# REMOTE HEART RATE ESTIMATION USING CONSUMER-GRADE CAMERAS

by

Nathan E. Ruben

A report submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Electrical Engineering

Approved:

| | |
|---|---|
| Dr. Jacob Gunther<br>Major Professor | Dr. Todd Moon<br>Committee Member |

Dr. Donald Cripps
Committee Member

UTAH STATE UNIVERSITY
Logan, Utah

2015

# Abstract

Remote Heart Rate Estimation Using Consumer-Grade Cameras

by

Nathan E. Ruben, Master of Science

Utah State University, 2015

Major Professor: Dr. Jacob Gunther
Department: Electrical and Computer Engineering

There are many ways in which the remote non-contact detection of the human heart rate might be useful. This is especially true if it can be done using inexpensive equipment such as consumer-grade cameras. Many studies and experiments have been performed in recent years to help reliably determine the heart rate from video footage of a person. The methods have taken an analysis approach which involves temporal filtering and frequency spectrum examination. This study attempts to answer questions about the noise sources which inhibit these methods from estimating the heart rate. Other statistical processes are examined for their use in reducing the noise in the system. Methods for locating the skin of a moving individual are explored and used with the purpose for acquiring the heart rate. Alternative methods borrowed from other fields are also introduced to find if they have merit in remote heart rate detection.

(76 pages)

# Public Abstract

Remote Heart Rate Estimation Using Consumer-Grade Cameras

by

Nathan E. Ruben, Master of Science

Utah State University, 2015

Major Professor: Dr. Jacob Gunther
Department: Electrical and Computer Engineering

This research explores the questions related to finding a person's heart rate using a video camera. This field is one which has a potentially large number of applications, but also has a large number of problems that need to be addressed. Simple and low-complexity signal processing techniques are studied to see how well they can detect heart rate on a variety of video samples. These are also explored alongside ways in which a person's face can be detected and tracked. Alternative methods are also proposed which take a different approach on this challenging problem.

To my long-suffering wife

# Acknowledgments

The topic for my research came about when my first-born son, Hyrum, was born. Sarah, my wife, was filled with first-time mother anxieties and so my first priority was to mitigate her concerns for our son's wellbeing at night. This came about in the form of building a baby monitor prototype which could detect a child's heart rate using a web camera. Without her admirable attention to our son I would have never pursued this exciting work. I thank her for being the conscientious mother that she is.

I also thank Dr. Don Cripps for his contribution to this endeavor. He may not remember, but it was he who first introduced me to the idea that heart rate could, indeed, be found through camera optics. Don taught me and my peers that engineering is an adventure that is as thrilling as you make it. He counseled me to understand the fundamentals of engineering rather than blindly apply tried-and-true techniques whose utility is only weakly grasped by the user. I have immensely enjoyed being tutored by Don.

The merit of my research would have only been mediocre at best if it were not for my major professor, Dr. Jake Gunther. It has been an extreme pleasure working with Jake in analyzing and designing techniques that would help further the global progress in this area. Jake was responsible for many of the key points and ideas found in our study. Jake's out-of-the-box thinking persuaded me that there is more than one way to think about the challenges related to heart rate estimation. His mentor-ship and friendship are invaluable to me.

I finally must thank the Electrical and Computer Engineering department at Utah State University. I have been given the treasured opportunity to learn and discover the many facets of engineering from very skilled professionals. The faculty has been outstanding and I feel I could not have received better instruction at any other institution. I also thank the department and the late David G. Sant for their contribution to my research; making it possible by financing the necessary equipment and providing the workspace needed. I count the department's hospitality and expended funds in my behalf as a debt I hope one

day to repay many fold.

Nathan Ruben

# Contents

# List of Tables

# List of Figures

# Acronyms

| | |
|---|---|
| ROI | Region-of-Interest |
| ICA | Independent Component Analysis |
| BSS | Blind Source Separation |
| SNR | Signal-to-Noise Ratio |
| FPS | Frames per Second |
| RGB | Red Green and Blue |
| MA | Moving Average |
| AR | Auto Regressive |
| FFT | Fast Fourier Transform |
| DFT | Discrete Fourier Transform |
| BPM | Beats per Minute |
| PSD | Power Spectral Density |

# Chapter 1

# Introduction

Over the years there have been many sophisticated techniques for observing the biological processes of the human body. These observations range from simple yet fundamental functions like breathing to the complex chemical firings of the nervous system. One of the most useful of these metrics, with application in both medical and personal areas, is heart rate.

Heart rate can be observed by several methods. First, it can be easily measured when an individual places the tips of their fingers against a large artery and counts the pulses of blood coming from the heart. It can also be detected more autonomously by measuring the electrical impulses generated by the heart through a well-established method known as an electrocardiography (ECG). Another common method used in pulse oximeters is photoplethysmography (PPG) where the use of optics helps determine the change of the arteries due to heartbeats.

These methods all require some form of physical contact. There are instances, however, when physical contact is undesirable. For example, in medical practice, burn victims have skin which is too sensitive to attach probes necessary for measuring their vitals. Another inconvenient scenario is when trying to track heart rate on exercise equipment. There are even cases where it is best that the subject being monitored without their knowledge. This might arise when tracking people in a high security area looking for suspicious persons exhibiting unusual anxiety.

In this research, different methods and analysis techniques for remote heart rate estimation will be explored. These methods will rely completely on 8-bit CMOS image sensors for data acquisition. Each approach will be applied in ideal and practical scenarios and, coupled with simple decision-makers, will be used as heart-rate estimators.

## 1.1 Prior Work

One noteworthy technique for remote heart rate detection is discussed by Obeid *et al.*, in which the person is subjected to a series of low-power electromagnetic pulses [1]. The reflected waves from the chest cavity are received and then processed using the well-known principle of the Doppler effect, which helps determine which direction and how fast an object was moving. In this case, the objects moving are the breathing lungs and the beating heart.

Another validated but vastly different approach is to use thermal imaging in order to detect small changes in skin temperature due to heart beat as demonstrated by Yang *et al.* [2]. One great advantage of this idea is that ROI selection is relatively simple. The skin, which stands out due to the black-body radiation, stands out from the environment, providing its own video segmentation. While the convenience of this method is promising, the technology is not financially and, in some cases, not physically feasible. These methods, due to cost and physical constraints, are only usable in very narrow fields.

### 1.1.1 Remote Photoplethysmography

The approach seemingly best-suited for a broad set of heart rate applications is remote photoplethysmography. By way of background, photoplethysomgraphy (PPG), as defined by Challoner, is an optical measurement technique that can be used to detect blood volume changes in the microvascular bed of tissue [3]. The pulse oximeter is based upon these principles. Hemoglobin, a combination of red blood cells and plasma, absorbs varying amounts of infrared light depending on the amount of oxygen present. As blood propagates through the arteries, higher concentrations of hemoglobin associated with each heartbeat can be measured [4]. Traditionally this is done by emitting light through an appendage such as the finger or toe and measuring unabsorbed light at the other side. This process uses what is known as *transmission-mode* PPG.

Because of the close contact required for transmission-mode, there is an interest in using *reflectance-mode* PPG in a remote, contact-less setting. As the name suggests, reflectance-mode measures the light reflected from the skin as opposed to light emitted through the skin. The use of this PPG mode becomes more feasible with development of inexpensive

cameras and the expansive work done in image processing. Both Allen [5] and Scalise [6] have made significant strides in the reflection-mode area, proving that the reflected light due to the expanding and contracting blood vessels is visible to sensitive modern optics.

The work by Allen, in particular, holds great value for forming a foundation for this study. In his experiments, Allen discovered that the same sources of physiological noise such as respiration and thermoregulation are found in remote PPG as in contact PPG. Allen continued to analyze the PPG waveform of the signal in hopes of characterizing its features and how they could be used to diagnose blood-related diseases.

### 1.1.2 Introducing the Use of Consumer-Grade Cameras

Further work on the topic of clinical diagnosis performed by Verkruysse *et al.* for optically determining skin disease validated the use of consumer-grade cameras [7]. In their study, power maps were used with respect to PPG signal strength on areas of the face to determine which regions are most suitable for heart rate extraction. These power maps will differ based on different types of skin conditions. This work has added value to the use of remote PPG in the clinical world.

This is a valuable discovery, because video cameras are ubiquitous. The technology of CMOS has improved so much in the recent years that they are on par with CCD sensors in terms of quality and the resolution is extremely high. Another advantage of cameras is cost. The cell phone industry and other markets have driven the price of cameras down. This essentially means that, provided a good algorithm, a camera could be turned into a very inexpensive bio-sensor.

An obvious, but nonetheless valuable, technique demonstrated by Verkruysse *et al.* was that of spatial averaging to increase PPG SNR. This method, which will be discussed later, is what allows for a seemingly invisible signal to become very clear and obvious. Spatial averaging is employed in every other research project on remote PPG involving cameras.

In a short study performed by Sun *et al.* the quality of signal was observed between that obtained by a high-resolution camcorder and a modern, inexpensive webcam [8]. Under various lighting conditions and even with varying heart rate signals, the two devices

performed the same with little to no statistical variation in their outputs. While intuitively one would suppose a higher-performance device would produce better results, it becomes clear that signal strength increase which comes from spatial averaging has little to do with the quality of the sensor used.

### 1.1.3 Blind Source Separation

In spite of these ground-breaking discoveries, the art of estimating heart rate from video was not fully appreciated until a popular paper produced by Poh *et al.* outlined a method using *independent component analysis* (ICA) as a means of dramatically increasing the *signal-to-noise ratio* (SNR) [9]. Poh treated the three time-series RGB channels as a mixed signal. By applying blind source separation (BSS), as seen in figure 1.1, a mixing matrix is found which decorrelates the three signals as much as possible, one of which is representative of the PPG signal.

From experimentation elsewhere and in this study, it has been found that Poh's method does, in fact, work very well in relatively noisy environments where the signal strength is very poor. However, in the extreme cases, when the signal is, in fact, very clean or it is heavily saturated with motion-induced noise, the algorithm does not perform very well.

They also introduced the use of face-tracking software to help reduce the noise due to head motion. They utilize the ICA approach in a heart rate tracker and dramatically decrease the variance of the heart rate estimate when observing the spectral magnitude. The utilization of motion-compensation and statistical algorithms spurred on several real-time applications some of which allowed individuals to check heart rate on their smartphone camera [10, 11].

### 1.1.4 Video Magnification

Rubenstein conducted research regarding the amplification of subtle changes in video [12]. His work was able to reveal very small movements which would have otherwise gone unnoticed by the viewer. A natural by-product of this video processing was also the amplification of color changes, such as those due to heart rate.

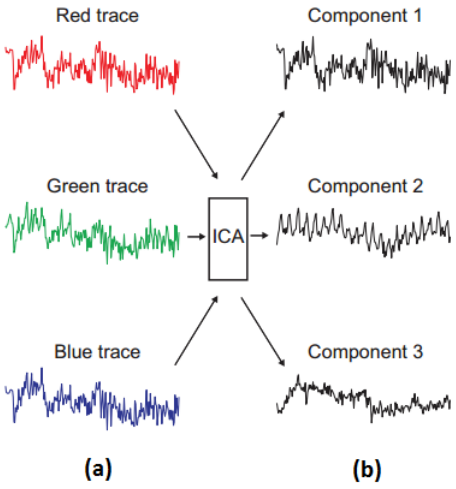



Fig. 1.1: Showing the effect of the ICA algorithm on noisy data. (a) When the mixing matrix is applied, one of the outputs is quite clearly the PPG signal. (b) This figure was borrowed from Poh *et al.*

The video magnification process is illustrated in figure 1.2. In part (a) of the figure, a time sequence of images is inserted into the magnification algorithm. Each image in part (b) is spatially decomposed into different frequency bands. Depending on the application, this is done either by using a laplacian or gaussian pyramid. In the case of heart rate amplification, the skin surface where the PPG signal lies is generally smooth thus a gaussian pyramid would be most appropriate.

The low-frequency spatial band is then temporally band-pass filtered in the range of a normal heart rate (.7 - 3.5 BPM). The output is then amplified by some magnification coefficient which, for heart rate, is in the range of 100 to 150 due to the weak nature of the signal. The pyramid is then collapsed and summed with the original input images to create a pulsating effect in the video as seen from part (c).

Because of the dual-amplification nature of this process (motion and color) it only performs well under zero-movement conditions. Even so much as eyes blinking can cause strange oscillations to be seen throughout the face in the output video. This makes it non-ideal for practical applications.

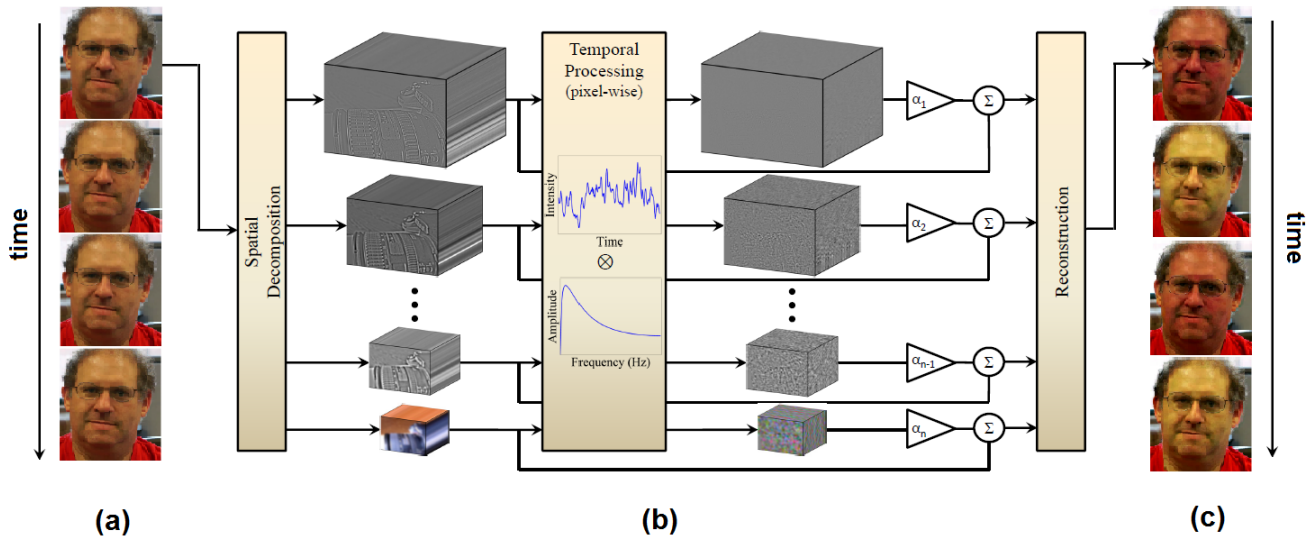



Fig. 1.2: This figure taken from Rubenstein's work shows the process by which video magnification is performed to amplify subtle changes in motion and color.

### 1.1.5 Other Contemporary Methods

After the initial success of Poh *et al.*, several other methods have been established including a *principal component analysis* version (PCA) which reduces the computational complexity [13]. This method, according to the authors, was reported to be just as efficient as the ICA approach. The authors also made use of analyzing the zero-crossing of signals to determine the period. They also examined the possibility of using the auto-correlation to determine a periodic correlation in the input signal. Both ideas proved to have merit but were not seriously pursued.

## 1.2 Research Objectives

There has been a long-standing need for a non-invasive way to measure these vitals; particularly heart rate. These most recent studies and algorithms go to show that the technology has reached the computational speed necessary to sustain usable, inexpensive applications. In spite of the great progress made by many pioneering researchers, there are some critical disadvantages that make this technology "experimental." In order to bridge the gap between clinical trial and real-world use, these have to be addressed.

### 1.2.1 Lower Latency

One problem with the algorithm proposed by Poh *et al.* and all other frequency-analysis techniques is the amount of data necessary to make a reasonable heart rate estimate. The blind source separation (BSS) approach (this includes ICA and PCA) requires a DFT of sufficient length in order to attain the proper resolution. Because of the low sample rate of standard video (around 30 fps) longer data sets translate to very long time intervals between each heart rate calculation. In theory, the two BSS techniques require around 15 seconds of data before the algorithm outputs any metric. In some cases, this frame rate could be even lower, costing an inordinate amount of time to acquire PPG data.

In this research, some standard and very simple analysis is performed to assess whether other non-Fourier techniques might succeed in shortening the time it takes to compute heart rate. These techniques, if successful, might be incorporated into larger more sophisticated methods for estimating and tracking the heart rate. There are, however, advantages to using frequency analysis and the latency impact in most real-world applications may be of minor importance.

### 1.2.2 Motion-Compensation

As suggested by Verkruysse *et al.* and adopted in all of the aforementioned algorithms, the heart rate signal is stronger when considering the average value across an area of skin as opposed to a single point. This area is known in literature as the *Region of Interest* (ROI). While computer vision has made dramatic progress in terms of object recognition (including face recognition), the coherency of the ROI from frame to frame requires a whole new level of accuracy. Subjects whose faces may rotate or be wholly uncooperative with the camera pose a challenge to collecting the desired data. The idea of reducing the error which is caused by movement will be referred to as motion-compensation. While the scope of this research does not include developing new object recognition algorithms, it will explore the state-of-the-art techniques and how effective they are when used in parallel with the heart rate estimation schemes.

### 1.2.3 Model-Based Design

While photoplethysmography is a very old topic, the camera-based acquisition of it is quite recent. Thus far the different modes for increasing the observed PPG SNR have all been analysis-based and have very little to do with the system model itself. Such models are characteristic of other more mature fields such as speech and pattern recognition. In this study a comprehensive yet simple model will be derived and treated as a minimization problem.

## 1.3 Experiments Overview

The remainder of this research will be divided into three experiments. The first will involve analyzing the PPG signal itself and its characteristics as observed from a camera. The commonly-used observation model will be introduced. Several basic analysis techniques, such as peak-detection and Fourier transforms, will be used as estimators to determine their utility for this problem. This will all operate on ideal, relatively noise-less, data sets.

The second experiment will be testing these same techniques on practical data which has a large amount of noise. Two modern approaches for motion-compensation will be explored and utilized to help counteract the effects of the subject's movement on the measure heart rate. There will be some metrics introduced to help quantify the performance of one technique versus another.

