As MRC based method averages PPG signal obtained from different regions, the accuracy of the derived PRV can not be any better than 5 – 10 ms even if SNR is improved further. This is a limitation of any camera-based PPG estimation algorithm which relies on averaging to improve quantization noise. A reduction in SNR can be traded for improved accuracy in PRV estimation by using smaller ROI.

Another limiting factor is the frame rate of the camera used (30 Hz for our experiments) for PPG acquisition. The PPG acquired using pulse oximeter is generally sampled at 500 – 1000 Hz. Even if one interpolates the acquired camera PPG signal (we interpolated camera PPG to 500 Hz using spline interpolation), high frequency details can not be recovered because of aliasing. One solution would be to use very high frame rate camera (e.g., 500 Hz). Since exposure time of higher frame rate camera would be very small (e.g. 2 ms), this could reduce the SNR of the estimated

camera PPG signal, making it ineffective for detecting pulse minima essential for PRV estimation. Thus, again a reduction in SNR can be traded for improved accuracy of PRV estimation by using higher frame rate cameras.

5.5 Limitation of distancePPG under motion

Even if we track the face faithfully using region-based motion tracker or some state-of-art non-rigid faced tracking method, there are other challenges present in realistic scenarios which would make motion tracking ineffective. First, most light sources have a spatial illumination gradient and when a tracked region moves in space, it leads to change in incident light intensity in that region over time. Such changes could corrupt the PPG extraction process.

The camera based PPG acquisition system proposed in this thesis do not model changes in skin reflectance due to motion. Motion of person relative to camera can lead to change in angle of the skin surface relative to the light source, or change in the angle of reflected light captured by the camera. Such change in angles lead to large changes in skin surface reflectance, and is quantified by the bidirectional reflection distribution function (BRDF) of the skin surface. This can completely overwhelm our PPG acquisition process, as we are not modeling the BRDF related changes and is an open challenge which need to be solved to truly realize the potential of camera PPG for in-situ non-obtrusive vital sign monitoring. This could be one of our future research goals.

5.6 Conclusion

In this thesis, we dived deeper into estimating the whole PPG signal from a camera-based system, and demonstrated that distancePPG can lead to a significant improvement in overall performance compared to prior systems. The key to the performance improvement is the new goodness metric, which captures the inherent quality of the signal from a particular region of the skin.

Two popular mobile phone app for measuring non-contact pulse rate (PR) using the color change (or PPG) signal from a persons' face are Philips Vital sign camera [28] and Cardiio [29]. Both these apps require users to be at rest facing the camera in a well lit environment to be effective [29]. The distancePPG algorithm discussed in this paper address these challenges and thus would extend the use case of mobile phone and computer apps for vital sign monitoring.

Our future work includes porting our code to popular mobile platforms (Android/iOS), and open-source the mobile codebase for wider use in actual research project.

Bibliography

- [1] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, Dec. 2008, PMID: 19104573 PMCID: PMC2717852. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2717852/>
- [2] M.-Z. Poh, D. McDuff, and R. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [3] L. A. M. Aarts, V. Jeanne, J. P. Cleary, C. Lieber, J. S. Nelson, S. Bambang Oetomo, and W. Verkruysse, “Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - a pilot study,” *Early human development*, vol. 89, no. 12, pp. 943–948, Dec. 2013, PMID: 24135159.
- [4] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” 1981, p. 674679.
- [5] C. Tomasi and T. Kanade, “Detection and tracking of point features,” Carnegie Mellon University, Technical Report MU-CS-91-132, 1991.
- [6] J. Allen, “Photoplethysmography and its application in clinical,” *Physiological*

- Measurement*, vol. 28, no. 3, p. R1, Mar. 2007. [Online]. Available: <http://iopscience.iop.org/0967-3334/28/3/R01>
- [7] F. P. Wieringa, F. Mastik, and A. F. W. van der Steen, “Contactless multiple wavelength photoplethysmographic imaging: a first step toward ”SpO₂ camera” technology,” *Annals of biomedical engineering*, vol. 33, no. 8, pp. 1034–1041, Aug. 2005, PMID: 16133912.
- [8] K. Humphreys, T. Ward, and C. Markham, “Noncontact simultaneous dual wavelength photoplethysmography: a further step toward noncontact pulse oximetry,” *The Review of scientific instruments*, vol. 78, no. 4, p. 044304, Apr. 2007, PMID: 17477684.
- [9] J. Zheng, S. Hu, V. Azorin-Peris, A. Echiadis, V. Chouliaras, and R. Summers, “Remote simultaneous dual wavelength imaging photoplethysmography: a further step towards 3-d mapping of skin blood microcirculation</title>;” F. S. Azar and X. Intes, Eds., Feb. 2008, pp. 68500S–68500S–8. [Online]. Available: http://spie.org/x648.html?product_id=761705
- [10] M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” *Optics Express*, vol. 18, no. 10, pp. 10762–10774, May 2010. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-18-10-10762>
- [11] A. A. Kamshilin, S. Miridonov, V. Teplov, R. Saarenheimo, and E. Nippolainen,

