

# Remote Vital Signs Monitoring with Depth Cameras

by

Ryan Chu

A thesis submitted in conformity with the requirements  
for the degree of Master of Health Science  
Institute of Biomedical Engineering  
University of Toronto  
Toronto, Ontario

© Copyright by Ryan Chu 2020

# Remote Vital Signs Monitoring with Depth Cameras

Master of Health Science

2020

Ryan Chu

Institute of Biomedical Engineering

University of Toronto

## Abstract

Remote monitoring of vital signs provides a low-cost, non-intrusive method of assessing an individual's health. These systems have wide-ranging applications, including detection of patient deterioration in clinical settings and home monitoring of at-risk individuals.

This thesis evaluates a novel remote vital signs monitoring system that integrates 3-D depth information and light intensity data to increase tolerance to subject motion. We use 3-D depth analysis and measurement of reflected light intensity due to blood volume variations to estimate heart rate and respiratory rate, two important vital signs. This method achieves a mean error of 1.7 beats per minute for heart rate and 0.05 breaths per minute for respiratory rate.

Furthermore, techniques are also described to quantify the severity of motion artifacts, and methods of motion correction are applied to salvage physiological data from noisy signals. Future work will include examining additional data scenarios and improving analysis methods and motion tolerance.

# Acknowledgements

I would like to acknowledge the numerous individuals who made this project possible and supported me throughout its development. Firstly, my sincere appreciation goes to my co-supervisors, Dr. Ofer Levi and Dr. Alex Mihailidis, for their guidance and dedication over the past two years. In particular, Dr. Levi's passion for the project helped inspire me and other members of the lab to continuously push forward. My thanks goes to Dr. Joe Cafazzo and Dr. John Granton for their roles as my thesis committee and helping guide the direction of research. I would like to thank the following lab members as well: Eric Zhu and Dene Ringuette for their mentorship, as well as Maxx Wu, Lindsay Kuramoto, Matthew Downing, and Apoorva Srivastava for their contributions as undergraduate students. Finally, I would like to thank my family and friends for their many forms of support in an uncertain and challenging time, whether it was constant check-ins or keeping me company during long nights of writing. This would not have been possible without them.

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                 | <b>1</b>  |
| 1.1      | Motivation . . . . .                                | 1         |
| 1.1.1    | Detecting Patient Deterioration . . . . .           | 2         |
| 1.1.2    | Disadvantages of Contact Methods . . . . .          | 3         |
| 1.1.3    | Telemonitoring . . . . .                            | 4         |
| 1.1.4    | Public Health Applications . . . . .                | 5         |
| 1.2      | Overview of Vital Signs . . . . .                   | 5         |
| 1.2.1    | Heart Rate . . . . .                                | 5         |
| 1.2.2    | Respiratory Rate . . . . .                          | 6         |
| 1.2.3    | Oxygen Saturation . . . . .                         | 9         |
| 1.3      | Vital Sign Measurement Techniques . . . . .         | 9         |
| 1.3.1    | Light Interaction with Tissue . . . . .             | 10        |
| 1.3.2    | Retrieving Vital Signs from Camera Images . . . . . | 14        |
| 1.3.3    | Head Ballistocardiography . . . . .                 | 18        |
| 1.3.4    | Depth Sensing . . . . .                             | 20        |
| 1.4      | Thesis Objectives . . . . .                         | 23        |
| <b>2</b> | <b>Experimental Methods</b>                         | <b>26</b> |
| 2.1      | Equipment . . . . .                                 | 26        |
| 2.1.1    | Intel RealSense Cameras . . . . .                   | 26        |

|          |   |           |
|----------|---|-----------|
| 2.1.2    | Camera Setup and Configurations . . . . .               | 29        |
| 2.1.3    | GE Dash Patient Monitor . . . . .                       | 31        |
| 2.2      | Extracting Physiological Data . . . . .                 | 32        |
| 2.2.1    | Retrieving Images . . . . .                             | 32        |
| 2.2.2    | Regions of Interest . . . . .                           | 32        |
| 2.2.3    | Physiological Signal Processing . . . . .               | 35        |
| 2.3      | Motion Tolerance Techniques . . . . .                   | 37        |
| 2.3.1    | Identifying Motion-Corrupted Signal . . . . .           | 37        |
| 2.3.2    | Motion Rejection . . . . .                              | 40        |
| 2.3.3    | Motion Compensation . . . . .                           | 40        |
| 2.4      | Summary . . . . .                                       | 41        |
| <b>3</b> | <b>Results and Analysis</b>                             | <b>42</b> |
| 3.1      | Overview . . . . .                                      | 42        |
| 3.2      | Heart Rate Results . . . . .                            | 45        |
| 3.2.1    | Extracting Blood Volume Pulse . . . . .                 | 45        |
| 3.2.2    | Comparing Results of Different ROIs . . . . .           | 46        |
| 3.3      | Respiratory Rate Results . . . . .                      | 49        |
| 3.3.1    | Extracting Respiratory Waveform . . . . .               | 49        |
| 3.3.2    | Comparing Results of Different ROIs . . . . .           | 50        |
| 3.3.3    | Strip Analysis for Respiration Classification . . . . . | 51        |
| 3.4      | Motion Analysis . . . . .                               | 56        |
| 3.4.1    | Rejecting Severe Motion . . . . .                       | 56        |
| 3.4.2    | Single-Camera Motion Compensation . . . . .             | 60        |
| 3.4.3    | Multi-Camera Motion Compensation . . . . .              | 63        |
| 3.5      | Participant Data Analysis . . . . .                     | 65        |
| 3.6      | Summary . . . . .                                       | 69        |