Lastly, the final experiment will be a complete reformulation of the heart rate estimation problem at large. A model-based approach will be introduced which significantly differs from every other proposed estimator to date. This experiment, if successful, will not exhaustively determine the merits of this method but will give direction to next steps for creating a more robust model-based algorithm than is currently available.

# Chapter 2

# Remote Heart Rate Analysis in Ideal Circumstances

The purpose of this first experiment is to identify useful and effective ways for extracting the heart rate signal using a common camera. The data, therefore, will be relatively noiseless to promote a high SNR. There will also need to be some ground-truth data for the signal trying to be estimated.

For the experiment setup, seven subjects were recorded at 30 FPS using a high-resolution camcorder. A pulse-oximeter was selected for providing the true data of the subject's heart rate. This was selected for both its relative accuracy and ease of use. The particular device used was the Contec C110 which outputs both the measured waveform and the calculated average heart rate. Both of these signals were used in the analysis. Wearing the pulse-oximeter on their finger, the participants were asked to hold still for several minutes while being recorded with an HD camcorder. In each frame an ROI was selected on the participant's face over which the pixels were averaged together into a single RGB value. The 2-D average is shown by

$$y(t) = \frac{1}{(c_r - c_l)(r_t - r_b)} \sum_{i=c_l}^{c_r} \sum_{j=r_b}^{r_t} I(i,j). \tag{2.1}$$

## 2.1 Heart Rate Signal Analysis

It is important to consider the mathematical model of the desired signal. Consider the equation

$$y(t) = DC + x_{hr}(t) + x_{rr}(t) + n_t, \tag{2.2}$$

where $y(t)$ represents the averaged RGB value measured from frame $t$ in a video sequence. The signals $x_{hr}$ and $x_{rr}$ represent the heart rate and respiration rate, respectively. If only

observing the small-signal AC components of this equation, it becomes clear, as noted by Nilsson *et al.* and Johansson and Oberg, that the observed data is a combination of respiration signal, heart rate signal, and noise [14, 15]

$$y_{AC}(t) = x_{hr}(t) + x_{rr}(t) + n_t. \tag{2.3}$$

Furthermore these signals can be approximated to be quasi-stationary sinusoids over an appropriate interval [4]. This can be modeled as

$$x_{hr}(t) = C_{hr} \sin(\frac{2\pi f_{hr}}{60} t + \phi_{hr}), \tag{2.4}$$

and

$$x_{rr}(t) = C_{rr} \sin(\frac{2\pi f_{rr}}{60} t + \phi_{rr}). \tag{2.5}$$

.

It is important to note that heart rate and respiration rate signals rarely share close frequency bands under normal activity. This allows for traditional filtering to help separate the two signals. Perhaps the most problematic component of the received signal is the noise. It is hypothesized that the noise can be further decomposed into three sub-signals

$$n_t = n_w + n_{motion} + n_{extern}, \tag{2.6}$$

where $n_w$ is the noise introduced by the camera electronics and quantization error. The $n_{motion}$ term signifies the noise due to voluntary and involuntary motion from the subject. All of the noise generated by environmental changes and un-mapped biological signals are represented by $n_{extern}$.

### 2.1.1 SNR and Signal Acquisition

In order to better acquire and detect the heart rate signal it is important to identify the contributing factors to the signal's SNR. Knowing which parameters are critical to

amplifying signal power will help determine hardware specifications which will maximize the SNR and thus give the observer more confidence in the signal he is trying to measure. Before proceeding, it is important to note that the green channel of RGB has historically given the strongest SNR of the three channels [7]. From this point forward, speaking about a scalar-valued PPG signal will almost exclusively be referring to the green channel. Figure 2.1 sheds light on the signal chain which will be used in the data acquisition.

It becomes clear that the signal of interest $x_{hr}(\alpha, \beta)$, which is continuous, is sampled and, more importantly, quantized. This quantized signal is then summed over $W$ pixels to produce the spatially-averaged heart rate signal, $x_{hr}[ROI]$. A topic of interest is to understand how the quantization of the camera's A/D converter affects the measured heart rate signal ($x_{hr}[i, j]$) SNR. Rice introduces a formula for calculating this quantity as it relates to quantization noise [16],

$$SNR_q = 4.8 + 6.02b - 20\log_{10}\frac{X_m}{X_p} - 20\log_{10}\frac{X_p}{X_{rms}}. \tag{2.7}$$

In this equation $b$ represents the number of bits of the quantizer which, in the case of normal commercial cameras, is 8 bits. $X_m$ is the largest value achievable from the quantizer which is $2^8 = 256$. $X_p$ is the largest peak of the signal (assumed to be zero-mean). Lastly, $X_{rms}$ is the RMS value for the signal which can be computed by simply finding the standard deviation.

An issue arises when attempting to calculate $x_{hr}[i, j]$ directly. Consistent with other literature and as shown in figure 2.2 the amplitude of the heart rate signal is smaller than one quantization level of the 8-bit CMOS ADC found in most cameras. This means that the signal $x_{hr}[i, j]$ is not directly measurable without finer resolution.

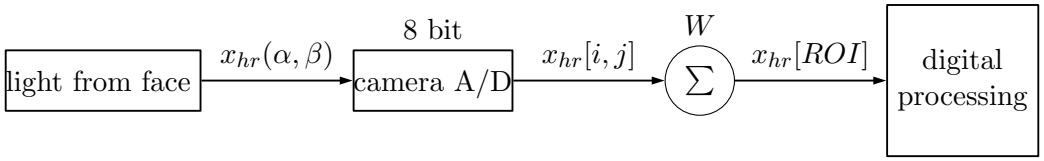



Fig. 2.1: Data acquisition flow diagram.

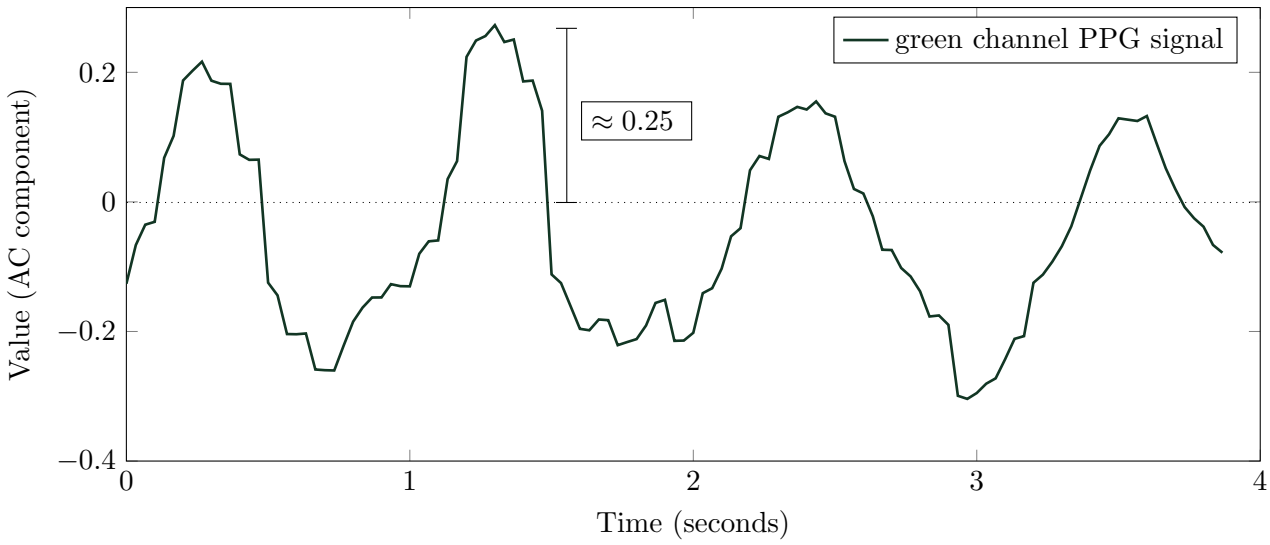



Fig. 2.2: The RAW green channel (zero-mean) shows amplitude of the PPG as measured from the output of the accumulator $x_{hr}[ROI]$.

The problem of insufficient resolution gives rise to the common method of averaging over a sufficient number of pixels in order to boost the SNR to an observable level. Considering the mathematics of the problem, let the following expression represent the observed quantized signal

$$x_n = s_n + v_n \qquad n = 1, 2, ..., W$$
$$v_n \sim U(0, \sigma_v^2),$$

where $W$ is the number of pixels in a region, $s_n$ is the heart rate signal and $v_n$ is the quantization noise. The noise is reasonably assumed to be spatially independent. Now, if the assumption can be made that the signal is coherent across all pixels (*i.e.* the region is contiguous on a area of skin) then it can be said

$$Wx_n = W(s_n + v_n)$$
$$\text{var}(Wx_n) = \text{var}(W(s_n + v_n))$$
$$W^2 P_{x_n} = W^2 P_{s_n} + W P_{v_n}$$

$$SNR_W = \frac{WP_{s_n}}{P_{v_n}} = W(SNR_q), \tag{2.8}$$

where $SNR_W$ is signal power measured by an averaging window of size $W$. This result suggests that the output PPG signal strength is proportional to the power of the PPG signal as seen by the quantizer. Converting the above formula into decibels produces:

$$SNR_q = SNR_W - 10\log_{10} W. \tag{2.9}$$

In order to accurately ascertain the SNR in any given video, one must recognize that the true signal power can never be measured as it is always in the presence of noise. A very common method to approximate the SNR of a channel is to use the following expression:

$$\hat{SNR} = \frac{\text{var}(Y) - \text{var}(N)}{\text{var}(N)}, \tag{2.10}$$

$$\text{where } Y = S + N.$$

The measured signal $Y$ can be thought to include both signal and noise power, hence the noise power is removed. The noise power itself is attained by the sampled covariance of areas of the image which contain pure noise (no signal). Using multiple window sizes, the results in figure 2.3 were attained.

It is clear that there is a linear correlation between $SNR_w$ and $W$ which coincides with the mathematical model. By fitting a line to the data and examining (2.8) one can determine that the slope of the line is $SNR_q$. This gives the result:

$$SNR_q = 0.0004.$$

Converting to decibels, we find

$$SNR_q = -34dB.$$

The power of the PPG signal is astonishingly small. It is so small that spatial pooling of pixels is absolutely necessary in order to raise the signal strength above that of the quantization noise, excluding all other noise. This gives a bit of perspective on the enormity

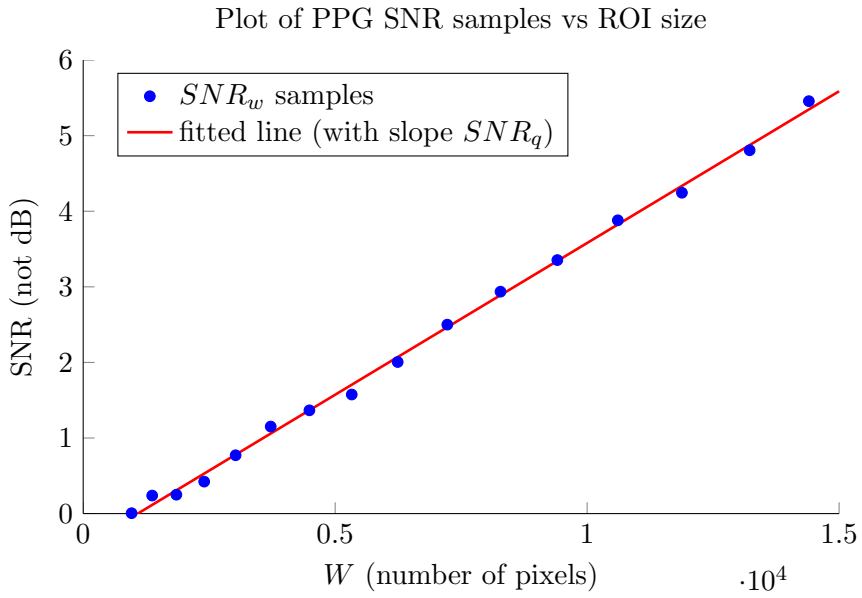



Fig. 2.3: The samples relating a measured PPG SNR with a given window size show the linear correlation between the two.

of the challenge at hand. This also provides a threshold for the minimum number of pixels to use in the window.

### 2.1.2 Effects of Video Compression

It is always important to consider sources of signal degradation when selecting equipment for experimentation/application. Video contains such a massive amount of data that recording and processing it can be an arduous task for even the most powerful computing architectures. It is therefore necessary to compress the video and throw away data which will be missed the least. There is always the possibility, however, that the compression meant to preserve the visual aesthetics of the video may be throwing away the human-invisible data which is used to find the PPG signal.

The most commonly used standard to date for video compression is H.264. This standard is well known for its high video quality achieved at very low bit-rates. For this reason it is ubiquitous in video recorders, video players, and video streamers. There are varying degrees of compression found within the H.264 family but the key principle is the same: to

remove inter-frame redundant data. This means, for example, that smooth surfaces whose appearance changes very little from one frame to the next will experience the greatest compression.

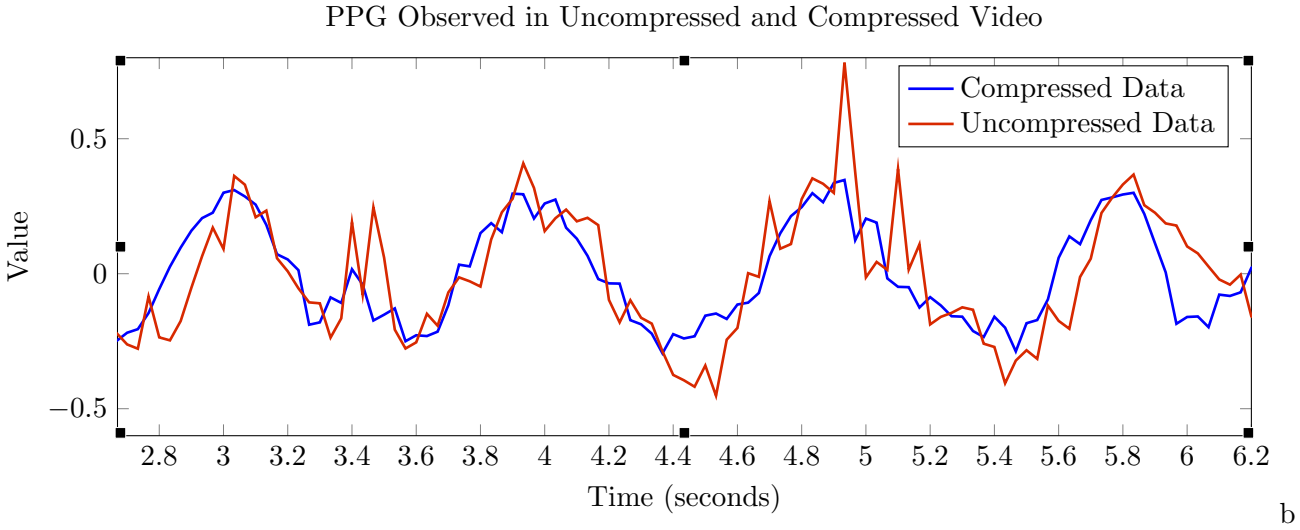
An exploratory test was run using a Cannon HD Camcorder (1920x1080 resolution) and a Microsoft Lifecam Studio HD webcam. The Cannon used the H.264 standard for video storage and the webcam was recording using Matlab's Image Acquisition Toolbox at full resolution in uncompressed AVI format. These two cameras were synchronously recording the same subject under optimal conditions (natural lighting, holding still). Both videos were processed using the same window size and visually inspected for the PPG signal. Figure 2.4 contains a portion of the results.

The heart rate signal quality, contrary to intuition, looks more favorable on the compressed camcorder. This result, though surprising, is consistent with the finding of Rustand that little to no signal data was lost due to compression [17]. From other samples it seems that the PPG may be stronger in the uncompressed video, but the noise also has more power. This would seem to suggest that while inter-frame compression does effect the temporal data, it has almost no effect on the signal of interest. This is a valuable result in that this allows experiments to be carried out with inexpensive devices and with significantly less storage. The participants in these experiments were all recorded using the same camcorder as was demonstrated in this test.

### 2.1.3 Underlying Signal Model

From the computed average over an area of pixels it is possible to hypothesize the amplitude of the heart rate signal as being less than one quantization level of an 8-bit CMOS camera. If it were assumed that a subject could hold perfectly still and no other environmental or physiological noise were present, it might be possible to model the quantization of the heart rate as follows.

From figure 2.5, it is obvious that there are other signals present in the AC signal received at the camera sensor, or else it would be impossible to amplify the signal through pixel averaging. However, it is important to establish from this concocted scenario that the




b

Fig. 2.4: Comparing the PPG signal measured from two videos using different storage formats.

presence of other time-varying signals (although considered noise) are critical in allowing the heart rate signal to be detected on a per-pixel basis.

A far more interesting model is found when considering the different noise structures and their roles in exposing the PPG signal. Now (2.3) will be used to initiate a new underlying analog signal. Combining motion noise, respiration rate, and heart rate create a more realistic model and lend credibility to the earlier hypothesis of the PPG signal's nature. Figure 2.6 demonstrates this simulated quantized input.

The waveform of the heart rate, while not clearly defined, is nonetheless present in the discrete output of the camera sensor. This would justify the simple yet intelligent practice of averaging over a sufficient number of pixels in order to achieve greater apparent resolution of the signal. Figure 2.7 comparison was performed between the simulated quantized output and several actual randomly-selected pixel outputs corresponding to the same time period.

The difference between the simulated pixel output and the actual pixel output is quite obvious. The true PPG signal is much stronger than expected; emitting obvious peaks of two to three quantization levels in the actual received data. This raises the question of whether the previous assumption of the signal's coherency was correct.