- “Photoplethysmographic imaging of high spatial resolution,” *Biomedical Optics Express*, vol. 2, no. 4, pp. 996–1006, Mar. 2011, PMID: 21483621 PMCID: PMC3072138. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3072138/>
- [12] B. D. Holton, K. Mannapperuma, P. J. Lesniewski, and J. C. Thomas, “Signal recovery in imaging photoplethysmography,” *Physiological Measurement*, vol. 34, no. 11, pp. 1499–1511, Nov. 2013. [Online]. Available: <http://m.iopscience.iop.org/0967-3334/34/11/1499>
- [13] M. Lewandowska, J. Ruminski, T. Kocejko, and J. 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)*, Sep. 2011, pp. 405–410.
- [14] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu, “Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise,” *Journal of biomedical optics*, vol. 16, no. 7, p. 077010, Jul. 2011, PMID: 21806290.
- [15] S. Hu, J. Zheng, V. Chouliaras, and R. Summers, “Feasibility of imaging photoplethysmography,” in *International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008*, vol. 2, May 2008, pp. 72–75.
- [16] S. Hu, V. Azorin-Peris, and J. Zheng, “Opto-physiological modeling applied

- to photoplethysmographic cardiovascular assessment,” *Journal of Healthcare Engineering*, vol. 4, no. 4, pp. 505–528, Dec. 2013. [Online]. Available: <http://dx.doi.org/10.1260/2040-2295.4.4.505>
- [17] William W Nichols, Michael F O’Rourke, and Charalambos Vlachopoulos, *McDonald’s Blood Flow in Ateries Theoreticla, Experimental and Clinical Principles*, 6th ed. CRC Press, 2011.
- [18] S. C. Millasseau, J. M. Ritter, K. Takazawa, and P. J. Chowienczyk, “Contour analysis of the photoplethysmographic pulse measured at the finger,” *Journal of hypertension*, vol. 24, no. 8, pp. 1449–1456, Aug. 2006, PMID: 16877944.
- [19] N. Selvaraj, A. Jaryal, J. Santhosh, K. K. Deepak, and S. Anand, “Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography,” *Journal of medical engineering & technology*, vol. 32, no. 6, pp. 479–484, Dec. 2008, PMID: 18663635.
- [20] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals,” *Current Cardiology Reviews*, vol. 8, no. 1, pp. 14–25, Feb. 2012, PMID: 22845812 PMCID: PMC3394104. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394104/>
- [21] John G. Proakis and Masoud Salehi, *Communication Systems*, 2001, no. 0130617938.

- [22] J. Shi and C. Tomasi, “Good features to track,” in *1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94*, Jun. 1994, pp. 593–600.
- [23] Z. Kalal, K. Mikolajczyk, and J. Matas, “Forward-backward error: Automatic detection of tracking failures,” in *2010 20th International Conference on Pattern Recognition (ICPR)*, Aug. 2010, pp. 2756–2759.
- [24] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, p. 381395, Jun. 1981. [Online]. Available: <http://doi.acm.org/10.1145/358669.358692>
- [25] M. T. Petterson, V. L. Begnoche, and J. M. Graybeal, “The effect of motion on pulse oximetry and its clinical significance,” *Anesthesia and analgesia*, vol. 105, no. 6 Suppl, pp. S78–84, Dec. 2007, PMID: 18048903.
- [26] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001*, vol. 1, 2001, pp. I–511–I–518 vol.1.
- [27] M. Nitzan, B. Khanokh, and Y. Slovik, “The difference in pulse transit time to the toe and finger measured by photoplethysmography,” *Physiological measurement*, vol. 23, no. 1, pp. 85–93, Feb. 2002, PMID: 11876244.

[28] “Philips vital signs camera,” <http://www.vitalsignscamera.com/>.

[29] “Cardiio,” <http://www.cardiio.com/>.