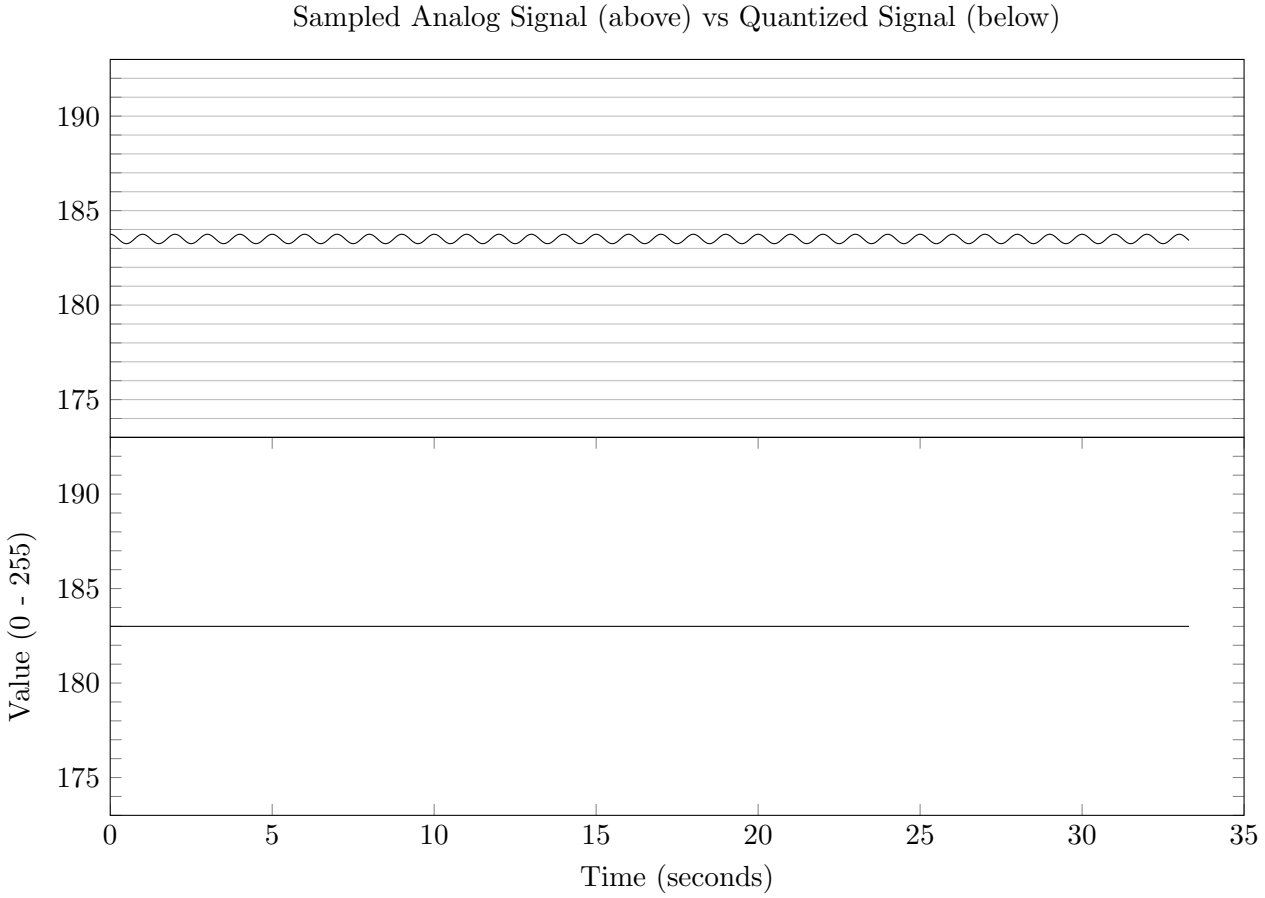



Fig. 2.5: The simulated analog input of the heart rate lies right between two quantization levels of the image sensor. Thus, without any perturbations the signal is invisible to the camera. Rounding is assumed to be truncation.

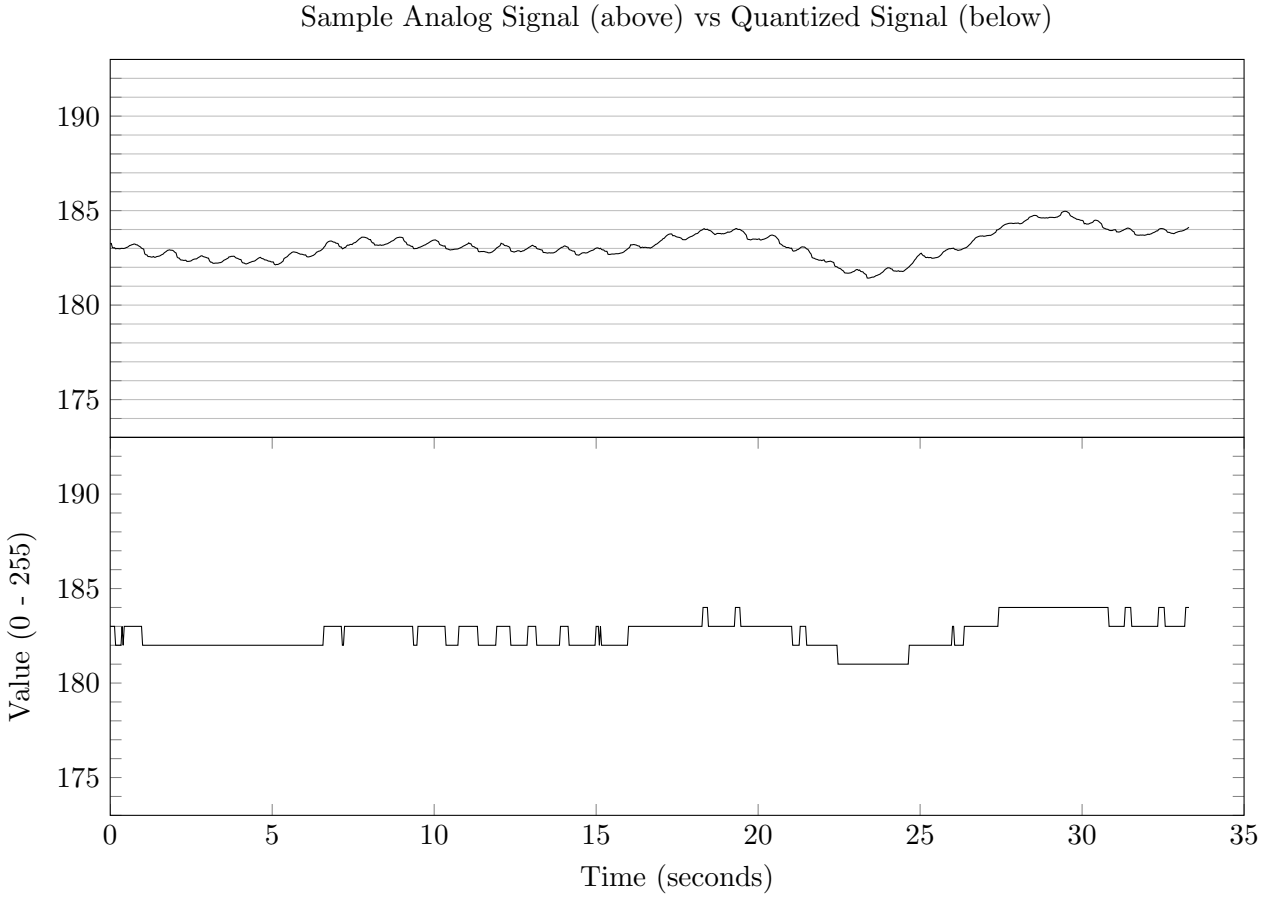



Fig. 2.6: This represents a more realistic case for the hypothetical PPG signal source with noise present. The low frequency noise allows the heart rate signal to toggle the bits of the quantizer.

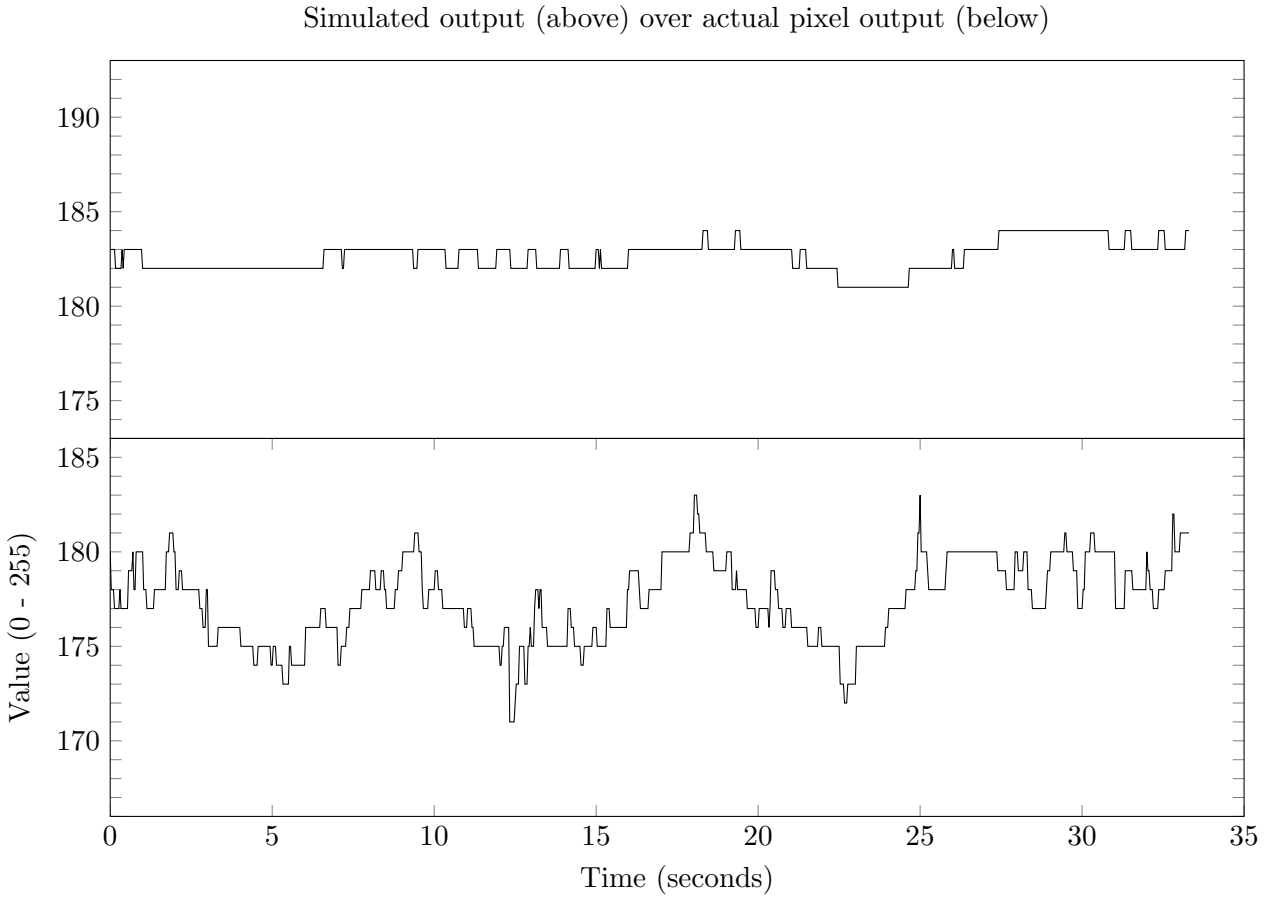



Fig. 2.7: The simulated pixel signal is compared to an actual pixel output corresponding to a point on the subject's forehead.

### 2.1.4 Future Work Regarding PPG Signal Propagation

Averaging has many statistical advantages when seeking to reduce zero-mean noise present in a signal across time or space. As shown in the derivation of (2.9), averaging is very effective if the signal is indeed coherent. However, the recent comparison of the hypothetical PPG signal as measured by an 8-bit quantizer and the actual measured signal suggests that the heart rate may not be uniformly in-phase across all regions of the face.

It is not hard to imagine that some regions of the facial tissue may emit the heart rate better than others [17]. It is also conceivable that there may be phase distortion due to blood arriving at facial capillaries at different times. If indeed there exists some propagation path for which the blood moves through the face, this would mean that a particular spatial linear combination is required to attain the maximum SNR.

It may be suggested that there is much more research to be done regarding the nature of the PPG signal as observed from the facial skin. There has yet to be a study performed on the effects of distance, image resolution, bit depth, level of illumination, and artificial lighting have on the signal. A greater understanding of these effects may be necessary to advance the art of the field. Also, the previous observations lead one to believe that the signal may not be coherent across the face. In order to maximize the observed heart rate signal's strength, the principles and techniques of array processing may introduce useful methods for biometric sensing.

## 2.2 Reporting Peak-Detection Estimates

To begin the exploration of techniques of heart rate estimation it seemed most fitting to start with the most intuitive and naive approach: peak detection. This is, after all, the method that medical staff and phlebotomists alike use when assessing an individual's pulse. The data from the green channel was filtered using a 4th order band-pass zero-phase Butterworth filter in order to remove the mean and respiration rate signals.

As seen in figure 2.8 a simple peak detector was then applied in order to calculate the time between two peaks ( $T_{hr}$) which could then be interpreted as the instantaneous period of the heart rate. The heart rate would naturally be calculated as $f_{hr} = \frac{1}{T_{hr}}$. There is

one major disadvantage to his method of estimation. Because of the inverse relationship between period and frequency, small errors in estimating shorter periods result in large errors in higher frequencies.

The accuracy of peak detection deteriorates for higher heart rate estimates. The error margin increases for shorter periods as shown in figure 2.9. This gives cause for concern when using this kind of processing for determining infant heart rate or in exercise scenarios where the PPG is elevated in frequency. One method to help reduce the error may be to employ a peak approximation formula derived by Jacobsen and Kootsookos [18]. The estimated peak is determined by

$$\delta = -\left[\frac{x_{k+1} - x_{k-1}}{2x_k - x_{k+1} - x_{k-1}}\right] \tag{2.11}$$

$$\hat{x}_{peak} = x_k - \delta.$$

By using the two adjacent values next to a detected peak, one can make a reasonable estimate as to where the true peak lies. Updating the calculated heart rate using only one peak-to-peak period, however, is in and of itself inaccurate because of the inherent variation from heart beat to heart beat. Remembering the common practice of counting pulses over a period of time leads one to believe that perhaps a moving-average (MA) model might be adopted for a local average in time:

$$\hat{y}_{MA} = \sum_{i=0}^{M-1} \beta_i x(k-i). \tag{2.12}$$

The terms in $x$ are the current and previous heart rates calculated from the periods. In this example, allow the coefficients to be conditioned as follows:

$$\beta_i = \tfrac{1}{M} \text{ such that } \textstyle\sum_{i=0}^{M-1} \beta i = 1.$$

There is also an element of auto regression in the problem seeing that the most recent heart rate estimates should influence the current estimate. For demonstration purposes, let a second linear predictor be defined as

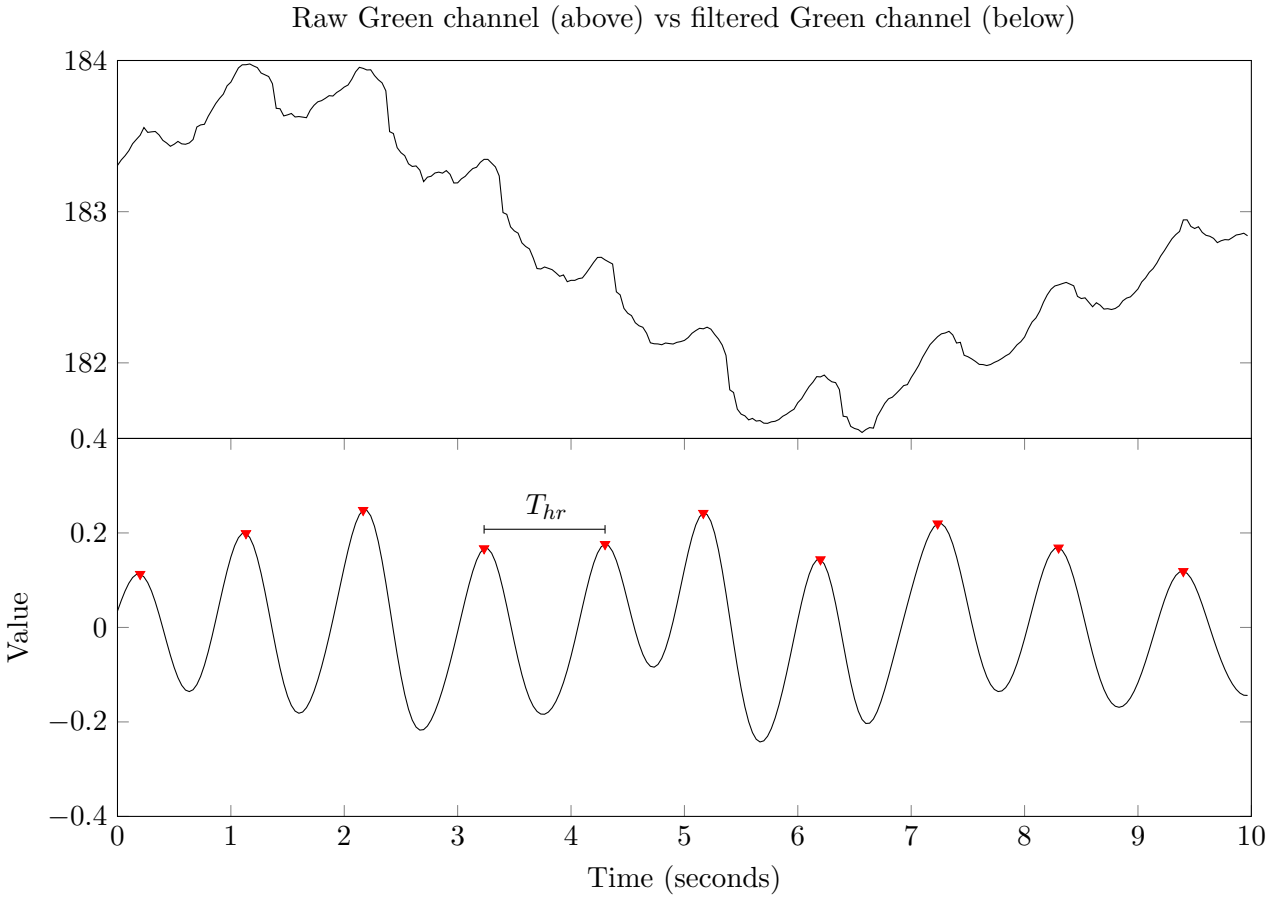



Fig. 2.8: The raw green channel is compared to the zero-phase filtered data. A peak detector is applied to the filtered data to determine $T_{hr}$ for every peak.

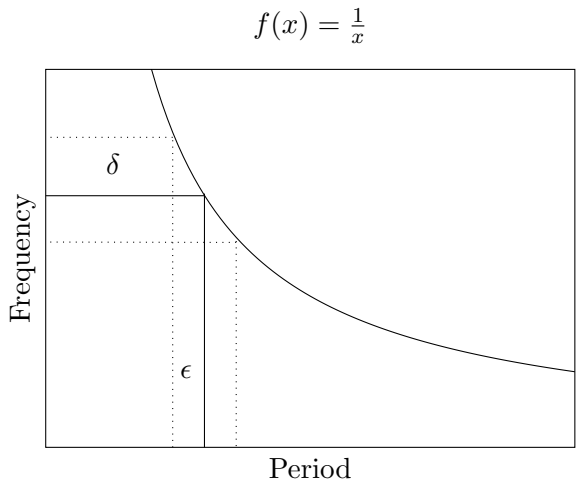



Fig. 2.9: This shows how the $\frac{1}{x}$ relationship creates a large error margin ( $\delta$) created from small errors ($\epsilon$) in the low-valued region of the x-axis.

$$\hat{y}_{AR} \equiv \sum_{i=1}^{M} \alpha_{i-1} y_{AR}(k-i). \tag{2.13}$$

For simplicity's sake, assume that the $\alpha$'s are equivalent to the $\beta$'s. These two non-optimized predictors were used across the subject's first stage recordings with results that are very similar to figure 2.10. Both predictors, in most cases, performed well enough for practical heart rate estimation. As was measured by the error power and can be visually inspected, the AR predictor usually faired better than that of the MA model. If the MA and AR models are thought of as FIR and IIR filters, this should come as no surprise. Both of these, however, still suffer from high frequency ringing which might make them mildly irritating to use for medical purposes.

As demonstrated by this experiment, peak detection may be a viable option for heart rate estimation under ideal circumstances. The advantages to peak detection is much lower latency depending on the coefficient length used by the predictor. However, the price for lower latency is increased error margin for elevated heart rates. It may also be difficult to robustly extract the actual signal peaks in a noisy environment. A further study on the PPG waveform as observed from the skin may allow the ability to provide an matched filter

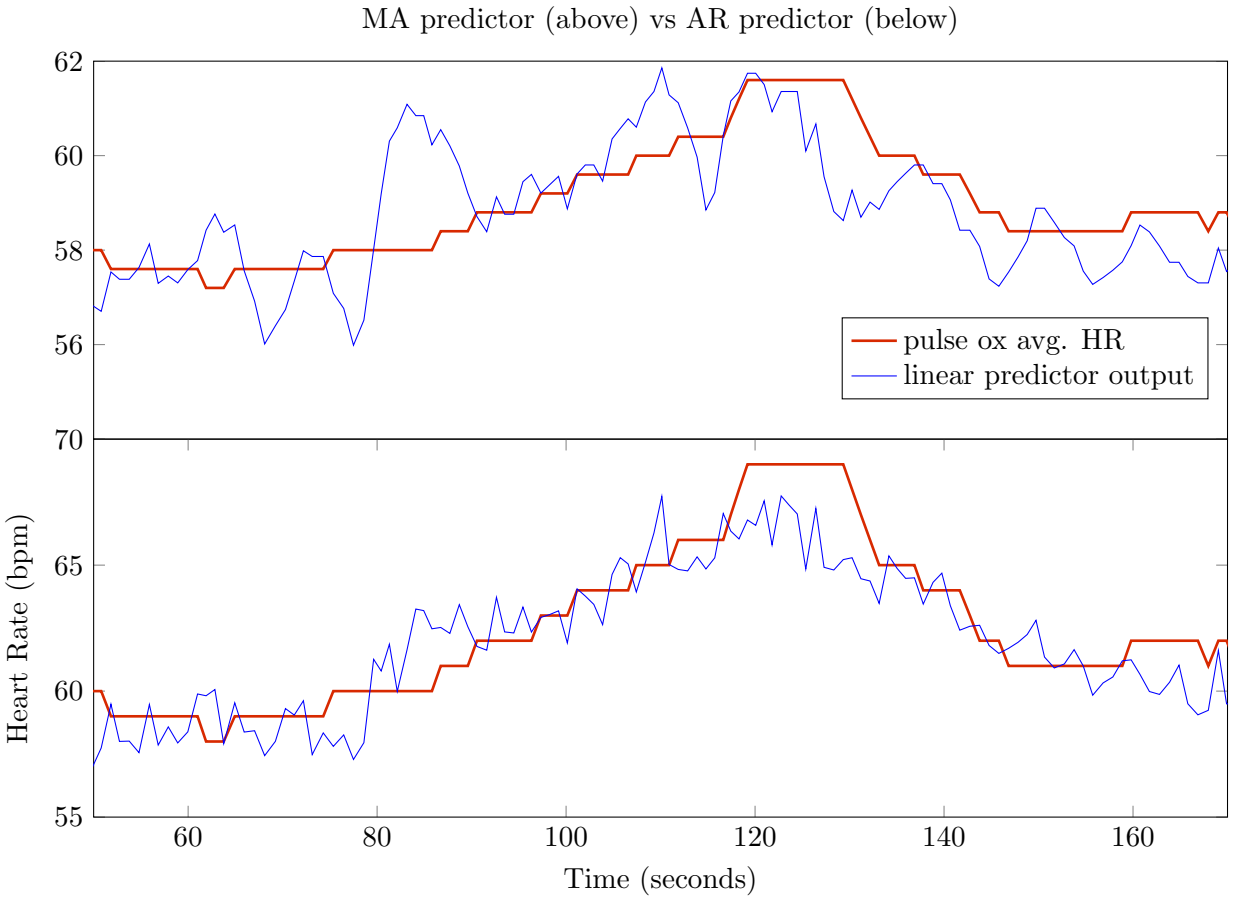



Fig. 2.10: Comparing the results of $y_{MA}$ and $y_{AR}$ to the measured heart rate as per the pulse oximeter.

to aide peak detection.

Beyond the scope of this thesis, it could be proposed to investigate further the use of filters with peak detectors. The difficult part of this signal is that it does not have a constant period between samples. Thus there is not only error in the signal but also in the time domain. In order to minimize the error power from the predictor, one might use a partial Total Least Squares approach to account for signal as well as sampling error.

## 2.3 Using the FFT

In any frequency analysis problem the FFT (or DFT) will be one of the first tools used. In most problems it does a very good job. However, in the case of heart rate estimation there are two opposing factors: sample rate and frequency resolution. While there are some cameras capable of capturing hundreds or even thousands of frames per second, most consumer-grade cameras, which are the focus of this thesis, can only capture at a lowly 30 FPS. This sample rate is more than adequate for the detection of heart rate seeing that it is well above Nyquist. However, searching for a signal with a narrow bandwidth in a slowly-sampled channel can be problematic. Consider the following equations:

$$\begin{aligned} \tfrac{F_s}{N} &= \tfrac{x_{hr}}{60k} \\ \frac{F_s}{N} 60k &= x_{hr}. \end{aligned} \tag{2.14}$$

In these two equations $F_s$ is the sample rate, $N$ is the total number of samples (used in the FFT), $k$ is the frequency bin, and $x_{hr}$ is the calculate heart rate (as measured in BPM). By fixing $k$ to 1, (2.14) gives the frequency resolution between any two bins in the FFT output. Using this same equation, if the desired accuracy of the heart rate estimator were 3 BPM,

$$\begin{aligned} \tfrac{F_s}{N} 60 &= 3 \\ \tfrac{30}{3} * 60 = N &= 600 \text{ samples}, \end{aligned}$$

the number of samples required for the desired accuracy would take 20 seconds to acquire at 30 FPS. This is problematic for two reasons. First, in some applications, like exercise, it

is important to have timely and frequent updates of your current heart rate. The latency to perform an FFT becomes undesirable in this case. Also, the heart rate cannot be assumed to be stationary over such a long interval.

One answer to the first problem would be to use a sliding window. As the data comes in, instead of completely starting a new data set, replace the oldest samples with the most recent data. This will allow for very rapid updates of the frequency spectrum. It is also entirely justified to use an update scheme which is composed mostly of the same data used in the previous estimate due to the averaging nature of the heart rate signal. There is enough variation between pulses as seen with the peak detector that averaging over the current and previous values produces the most useful results.

The PPG signal, before it can be inspected for frequency content must first be filtered. A bandpass filter is applied over the range where a normal heart would be expected (0.6 - 3.5 Hz). This is necessary because the massive DC and respiration rate components wash out the PPG signal with spectral leakage and spectral spreading otherwise. The data is also zero-padded up to 1024 in order to allow for a more accurate picture of the frequency content. The DFT output of the filtered green channel from the ROI, as shown in figure 2.11, readily exploits the dominant heart rate signal.

In order to test its utility, the output of a sliding-window FFT estimator was compared to the average heart rate reported by a pulse oximeter. The PPG as measured per the finger by the pulse oximeter was also reported.

Figure 2.12 demonstrates that under ideal conditions simply taking the FFT of the green channel has a comparable output to the pulse oximeter. This method only differs from the BSS technique in that it does not require a mixing matrix in order to statistically separate the red, green, and blue channels [9]. From inspection, the estimator, in most cases, never differed more than 5 BPM from the true heart rate.

While the sliding windows gives frequent updates, it is still dealing with a quasi non-stationary signal. While the heart rate, in general, varies only a little over an extended period, there are times where it transitions very quickly. Such a case might be an adrenaline

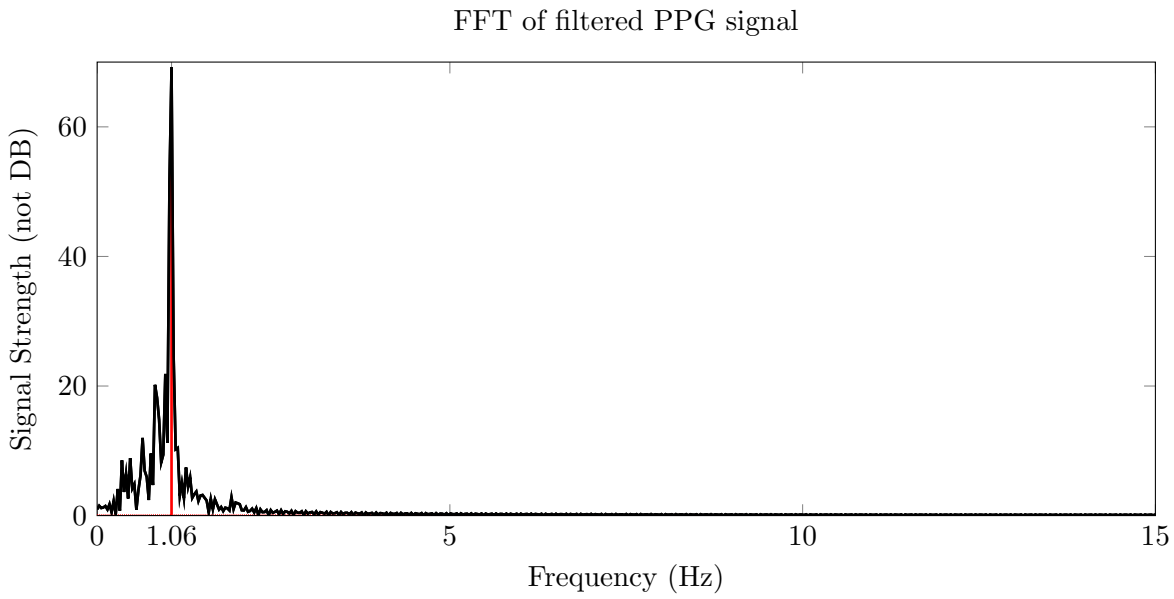



Fig. 2.11: Shows the dominant frequency of the ROI is 1.06 Hz which corresponds to 64 BPM heart rate.

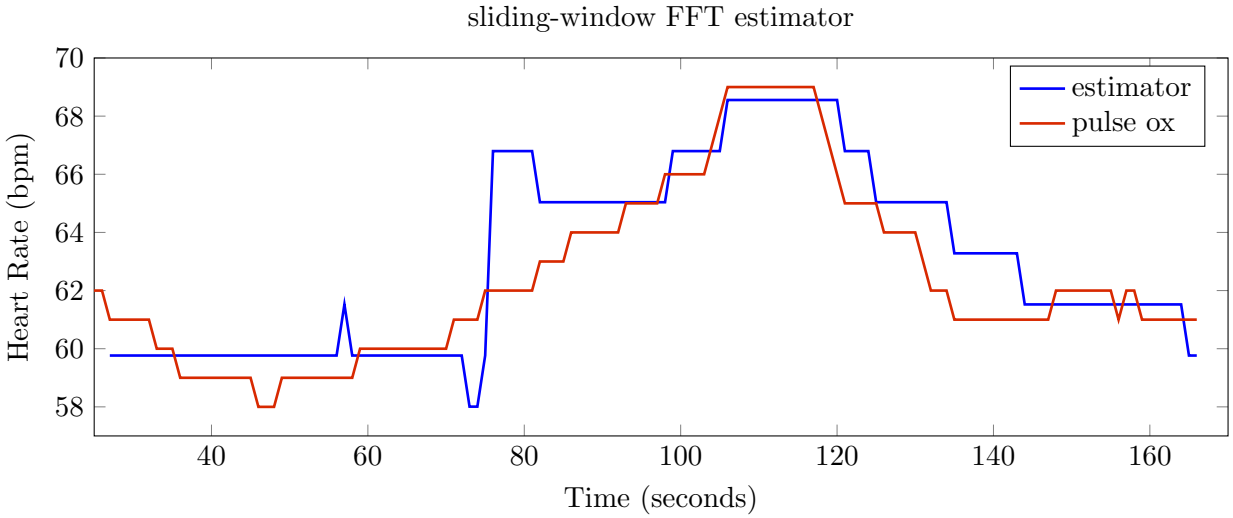



Fig. 2.12: A comparison between the sliding-window FFT estimator and the average heart rate reported by the pulse oximeter.

rush when someone is asked a sensitive question. The sliding window, in many ways, acts as a low-pass filter and would thus miss any "transient" behavior in heart rate.

One approach would be to shrink the FFT window size. This would allow for sudden or rapid shifts in frequency to surface sooner in the spectrum output. This, however, conflicts with the original problems of window size and frequency resolution. The common practice of zero-padding still applies to a smaller window, however this does not technically gain any frequency resolution. However, the peak-approximation formula (2.11) used with the peak detector may possibly provide the additional accuracy needed.

It can be seen from figure 2.13 that the short window FFT performed surprisingly well in the heart rate tracking test. While not quite on par with the full window tracker, it was exceptionally close considering that it used half of the data as the former and would be expected to have a frequency resolution of 6 BPM. There are a couple of interesting points worth noting from the comparison of these two estimators.

First, there appears to be a trade-off between estimating the resolution error using the peak approximation and introducing noise into the average heart rate. The small window estimator appears to have a high frequency noise component perhaps due to the approximation error. It would be expected that this "noise" would increase if the window size decreased and the estimator relied more heavily on the peak approximations.

Secondly, there is a noticeable phase lead in the short window estimator. This is due to the fact that it relies less on previous data which introduced a lag in the update step of the long window estimator. Intuitively the smaller the window size the closer one comes to finding the instantaneous heart rate.

## 2.4 Power Spectral Density Frequency Analysis

One key feature to exploit when modeling the observed PPG signal (assumed to be stationary) is that it's periodic. This is significant because it means that the signal will also periodically correlate with itself. The cyclo-stationary nature of this signal suggests that examining the auto-correlation function will boost the SNR in the presence of uncorrelated noise.

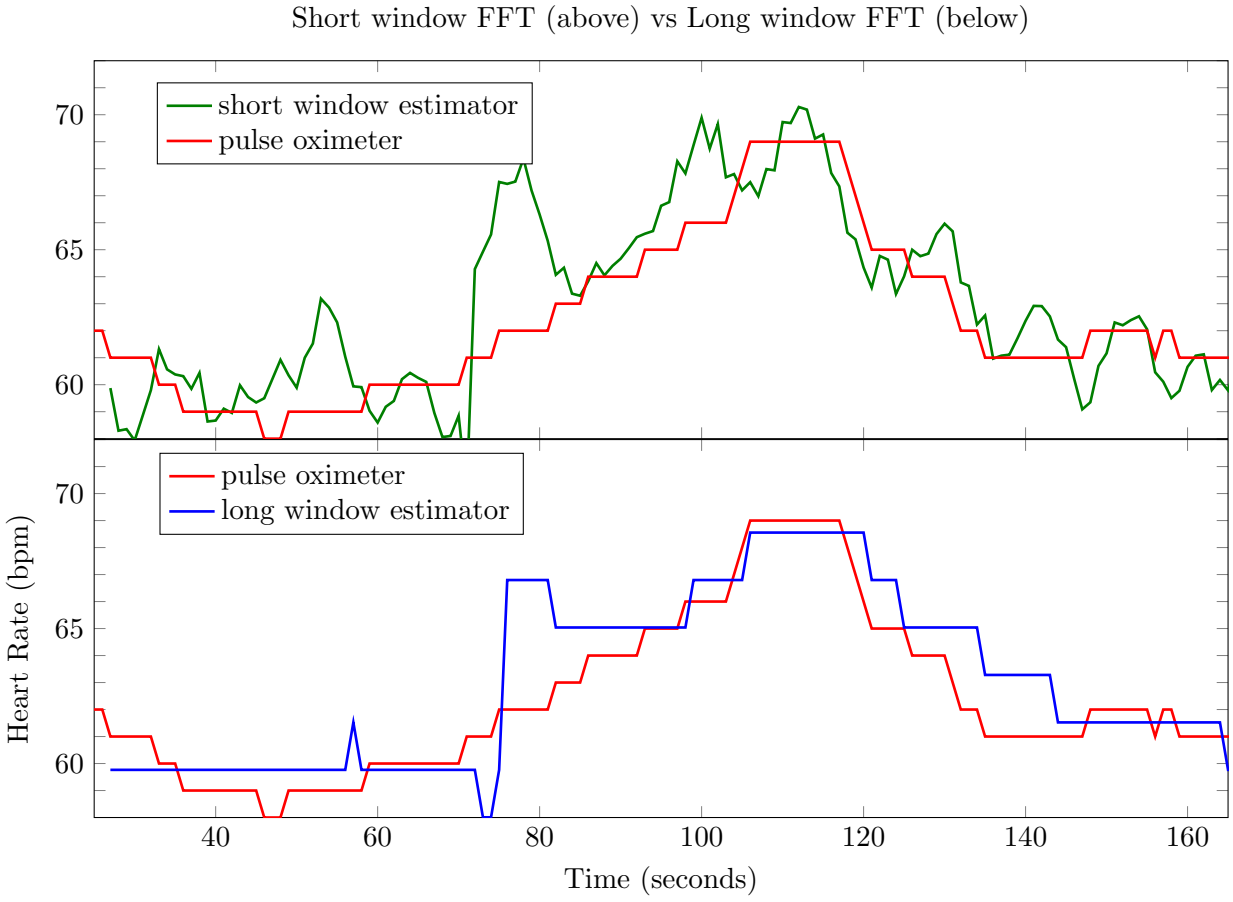



Fig. 2.13: Providing a comparison between two estimators using different window lengths. The top uses 300 data points while the bottom uses 600 data points.

### 2.4.1 Least-Squares De-Trending

In order to perform a relatively accurate calculation of the autocorrelation function, the signal should have the large low-frequency artifacts removed as to clearly see the AC signals of the PPG and noise. This can be done through conventional filtering, however this starts to become a problem when dealing with a slowly-sampled signal. The group delay of the filter can be on the order of several seconds, which adds to the latency of the estimator. One possible avenue to remove the mean and low-frequency signals is to simply apply a least-squares fit. By forming a subspace spanned by the frequencies that are to be removed, one can project the data onto the subspace and then remove the projection from the signal. Sampled low-frequency signals are used to construct the projection matrix

$$A = \begin{bmatrix} a_1 & a_2 & ... & a_n \end{bmatrix}$$

where

$$a_i(k) = e^{\frac{j2\pi f_i k}{f_s}}$$

$$s_{fit} = s - AA^{\dagger}s. \tag{2.15}$$

The results from (2.15) can be seen in figure 2.14. The columns of $A$ contain the sampled sinusoids (both the real and imaginary portions). This de-trending can occur iteratively in "chunks" or blocks of sequential data to give the appearance of filtering. While the initial starting phase of the sinusoidal basis functions is inconsequential (just start at zero), the phase between processing blocks should be continuous in order for the de-trended output to be continuous. As seen in the spectral plot, the low frequencies are drastically reduced. This method of "filtering" is just as effective as a high-pass, low-cutoff filter with substantially less latency.

### 2.4.2 Observed Periodicity in the Autocorrelation Function

The results from the previous signal de-trending prove very useful for computing the auto-correlation of a data set. Without the low-frequency removal, the AC ripple of the PPG signal would be drowned out by the larger also-correlated signals due to respiration

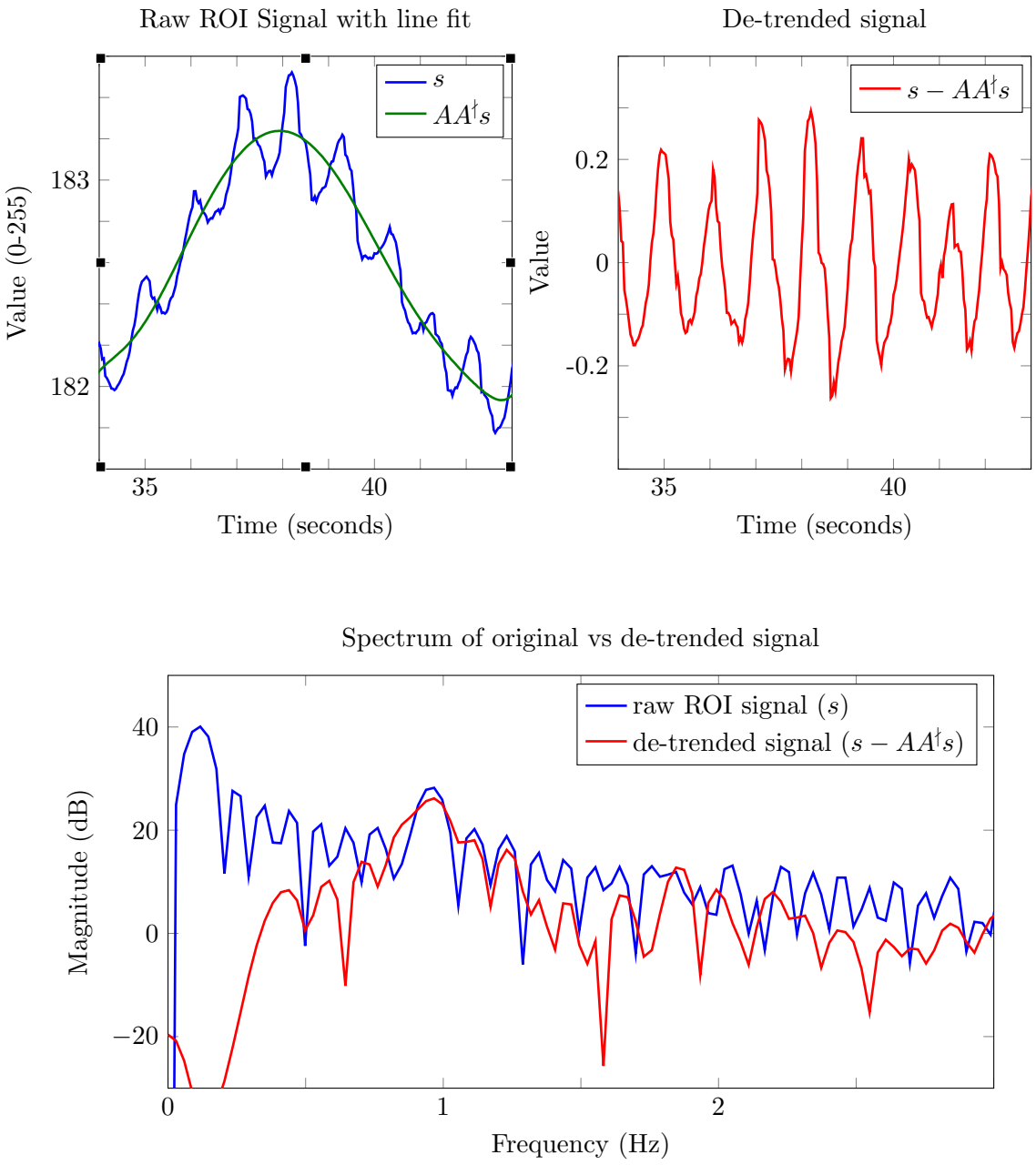



Fig. 2.14: Demonstrating the use of least-squared de-trending to obtain the spectral results (below) which are comparable to those achieved by filtering.

and motion. The magnitude spectrum of the auto-correlation function, or power spectral density (PSD) function, can be observed in figure 2.15. As seen from comparing the spectral output of the de-trended data, the auto-correlation perceptibly removes uncorrelated noise that is present with the signal of interest. It can also be clearly seen from the time-domain auto-correlation that a periodic signal is present.

The autocorrelation is simply computed using two windows, one long and the other short. The length of the shorter window will determine the number of points used to estimate the auto-correlation whereas the difference in size between the longer and shorter windows determines the length of the function. The longer window sizes lead to better accuracy but also lead to more samples being needed. There is a delicate balance in the size and accuracy of the autocorrelation function. In this case, the longer window size is 400 samples which is approximately 13 seconds if data. The shorter window was 75 sample which is 2.5 seconds of data. This would mean that after an initial "warm up" period used to fill the longer window, the auto-correlation and consequently PSD could be computed every 2.5 seconds. This is a greatly improved latency factor.

### 2.4.3 PSD Heart Rate Estimator

The next and final step for using the auto-correlation function in heart rate estimation is to track the frequency changes over time as done in the FFT approach. This will be referred to as the PSD heart rate estimator. Figures 2.16 and 2.17 show the results of using this processing on the three different channels. This example vindicates the choice to exclusively use the green channel when analyzing the PPG signal.

While the other channels (red and blue) carry a hint of the PPG signal, the heart rate is very dominant and clear in the output of the green channel as seen in figure 2.16. A visual comparison between the green channel spectrum over time and the average heart rate reported by the pulse oximeter are compared in figure 2.17.

As seen above, the frequency of the periodic auto-correlation function does indeed coincide with the PPG signal. There is, however, a noticeable phase lead in the estimator. This may be due to the fact that the internal calculations of the pulse oximeter rely on real-

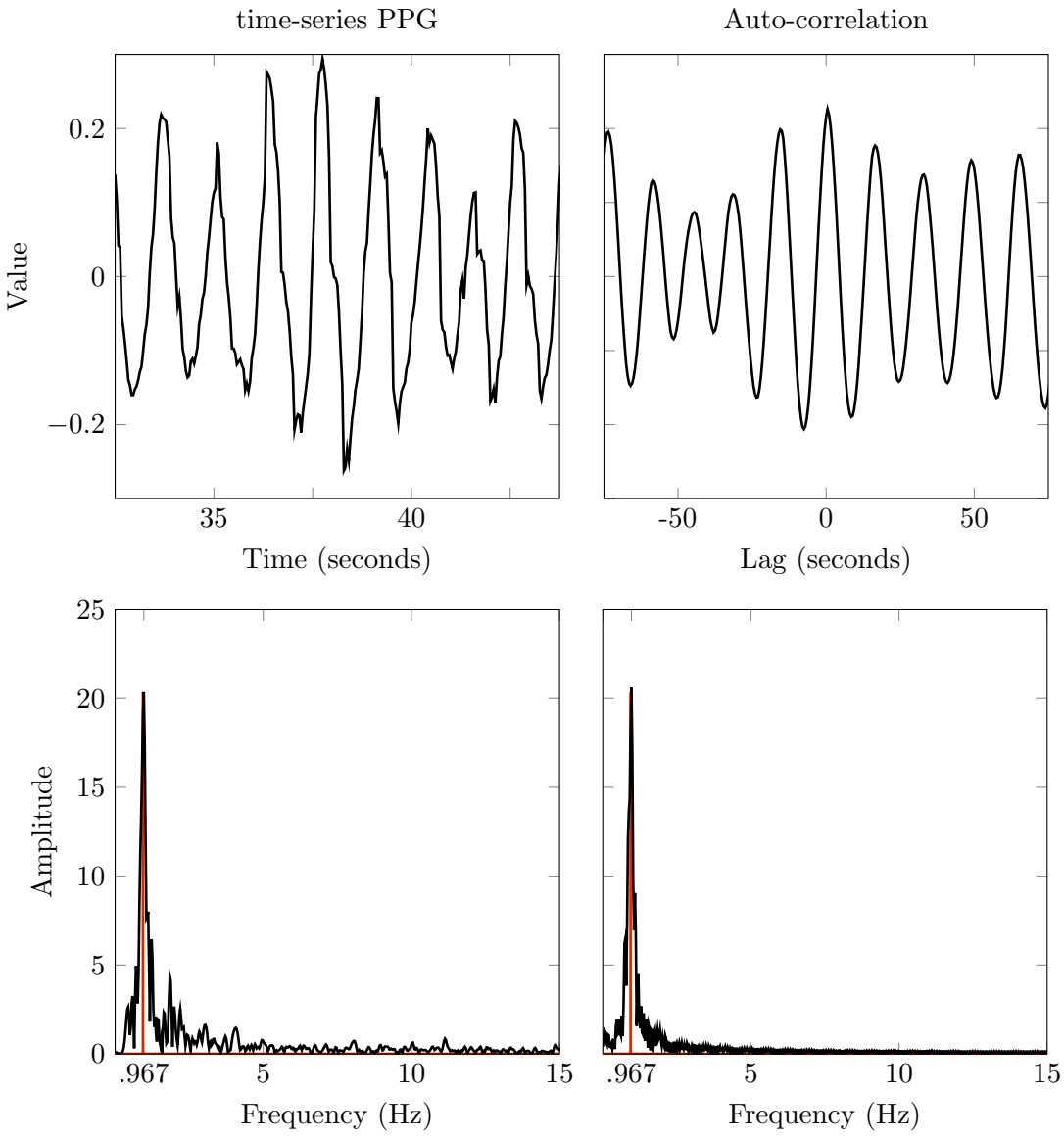



Fig. 2.15: Comparing the fitted measured PPG signal (left) to the estimated auto-correlation function of the signal (right). Both have a respective time plot (above) and spectrum plot (below).

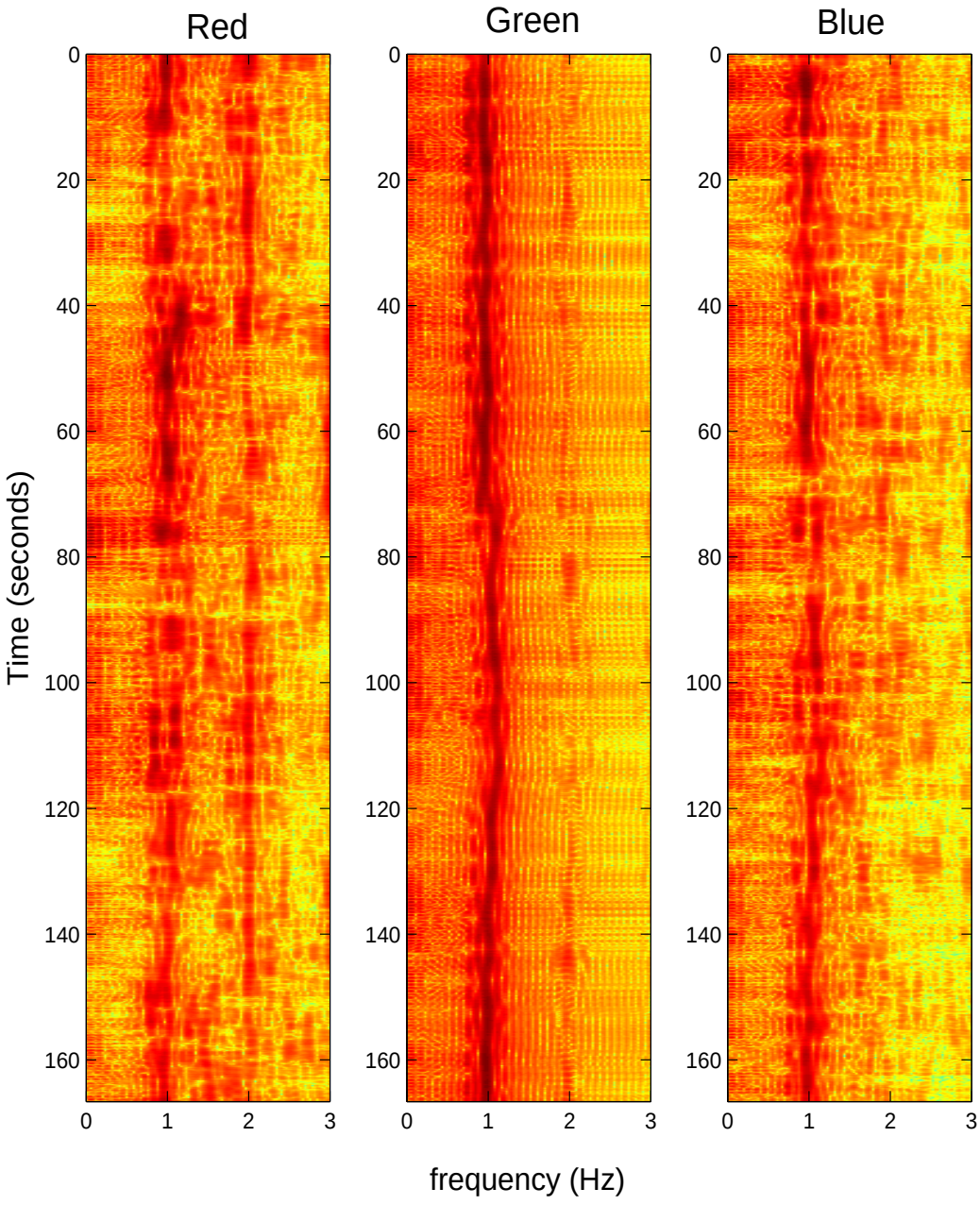



Fig. 2.16: Comparing the frequency magnitudes of the three RGB channels over time. Dark red indicate higher elevations. The magnitudes are plotted in dB.

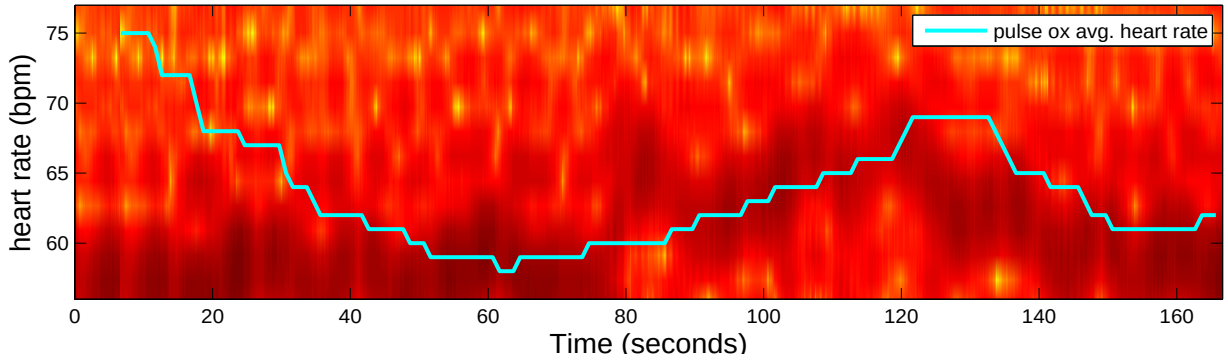



Fig. 2.17: The average heart rate is plotted over the frequency magnitude graph of the auto-correlation function.

time filters which introduce a phase delay whereas the current processing utilizes the zero-phase causal "filtering" from the least-squares de-trending. It is clear from this experiment that auto-correlation analysis is just as useful in the ideal case for monitoring heart rate as other contemporary methods.

# Chapter 3

# Heart Rate Estimation in Non-ideal Circumstances

Remote heart-rate estimation is not terribly practical when the subject is expected to remain completely still. A wide variety of applications become readily available when discussing the need for heart rate detection of persons who are in motion. While the number of use cases for acquiring the heart rate of moving subjects is great, so is the complexity of the problem. In order to prepare to minimize the effects of motion-induced noise on the signal, two motion-compensation techniques will be explored.

## 3.1 Kanade-Lucas-Tomasi (KLT) Algorithm

The KLT algorithm was developed over 20 years ago in a combined effort to estimate the rotation and translation of a detected object over a video sequence [19, 20]. The procedure attempts to identify features within a frame corresponding to the lowest eigenvalues in an image (because they are less like the total image, thus they are distinctive). Between consecutive frames a rotation matrix is estimated which will minimize the distance of the respective feature sets. This technique becomes valuable because it lends itself to facial tracking very easily.

In what has become traditional facial-tracking, classifiers are trained to identify Haar-like facial features in images. This method was first introduced by Viola and Jones to be used in cascaded classifiers to quickly and decisively rule out areas of an image which do not contain a face [21]. This method is most widely used in real-time applications because of its low computational complexity.

During the second stage of experiments, the same seven subjects from Experiment 1 were allowed to act naturally as they were asked a series of questions which would help elicit an emotional response. During the course of the interview the subjects' heart rates varied

as well as the motion-induced noise. By applying the same tests used in the first stage, it was discovered whether these techniques are effective in high-noise, practical scenarios.

As shown in figure 3.1, the reference frame (or the window rotated by the KLT algorithm) was used to provide the motion compensation for the inner ROI window used to extract data. If the compensation was ideal, the ROI would correspond to the same region of the forehead from frame to frame regardless of how the head moved. The initial facial frame must first be set via traditional face detection techniques reported on earlier.

Once the frame is set, the eigen-features are computed within the boundaries of the frame. In subsequent images, the eigen-features are computed within the previous image's reference frame. The rotation and translation between the two feature sets are then applied to the reference frame.

The greatest advantage of using the KLT algorithm is the higher accuracy and almost non-existent "jitter" the reference frame had in tracking the face. In other motion-compensation methods used for this same purpose a lowpass zero-phase filter must be applied to the path of the window in order for it to smoothly follow the subject. Because KLT does not rely on filtering, it is very applicable for real-time applications. One other advantage is that is it can be highly automated. Relying on already robust face detection, the algorithm does not need any tuning in order work on a variety of different subjects out of the box. The one caveat is that the initially frame of reference must be set when the subject's head is upright and looking directly at the camera.

There are some disadvantages in using the KLT algorithm in practical situations. While the KLT can adapt to the translational and rotational movements of the subject, it does not account for left-to-right or up-and-down motions of the head. In some cases, the movements can be so severe that the reference frame is lost altogether and must be reset.

Additionally the frame of reference is prone to "drifting" and shrinking if the subject undergoes rapid movements (like jogging in place). This may be due to the higher degree of feature disparity between two images which causes the estimator to maker larger errors. Over time these errors result in an offset or skewness of the original frame of reference.

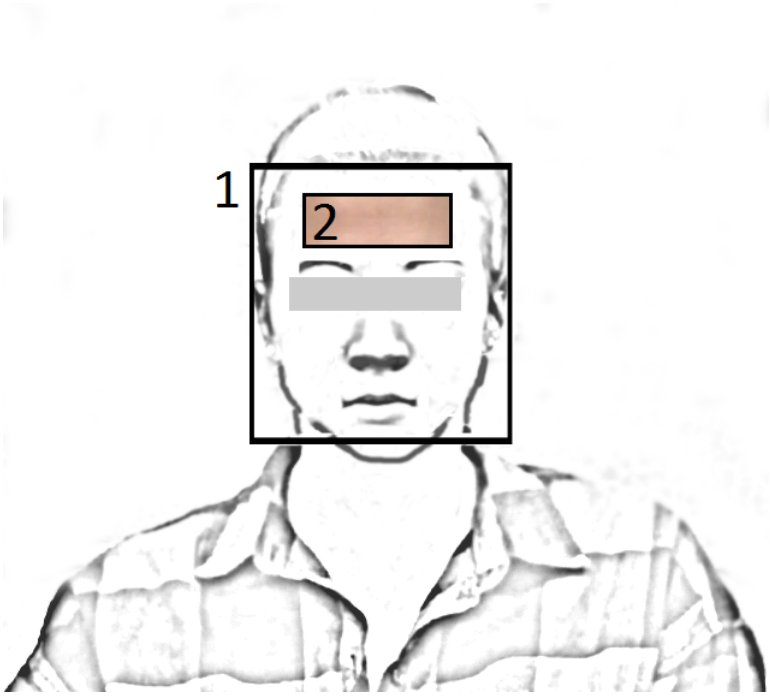



Fig. 3.1: A captured image of the KLT algorithm tracking a subject. Box 1 indicates the reference frame which is rotated according to the subject's motion. Box 2 is the ROI selected within the window.

This condition can be circumvented if the frame of reference is periodically reset by a face detector.

### 3.2 Video Segmentation

An alternative solution to motion compensation is making use of video segmentation. Image segmentation is a maturing field which makes use of minimum spanning trees which allow for similar pixels to be grouped into regions. For more consistent regions, the segmentation is usually run again over the regions to allow for small obscure regions to be combined with larger ones.

In a recent project undertaken by Grundmann *et al.*, this method was extrapolated to three dimensions applied across video segments using spatio-temporal trees [22]. In their approach, the color distances of the adjacent pixels in space and time were used in the graph construction. As in the previous two dimensional case, the video was processed in a hierarchical manner to ensure less disparity of regions over time.

In figure 3.2, the original and segmented video frames are placed next to each other to give a sense for the outcome of video segmentation. It is clear from the image to the right which area would be used as the ROI. This method of motion compensation is entirely different from the KLT method with its own set of strengths and weaknesses.

The obvious strength is that, if properly selected, the ROI will always contain the maximum number of eligible pixels without including bad pixels (like those associated with hair or eyebrows). It may not provide a mapping of a set of pixels in one time frame to another time frame, but this is irrelevant in the case where spatial averaging is used. The greatest advantage of video segmentation is that it will always associate like-pixels to the same region, almost without any regard to the orientation of the head. That is more than can be said for the KLT approach.

There are, however, a few disadvantages with this form of motion compensation as well. It is a little less practical in that the selection of a facial region is not very autonomous. One could apply a face detector and select the ROI as the largest region above the eyes, but there are no guarantees that multiple regions won't fit that criteria. The parameters


Fig. 3.2: A comparison between original anonymous image (left) and segmented image (right). The "ROI" marker indicates which shaded area would be used to collect pixel data.

associated with an optimal segmentation are also different depending on the scene of the video. For example, a subject who is standing in front of a smooth skin-colored backdrop requires finer segmentation as to not mistakenly include background pixels in the ROI. There is an optimal level of segmentation that must be carefully chosen for each case.

It was also found in scenarios ranging from little to extreme motion that the regions had a tendency to "shift" over time. For many subjects, a region which was manually selected as the ROI could move from the forehead down to the cheek or begin including non-PPG noisy areas like the eyes. To compensate, as with the KLT method, the ROI should be periodically checked by a face detector and, if necessary, changed.

### 3.3 Evaluating the Performance of Experiment 1 Analysis Techniques

During the second stage of experiments, the same seven subjects from Experiment 1 were allowed to act naturally as they were asked a series of questions which would help elicit an emotional response. During the course of the interview the subjects' heart rates varied as well as the motion-induced noise. One two-minute video segment from each participants' interview was selected which would allow for both motion-compensation approaches to be used. By applying the same tests used in the first stage, it was discovered whether these

techniques are effective in high-noise, practical scenarios.

In order to prove the effectiveness of one technique over another, there must be a set of metrics to compare. It is difficult, however, to find a sensible metric which really describes how an estimator performs. Most of the analysis, so far, has centered on how closely the estimator path conforms with the pulse oximeter average heart rate. This naturally leads one to think in terms of error. Heart rate monitoring is not about finding the exact instantaneous heart rate but rather about following the trend of heart rate over time. Thus large deviations in the heart rate would cause a misinterpretation of the trend whereas small errors are acceptable. The metric of choice, then, will be the percentage of time that the estimator lies within a given error tolerance.

This metric alone may not encompass the essence of a good heart rate estimator. After all, one could conceive of an estimator being correct only every other sample yet still garnering an accuracy score of 50%. The metric of average time of consecutive accurate readings should also be taken into account in order to determine the sparseness of the estimator's accuracy.

### 3.3.1 Revisiting Peak Detection

One of the main drawbacks of peak detection is it is very sensitive to noise. In order to observe the periodic signal of interest in the time domain it must be dominant in the observed data sequence. With such low expectations, the results in Table 3.1 are not surprising.

Even though the segmentation motion-compensator marginally outperformed the KLT implementation, both estimators functioned very poorly. In some cases a random number generator would have better statistics than those found in Table 3.1. This data reflects the peak detector's greatest weakness. While delivering rapid and frequent updates, the peak detector ignore a vast amount of the data available for estimating the signal's dominant frequency. Such methods like the FFT use every data sample for frequency analysis but the peak detector is only concerned with the largest points.

The sudden spikes and drops in the signal due to the changing orientation of the head

Table 3.1: Accuracy statistics for the peak detector estimator

| Subject | KLT | | Segmentation | |
|---|---|---|---|---|
| | % Accurate | Avg. Time | % Accurate | Avg. Time |
| 1951 | 6.5% | 1.5 s | 22.6 % | 1.4 s |
| 1952 | 4.5% | 1.0 s | 3.4 % | 1.0 s |
| 1953 | 15% | 2.33 s | 16 % | 1.25 s |
| 1954 | 17% | 1.23 s | 19 % | 1.5 s |

create many false peaks which give the estimator a very noise look. Figure 3.3 is a sample of the estimators' output compared to the reference heart rate. This demonstrates the need for using both the accuracy percentage and the correct-consecutive-reading time average metric when comparing techniques. The shorter average time windows indicate that the estimator is much more sporadic and less dependable as seen from the previous figure. More metrics could be used such as the variance in time of the correct consecutive readings, but for the purposes of this study these two metrics have been found to be both simple and sufficient.

### 3.3.2 FFT and PSD Revisited

Frequency analysis has its own set of strengths and weaknesses. As mentioned previously, the FFT uses all discrete data points in order to assess the spectrum of the signals. This makes the best use of the data already on hand. There are, however, two detractors to the signal quality.

First, spectral leakage caused by large, low-frequency signals can distort the frequency of interest. Despite enormous efforts to the contrary, there is still a great deal of noise due to the motion of the subject. Motion-compensation, while keeping the ROI fixed regardless of the subject's position, cannot account for the changes in lighting. Though lighting variation may seem subtle, one must remember that the extremely weak PPG signal is even smaller still.

The second source of spectral noise is due to spectral spreading. Abrupt movements of the head, even seemingly small ones, cause the data to shift in amplitude. These shifts, being seen as a step function, cause large harmonic ripples to be seen throughout the

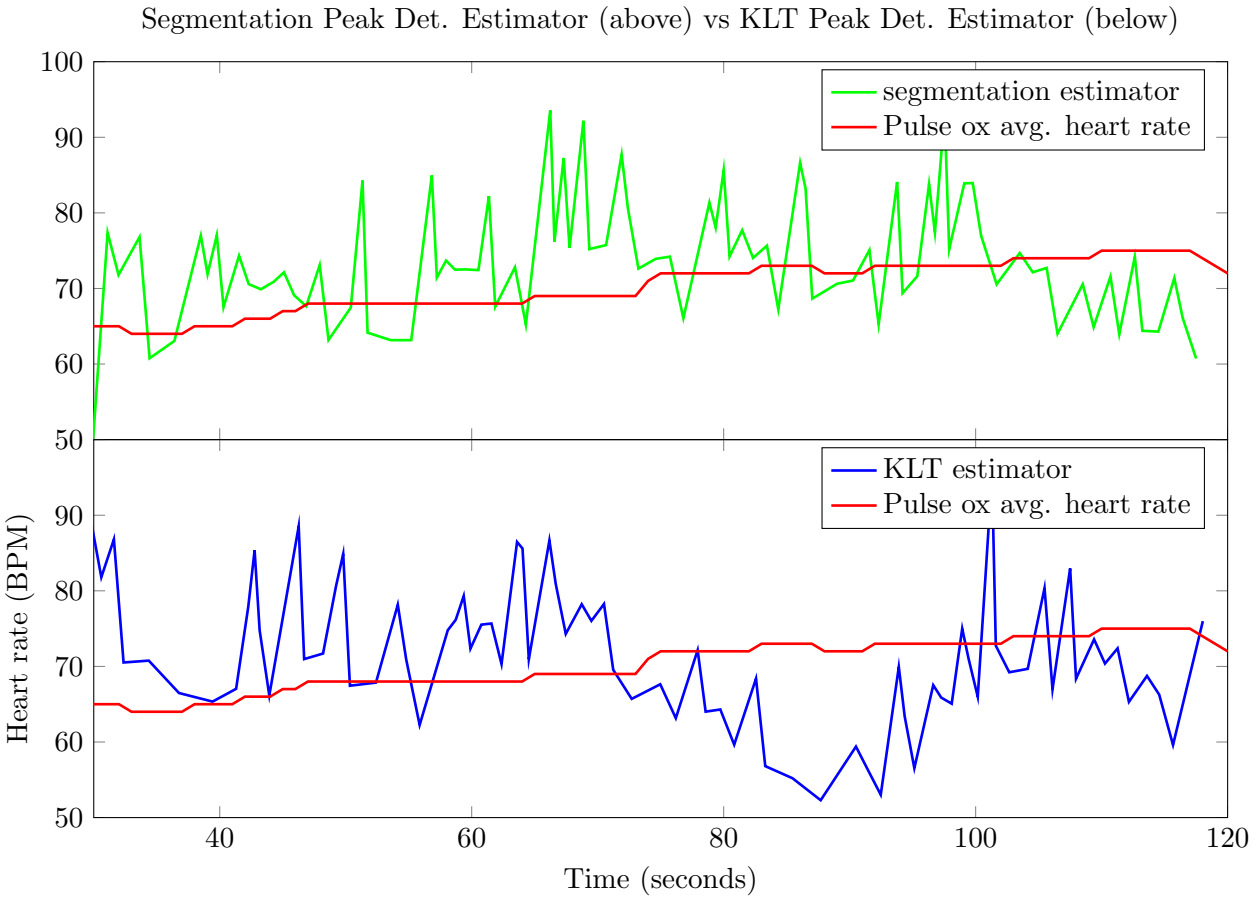



Fig. 3.3: The auto-regressive peak detection estimators were used with segmentation and KLT motion-compensation (separately).

spectrum. Results were collected for both the sliding FFT and auto-correlation techniques and displayed in Table 3.2 and Table 3.3. Both types of motion-compensation were used.

The average time of correct consecutive frames in the autocorrelation approach appears to be much lower than that of the sliding-window FFT. This is due to the fact that the update time for the PSD estimator is 33 ms as opposed to one full second in the FFT case which cause that particular metric to appear less appealing.

After experimentation, it is found that the PSD estimator coupled with the segmentation motion-compensator consistently produce the most accurate results. The results, however, are not far from those of the peak-detection estimator and are still fall very short of the ideal heart rate estimator.

As seen in figure 3.4, there are abrupt shifts on the order of several pixel values which cause a great deal of spectral noise in the frequency domain. These shifts can be associated with movements of the forehead (i.e. raising an eyebrow) or the surface of the head picking up reflective glare. Having demonstrated this, it comes as no surprise that the large noise harmonics can dominate the heart rate signal (see figure (b)).

Table 3.2: Accuracy statistics for the sliding-window FFT estimator

| Subject | KLT | | Segmentation | |
|---|---|---|---|---|
| | % Accurate | Avg. Time | % Accurate | Avg. Time |
| 1951 | 23.7% | 5.5 s | 24.7 % | 7.7 s |
| 1952 | 0% | 0 s | 0 % | 0 s |
| 1953 | 4.7% | 1.5 s | 7.5 % | 1.75 s |
| 1954 | 7.5% | 1.75 s | 2.2% | 2 s |

Table 3.3: Accuracy statistics for the sliding-window PSD estimator

| Subject | KLT | | Segmentation | |
|---|---|---|---|---|
| | % Accurate | Avg. Time | % Accurate | Avg. Time |
| 1951 | 20% | 0.77 s | 30 % | 0.92 s |
| 1952 | 6.1% | 0.726 s | 1.44 % | 0.27 s |
| 1953 | 4.5% | 0.25 s | 20 % | 0.764 s |
| 1954 | 10% | 0.46 s | 16% | 1.4 s |

Abrupt shifts also effect the computation of the auto-correlation function. Figure 3.5 shows how large artifacts in the input signal (upper) translates a large, periodic auto-correlation function (lower). To clarify, the autocorrelation function is represented along the vertical axis while x axis indicates the time index of the data from which the function was computed. Thus, by taking a cross-section of the graph with respect to some time index $t$ one would be left with the autocorrelation associated with that time. This function is then operated on by the FFT and analyzed for the heart rate frequency component.

In each red box it can be seen that the input waveform experiences some large change which may appear to be similar to a step or rect function. When the signal is de-trended, the step function is projected to a large low-frequency harmonic which statistically washes out the under-lying PPG signal. This then causes the resulting frequency spectrum function to suffer from the same problems found with the sliding-window FFT. The larger and steeper the artifact is, the greater effect it has on the PSD output.

### 3.3.3 Insufficient Techniques

From Experiment 2 it is clear that these simple approaches for uncovering the subtle PPG signal may have worked well under ideal circumstances but are insufficient for more real-world practical scenarios. The signal power is simply too weak in most of these noise-ridden cases. Using basic frequency and statistical techniques are not enough to build a quality and reliable estimator of the heart rate. These results, although disappointing, provide a valuable insight in what direction to take next. One must now take this problem out of the realm of simply analyzing the input signal and place it in that of statistical models and parameter estimation. To do this problem justice it must be considered in the light of probabilistic error minimization.

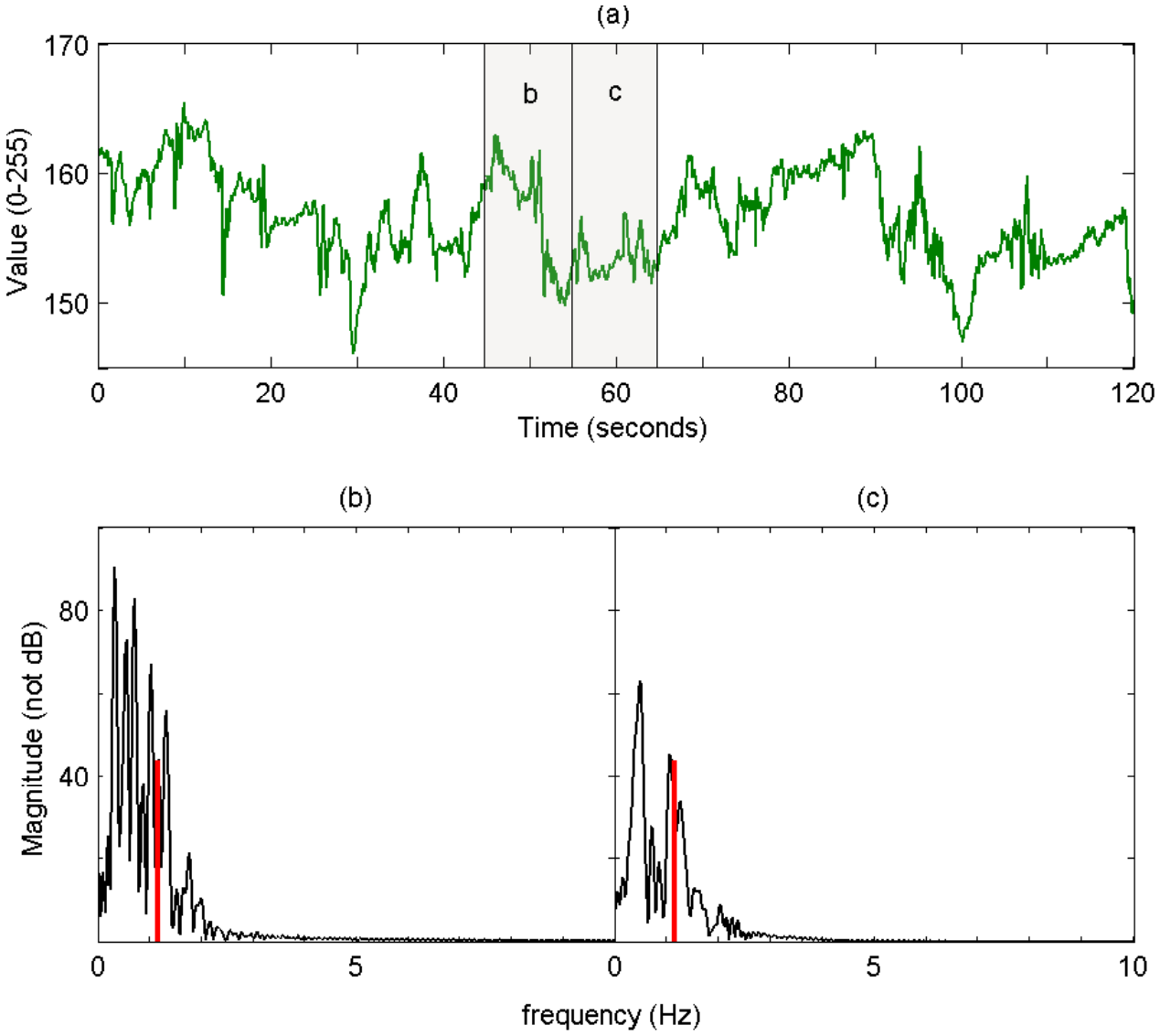



Fig. 3.4: Comparing the effects of spectral spreading in the output of the sliding-window FFT heart rate estimator. Figure (b) and (c) represent the frequency magnitude plots of the respective time-domain windows indicated in Figure (a). Figure (a) represents the green-channel collected from the ROI using segmentation motion-compensation.


Fig. 3.5: Comparing the transient effects of the time-domain data to the large auto-correlation samples. The red boxes highlight the artifacts of interest.

# Chapter 4

# A Model-Based Approach

Under realistic circumstances, the heart rate is simply immersed in large quantities of noise. It is clear that pure analysis of the data is not enough to separate the signal of interest from the rest of the channel. Experiments 1 and 2 have added insight into the challenging task at hand. The problem will now be considered in a different light. Consider the following system level equation which is a combination of (2.3) and (2.6).

$$x_i(t) = h_i(t) + m_i(t) + n \tag{4.1}$$

$$n \sim N(0, \sigma_n) \text{ i.i.d.}$$

In this model, $h_i(t)$ is the discretized PPG signal at pixel $i$ on the forehead. The term $m_i(t)$ represents the underlying motion, lighting interference, and other physiological signals (such as respiration) all lumped into one signal. The Gaussian random variable $n$ represents the noise due to the camera electronics. The observation equation is simply put as:

$$y_i(t) = Q(x_i(t)) = v_i(t) \in \{0, 1, 2, \ldots, 2^b - 1\}. \tag{4.2}$$

The function $Q(.)$ represents the quantization of the system variable where $b$ is the number of bits in the quantizer. The system and observation equations are consistent with those used in other studies and are certainly not unreasonable. This now provides a framework for error minimization

## 4.1 Design Constraints

In these equations, let it first be considered that $u_i$ is, in fact, deterministic and not the output of a random process. This assumption seems reasonable when considering what

is entailed in $u_t$. Noise due to physiological signals, sudden movements, and illumination changes are not actually random. From a signal perspective, the sources of this noise appear to follow a first-order Markov model.

Having determined that there is only one random noise source governing this problem, it is necessary to identify what constraints must be placed on this system as an attempt is made to minimize the error between the true signal and the estimate. Three key observations are made to help formulate an algorithm.

### 4.1.1 Likelihood Function

It is natural to start with defining a likelihood function which will help determine the most probable value of the input signal given an observation. The only random signal in the observation equation is assumed to be Gaussian which means the likelihood function is of the same distribution. Due to the quantization effect of $Q(.)$, the following formulation must be made:

$$\begin{aligned} P(y_i(t) = v|h_i(t), u_i(t)) &= (2\pi\sigma_n)^{-\frac{1}{2}} \int_{v-\frac{1}{2}-u_i(t)-h_i(t)}^{v+\frac{1}{2}-u_i(t)-h_i(t)} e^{-\frac{n^2}{2\sigma_n}} \mathrm{d}n \\ &= G(v + \frac{1}{2} - u_i(t) - h_i(t), \sigma_n) - G(v - \frac{1}{2} - u_i(t) - h_i(t), \sigma_n), \end{aligned} \tag{4.3}$$

where $G(., \sigma_n)$ is a zero-mean Gaussian distribution evaluated at the first argument with variance $\sigma_n$. This expression is easy enough to compute, however there must be another constraint to help separate $h_i(t)$ and $u_i(t)$.

### 4.1.2 System Noise Markovity

As mentioned before, $u_i(t)$ is assumed to be first-order Markov. This assumption drastically reduces the complexity of the algorithm. Some linear combination of the previous frame's pixels forms the current frame's pixels with some additive noise. This leads one to the following set of equations:

$$\mathbf{u}(t+1) = A\mathbf{u}(t) + B\mathbf{w}(t) \tag{4.4}$$

$$\text{where } \mathbf{w} \sim N(0, Q_w)$$

$$\text{and}$$

$$\mathbf{u}(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_M(t) \end{bmatrix},$$

where $M$ is the number of pixels being considered. In this case, the noise can be treated as an error. The $A$ and $B$ matrices are assumed to be known. To reduce the noise and improve the estimate of $\mathbf{u}(t)$, the following minimization may be considered with respect to the L2 norm:

$$\begin{aligned} \hat{\mathbf{u}}(t) &= \underset{\mathbf{u}(t)}{\operatorname{argmin}} \|B\mathbf{w}(t)\|_2^2 \\ &= \underset{\mathbf{u}(t)}{\operatorname{argmin}} \|\mathbf{u}(t+1) - A\mathbf{u}(t)\|_2^2 . \end{aligned} \tag{4.5}$$

This is a trivial least squares problem to solve. The effect of applying this constraint would be to "smooth out" all of the AC components in $u_i(t)$ making it representative of simply the combination of all deterministic non-PPG signals in the system.

### 4.1.3 Signal Predictability

It is reasonably assumed that $h_t(t)$ is periodic. Periodic signals can be approximated by a truncated Fourier series of harmonically-related sinusoids. One fact about sinusoids is that a linear combination of previous samples can exactly compute future samples. In other words, they're predictable.

$$\sum_{k=0}^{2L-1} h_i(t-k)p_k \approx h_i(t+T). \tag{4.6}$$

In (4.6), $T$ is an arbitrary time constant (not necessarily the period as that is not given). As a rule of thumb, it should stay relatively small so as to increase the amount of data that can be used in the minimization. The term $p_k$ represents a set of coefficients of length $2L$ that also can be estimated in this problem. Finally a set of equations have been derived which help maximize the likelihood of the input signal as well as separate it into the interpreted noise and heart rate signals.

## 4.2 Super Pixel Model

Before attempting to identify a solution for the previous set of equations, there are some simplifications that must be performed. The first major reduction in the model is to assume that the heart rate signal $h_i(t)$ is in fact uniform across all pixels. While this may not be a strictly valid assumption as per the discussion on page 15, it does greatly reduce the complexity of the algorithm which is of greater importance in this experiment.

The other simplification comes by way of sheer data volume. The number of pixels that can be found on a forehead in HD video is large. That would make the size of the vectors and matrices used in this minimization unwieldy. In an effort to minimize the number of samples without having to discard hard-won data, the principle of local pixel-averaging is employed once again. These local averages of pixels will be referred to as super pixels. These benefit the model for two reasons.

First, the aforementioned data size is reduced significantly when only considering super pixels making the problem more computationally feasible. Second, by averaging one gains an effective resolution which is much higher than the bit depth of the quantizer. With a sufficient number of pixels, the observed data becomes nearly continuous, thus (4.2) and (4.3) simply become

$$\begin{aligned} y_i(t) &= x_i(t) \\ &= h(t) + u_i(t) + n, \end{aligned} \tag{4.7}$$

and

$$P(y_t(t) = x_i(t)|h(t), u_i(t)) = G(x_i(t) - h(t) - u_t(t)). \tag{4.8}$$

The final assumption for this simplified model is to simply approximate $\mathbf{u}(t+1)$ as

$$\mathbf{u}(t+1) = \mathbf{u}(t) + B\mathbf{w}(t), \tag{4.9}$$

which makes for a very simple minimization problem:

$$\hat{\mathbf{u}}(t) = \underset{\mathbf{u}(t)}{\operatorname{argmin}} \|\mathbf{u}(t+1) - \mathbf{u}(t)\|_2^2 . \tag{4.10}$$

### 4.3 Gradient Descent Optimization

Despite the efforts made to create the most simplified model, there is no closed-form solution which allows one to compute $h(t)$ and $u_i(t)$ directly. When taking into account the likelihood, predictability constraint, and Markovity constraint, the solution is simply intractable.

The easiest solution, in this case, is to use gradient descent to optimize the estimates. Gradient descent rests on the ability to differentiate some cost function with respect to the parameters of interest and then advance those parameters in the negative slope direction (because this is a minimization problem). In creating the cost function, the likelihood function is modified to use the negative log-likelihood function. The sign change is to change it from maximization or minimization. The log-likelihood makes differentiation within Gaussian distributions simple and is justified due to the monotonically increasing property of logarithmic functions. Let the cost function $J$ be defined as follows:

$$\begin{aligned} J(\mathbf{h}, U, \mathbf{p}, Y) = &- \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{\frac{1}{2\sigma_n^2}(y_{i,j}-h_i-u_{i,j})^2} + \lambda_1 \|H\mathbf{p} - \mathbf{h}_{t+T}\|_2^2 \\ &+ \lambda_2 \sum_{i=1}^{N-1}\sum_{j=1}^{M} \|p_j(i+1) - p_j(i)\|_2^2 , \end{aligned} \tag{4.11}$$

where

$$H = \begin{bmatrix} h(0) & h(1) & \cdots & h(2L-1) \\ h(1) & h(2) & \cdots & h(2L) \\ \vdots & & \ddots & \vdots \\ h(N-2L) & h(N-2L-1) & \cdots & h(N-1) \end{bmatrix}$$

$$\mathbf{p} = \begin{bmatrix} p(0) \\ p(1) \\ \vdots \\ p(2L-1) \end{bmatrix}$$

$$Y = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_M \end{bmatrix}$$

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_M \end{bmatrix}.$$

Both $\lambda_1$ and $\lambda_2$ are user-defined parameters to ensure equal weighting among the different constraints. Fortunately, due to the previous simplifications, this cost function is differentiable. Exploiting this fact, the update equations for both $\mathbf{h}$ and $\mathbf{u_i}$ will appear as:

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu \nabla_{\mathbf{h}_k} J, \tag{4.12}$$

$$\mathbf{u}_{j,k+1} = \mathbf{u}_{j,k} - \mu \nabla_{\mathbf{u}_{j,k}} J, \tag{4.13}$$

$$\nabla_{\mathbf{h}_k} J = \frac{\partial}{\partial \mathbf{h}_k} J = \frac{1}{MN} \sum_{j=1}^{M} \left[ \frac{1}{\sigma_n^2} (\mathbf{h}_k - \mathbf{y}_j + \mathbf{u}_j) \right] + \lambda_1 P^T P \mathbf{h}_k, \tag{4.14}$$

and

$$\nabla_{\mathbf{u}_{j,k}} J = \frac{\partial}{\partial \mathbf{u}_{j,k}} J = (\frac{1}{\sigma_n^2} + 4\lambda_2)\mathbf{u}_{j,k}(t) - \frac{1}{\sigma_n^2}(\mathbf{y}_j(t) - \mathbf{h}(t)) - 2\lambda_2(\mathbf{u}_{j,k}(t+1) + \mathbf{u}_{j,k}(t-1)), \quad (4.15)$$

where

$$\mathbf{u}_j = \begin{bmatrix} u_j(0) \\ u_j(1) \\ \vdots \\ u_j(N-1) \end{bmatrix}, \mathbf{y}_j = \begin{bmatrix} y_j(0) \\ y_j(1) \\ \vdots \\ y_j(N-1) \end{bmatrix}, \mathbf{h} = \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(N-1) \end{bmatrix},$$

$$P = \begin{bmatrix} p(0) & p(1) & \cdots & p(2L-1) & 0 & \cdots & -1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \cdots & \ddots & \ddots & \cdots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & & \ddots & \ddots & \cdots & \ddots & \vdots \\ 0 & \cdots & & & & & & & 0 & -1 \end{bmatrix}.$$

Meanwhile, there is a closed-form solution for $\mathbf{p}$ since it neither depends on the Markovity-constraint or likelihood:

$$\mathbf{p}_{k+1} = 2\lambda_1 (H^T H)^{-1} H^T \mathbf{h}_{t+T}, \quad (4.16)$$

where

$$\mathbf{h}_{t+T} = \begin{bmatrix} h(T) \\ h(T+1) \\ \vdots \\ h(N-1) \end{bmatrix}.$$

## 4.4 Experimentation

This algorithm designed on the simplified system model was put into practice on the same data sets used in Experiment 1. No formal estimator was built from this algorithm as there is yet no iterative solution base. The primary focus of this experiment is to implement the adaptive formula and observe it the input signal is indeed extracted from the observed samples.

### 4.4.1 Setup

Ten seconds of data recorded at 30 FPS were collected from each of the super pixels outlined in figure 4.1. Each super pixel was the average of approximately 12,000 actual pixels. By experimentation, $\lambda_1$, $\lambda_2$, and $\mu$ were determine to deliver the optimal results such that the minimization occurred in a balanced way across all of the constraint factors and converged the quickest. A reiteration of the parameters:

$$\text{total number of samples } (N) = 300$$

$$\text{sensor count } (M) = 21$$

$$\text{super pixel area} = 11800$$

$$\text{estimate } \sigma_n = 0.0266$$

$$\lambda_1 = 100$$

$$\lambda_2 = 10000$$

$$T = 25$$

$$L = 10$$

$$\text{step size } \mu = .000025$$

$$\text{iterations} = 100$$

Initialization:

$$U = Y$$

$$\mathbf{h} = \mathbf{0}$$

$$\mathbf{p} = \frac{1}{2L} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

### 4.4.2 Results

This gradient-descent algorithm was applied to several Experiment 1 videos, all using close to the same parameters mentioned under the setup. Figure 4.2 shows the observed data measured by the super pixel array.

After many iterations, the cost function significantly reduced which indicates the algorithm is operating as it was designed. Figures 4.3, 4.4, and 4.5 contain the results of the algorithm.

From figure 4.5 it is obvious that the PPG signal is indeed present marking this experiment as a success. The fundamental frequency and its harmonics correspond to a 60 BPM which is exactly the heart rate associated with this particular data set. The noise data $U(t)$ appears to have captured the powerful low-frequency underlying noise while rejecting the AC periodic signals. While the heart rate was not hard to compute using traditional analysis in experiment 1, the model-based approach makes one important point: there is merit to solving the problem of remote heart rate estimation using intelligent models as found in other mature fields.

This experiment demonstrates that perhaps the answer to tracking heart rate in scenarios where conventional methods fail lies in accurate system modeling and minimization. The ability to make a reasonable heart rate estimate in the presence of great amounts of noise is out of reach with current analysis techniques. By moving the art of the field into using proven algorithms designed for dealing with noise such a the Kalman filter, the insurmountable challenge may actually become achievable.

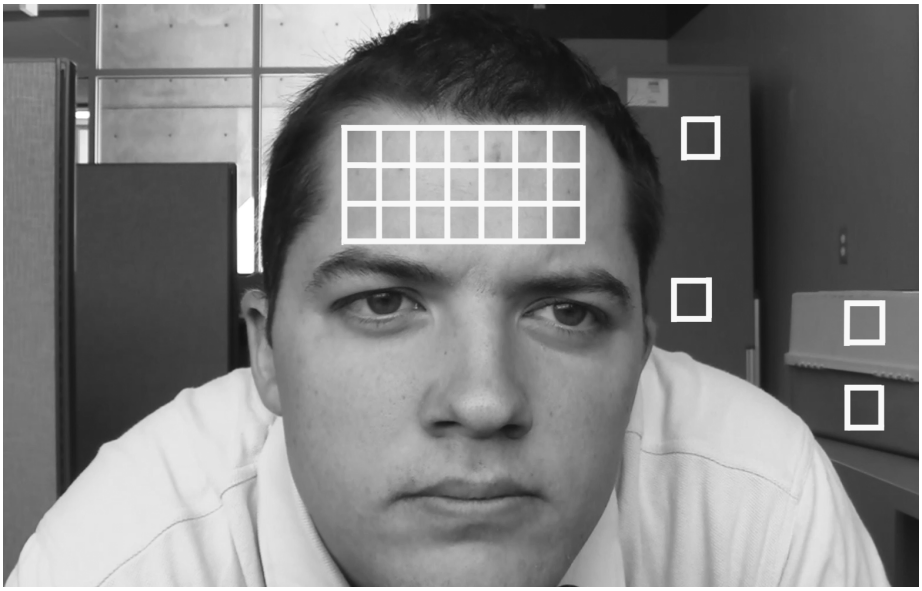


Fig. 4.1: A 3 by 7 pixel grid was used to collect "super pixels" for the algorithm. Four independent noise samples were taken to estimate $\sigma_n^2$.

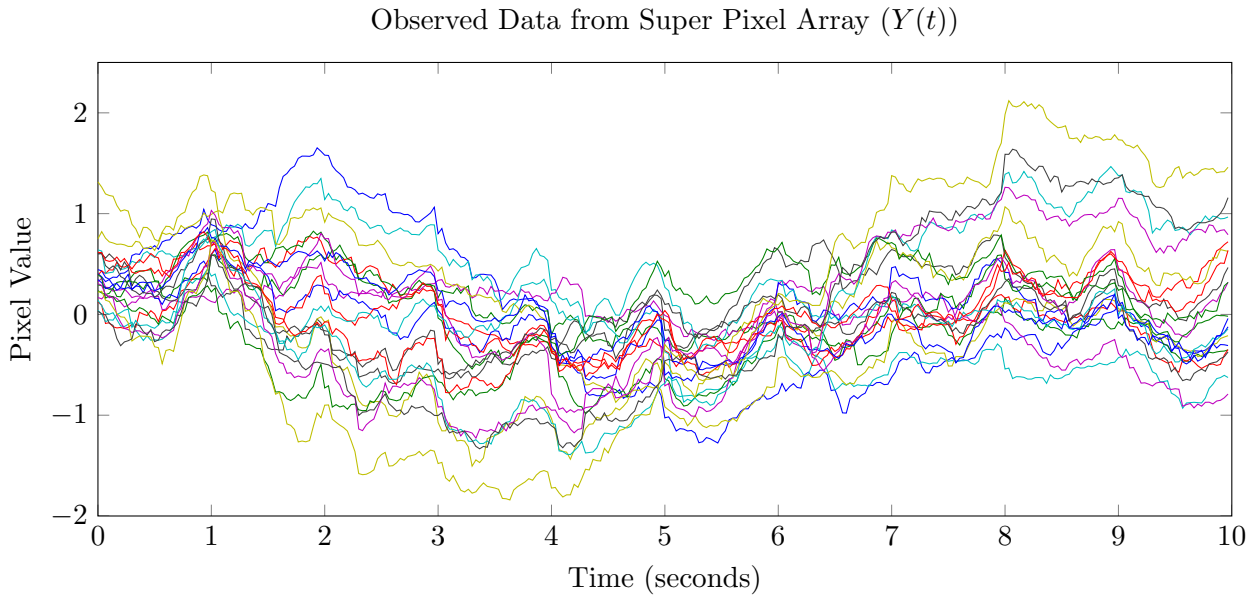



Fig. 4.2: The measured data of the M forehead super pixels is plotted over time.

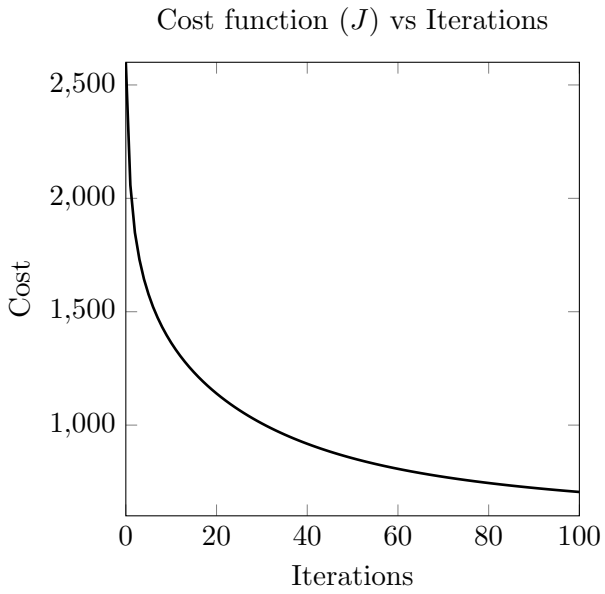



Fig. 4.3: The cost function $J(\mathbf{h}, U, \mathbf{p}, Y)$ as measured over iterations.

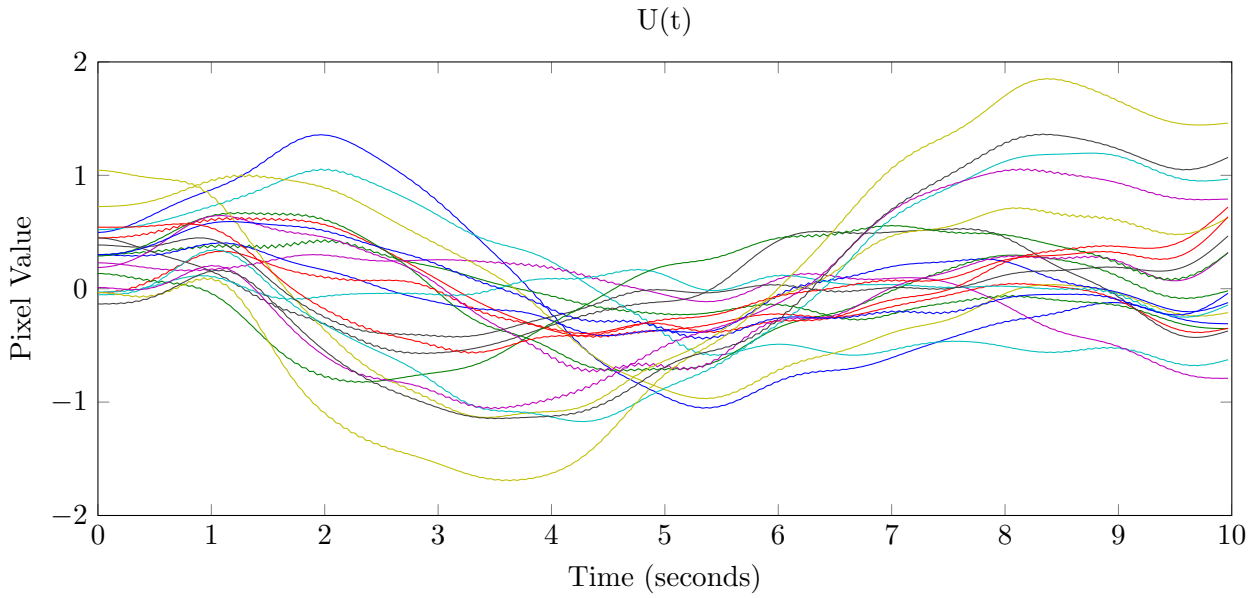



Fig. 4.4: The estimated 1st-order Markov noise $U(t)$.

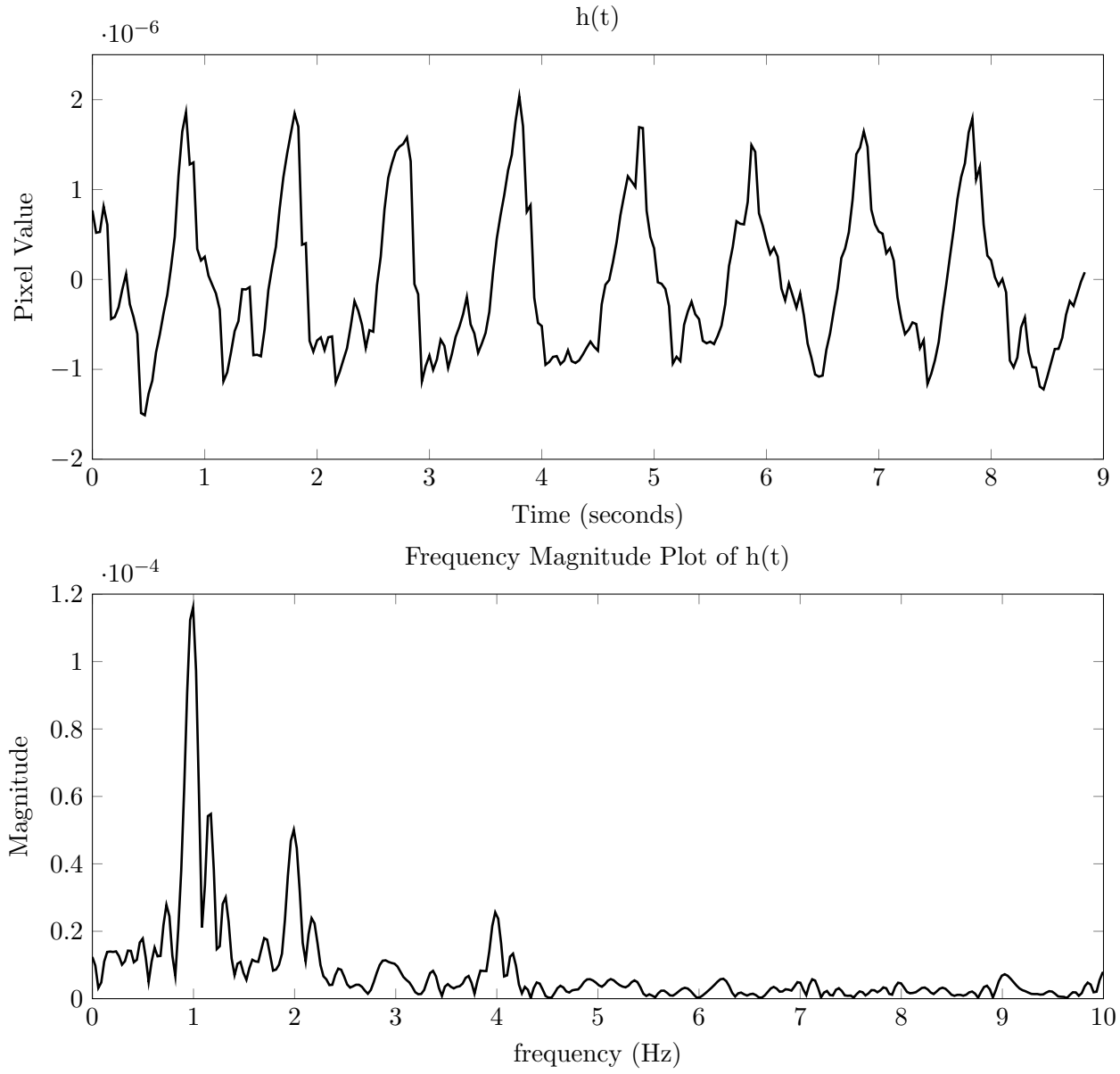



Fig. 4.5: The self-predicting signal $h(t)$ which represents the heart rate is shown along with its frequency spectrum.

# Chapter 5

# Research Conclusion

The goal of this research was to determine what could be learned about estimating and tracking the human heart rate using a consumer-grade camera. The use cases for this technology have a vast range which partially motivates the question in the first place. One can conceive of its utility when wanting to monitor an infant at risk of SIDS without placing wired sensors around or near the baby's body. Another might see this in light of security applications by using the heart rate from cameras as a preliminary lie detector. There are various ways in which the solution to this difficult problem might find itself in medical, research, and industrial cases alike. The advances in computer vision and remote biometrics further adds value to this field.

The novelty of remote heart rate tracking has encouraged several new algorithms in recent years. All of these algorithms, while working very well in some circumstances, have certain deficiencies which were addressed in this research. It was a goal to help reduce the latency of the heart rate update which plagues every Fourier-based estimator. While this was accomplished using peak-detection methods, there was a trade-off with the estimate reliability.

Motion compensation was also addressed which helps to acquire the signal of interest and also reduce the noise due to the subject's movements. The two techniques explored, KLT and video segmentation, were found to be viable options for tracking the ROI. The use of such trackers in heart rate estimation has never been examined before this study.

It was also discovered that heart rate estimation based purely on band-pass filtering and spectrum analysis is simply insufficient for tracking the heart rate in most practical scenarios. A model-based approach, while not thoroughly explored, was deemed to be a viable option for heart rate extraction in the most generic sense. From Experiment 3 it

became clear that an alternative method exists which does not rely upon temporal filters to remove the noise. It has been identified that a more thorough examination of the PPG system model may help improve upon the model's accuracy and even make it robust in areas where conventional analysis-based estimators fail.

In conclusion, the world is rapidly changing and so is the way in which information is collected. The rapid growth of inexpensive cameras has prompted their innovative uses in ways which were never before imagined. The field of remote photoplethysmography is itself a very recent and challenging field. If mastered, it could espouse great change in the respective areas of the medical, security, and exercise communities. This study may not have satisfied the unanswered questions, but it has provided a well-needed step in the advancement of the art.

# References


[1] D. Obeid, G. Zaharia, S. Sadek, and G. El Zein, "Microwave doppler radar for heartbeat detection vs electrocardiogram," *Microwave and Optical Technology Letters*, vol. 54, no. 11, pp. 2610–2617, 2012. [Online]. Available: http://dx.doi.org/10.1002/mop.27152

[2] M. Yang, Q. Liu, T. Turner, and Y. Wu, "Vital sign estimation from passive thermal video," *IEEE Conference on Computer Vision and Pattern Recognition.*, pp. 1–8, June 2008.

[3] A. V. J. Challoner, "Photoelectric plethysmography for estimating cutaneous blood flow," *Non-Invasive Physiological Measurements*, vol. 1, p. 125, 1979.

[4] N. G. Roald, "Estimation of vital signs from ambient-light non-contact photoplethysmography," Master's thesis, Norwegian University of Science and Technology, Department of Electronics and Telecommunications, 2013.

[5] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, p. R1, 2007. [Online]. Available: http://stacks.iop.org/0967-3334/28/i=3/a=R01

[6] L. Scalise, *Advances in Electrocardiograms - Methods and Analysis*, P. R. Millis, Ed., 2012. [Online]. Available: http://www.intechopen.com/books/advances-in-electrocardiograms-methods-and-analysis/non-contact-heart-monitoring"title="NonContactHeartMonitoring"⟩ NonContactHeartMonitoring

[7] 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. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-16-26-21434

[8] Y. Sun, C. Papin, V. Azorin-Peris, R. Kalawsky, S. Greenwald, and S. Hu, "Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam," *Journal of Biomedical Optics*, vol. 17, no. 3, pp. 037 005–1–037 005–10, 2012. [Online]. Available: http://dx.doi.org/10.1117/1.JBO.17.3.037005

[9] 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. 10 762–10 774, May 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-10-10762

[10] "Vital signs camera," 2012. [Online]. Available: http://www.vitalsignscamera.com

[11] "iphone heart rate monitor app," 2013. [Online]. Available: http://cardiio.com

[12] M. Rubinstein, "Analysis and visualization of temporal variations in video," Ph.D. dissertation, Massachusetts Institute of Technology, Feb. 2014.

[13] M. Lewandowska, J. Ruminski, T. Kocejko, and J. Nowak, "Measuring pulse rate with a webcam; a non-contact method for evaluating cardiac activity," *Federated Conference on Computer Science and Information Systems*, pp. 405–410, Sept. 2011.

[14] L. Nilsson, A. Johansson, and S. Kalman, "Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique," *Journal of Clinical Monitoring and Computing*, vol. 16, no. 4, pp. 309–315, 2000. [Online]. Available: http://dx.doi.org/10.1023/A%3A1011424732717

[15] A. Johansson and P. Oberg, "Estimation of respiratory volumes from the photoplethysmographic signal. part 2: a model study," *Medical Biological Engineering Computing*, vol. 37, no. 1, pp. 48–53, 1999. [Online]. Available: http://dx.doi.org/10.1007/BF02513265

[16] M. Rice, *Digital Communications: A Discrete-Time Approach.* Saddle River: Prentice Hall, 2009.

[17] A. Rustand, "Ambient-light photoplethysmography," Master's thesis, Norwegian University of Science and Technology, Department of Electronics and Telecommunications, 2012.

[18] E. Jacobsen and P. Kootsookos, *Fast, Accurate Frequency Estimators.* Hoboken: John Wiley Sons, Inc., 2007. [Online]. Available: http://dx.doi.org/10.1002/9780470170090.ch10

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, pp. 674–679, 1981. [Online]. Available: http://dl.acm.org/citation.cfm?id=1623264.1623280

[20] C. Tomasi and T. Kanade, "Detection and tracking of point features," *International Journal of Computer Vision*, 1991.

[21] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, 2001.

[22] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient hierarchical graph based video segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.