|          |  |           |
|----------|--|-----------|
| <b>4</b> | <b>Future Work</b>                                       | <b>71</b> |
| 4.1      | Recruitment of External Participants . . . . .           | 71        |
| 4.2      | Improved Scenarios and Analysis Methods . . . . .        | 73        |
| 4.3      | Multi-Camera System Expansion . . . . .                  | 75        |
| 4.4      | Motion Compensation Analysis for<br>Heart Rate . . . . . | 77        |
| 4.5      | Remote Measurement of Oxygen Saturation . . . . .        | 78        |
| 4.6      | Next Generation Imaging Systems . . . . .                | 79        |
| <b>5</b> | <b>Conclusions</b>                                       | <b>81</b> |
| <b>6</b> | <b>Appendix</b>  | <b>85</b> |
| 6.1      | Dash Monitor Data Extraction . . . . .                   | 85        |
| 6.1.1    | Numerical Values . . . . .                               | 85        |
| 6.1.2    | ECG Waveform . . . . .                                   | 85        |
| 6.2      | USB Bandwidth Limitations . . . . .                      | 86        |
| 6.3      | RealSense Camera Hardware Triggering . . . . .           | 86        |

# List of Tables

|     |   |    |
|-----|---|----|
| 3.1 | Summary of Bland-Altman analysis for HR estimation in various ROIs. Values are in BPM. . . . .                        | 70 |
| 3.2 | Summary of Bland-Altman analysis for RR estimation in various ROIs. Values are in BrPM. . . . .                       | 70 |
| 5.1 | Comparison of the novel system using both intensity and depth to other intensity-based systems in literature. . . . . | 82 |
| 5.2 | Comparison of the novel system to depth-based systems in literature. . . . .  | 82 |

# List of Figures

|      |   |    |
|------|---|----|
| 1.1  | Conventional equipment for vital signs monitoring. [13], [14]. . .                                    | 3  |
| 1.2  | Example of heart rate variability [27]. . . . .   | 7  |
| 1.3  | Contact-based techniques for measuring respiratory rate by method [35]. . . . .                       | 8  |
| 1.4  | Skin reflectance model illustrating specular and diffuse reflection [45]. . . . .                     | 10 |
| 1.5  | Extinction spectra for Hb and HbO <sub>2</sub> [46]. . . . .  | 11 |
| 1.6  | Model of the skin as stacked layers [43]. . . . .   | 13 |
| 1.7  | Different depths of light penetration depending on wavelength [50].                                   | 14 |
| 1.8  | AC and DC components of a PPG signal [51]. . . . .  | 15 |
| 1.9  | Creation of a goodness metric for judging the PPG signal strength for areas of the face [55]. . . . . | 16 |
| 1.10 | Respiratory features illustrated in an example PPG waveform [57].                                     | 17 |
| 1.11 | Example of motion rejection [60]. . . . .   | 18 |
| 1.12 | BCG waveform [64]. . . . .  | 19 |
| 1.13 | Accelerometer signals taken from the head in three axes [65]. . .                                     | 20 |
| 1.14 | Illustrations of depth sensing methods [68]. . . . .  | 21 |
| 2.1  | The Intel RealSense D400 series [75]. . . . .   | 26 |
| 2.2  | Example of IR and depth channel images. . . . .   | 28 |
| 2.3  | Difference between depth and range [77]. . . . .  | 28 |
| 2.4  | Side-view schematic of a typical single-camera acquisition setup.                                     | 29 |



|      |  |    |
|------|--|----|
| 2.5  | Multi-camera configurations. . . . .   | 30 |
| 2.6  | GE Dash 3000 Patient Monitor. . . . .  | 31 |
| 2.7  | ECG adhesive electrode placement [78], [79]. . . . .   | 32 |
| 2.8  | Generating ROIs for chest and abdominal regions. . . . .   | 33 |
| 2.9  | Example of sub-ROIs of the face. . . . .   | 35 |
| 2.10 | Identifying motion artifacts in a depth signal. . . . .  | 38 |
| 2.11 | Frequency bands used for the calculation of SNR [83]. . . . .                                      | 39 |
| 3.1  | Flowchart for processing pipeline of remote vital signs monitoring system. . . . .                 | 43 |
| 3.2  | Examples of HR estimation from IR and depth data. . . . .  | 45 |
| 3.3  | Bland-Altman analysis of different ROIs for HR estimation. . . . .                                 | 47 |
| 3.4  | Examples of artifacts or obscurations within ROIs. . . . .   | 49 |
| 3.5  | Examples of RR estimation from IR and depth data. . . . .  | 50 |
| 3.6  | Bland-Altman analysis of RR agreement using different ROIs/channels. . . . .                       | 52 |
| 3.7  | Negative correlation of chest and face movement during respiration. . . . .                        | 53 |
| 3.8  | Raw depth waveforms for frontal body strip analysis. . . . .                                       | 54 |
| 3.9  | Evaluation of strip respiratory signal for both normal and distressed breathing scenarios. . . . . | 55 |
| 3.10 | Examples of simulating severe motion events in a dataset. . . . .                                  | 56 |
| 3.11 | Results of RR estimation on a motion-contaminated dataset before rejection. . . . .                | 58 |
| 3.12 | Motion rejection applied to a contaminated dataset for RR estimation. . . . .                      | 59 |
| 3.13 | Results of HR estimation on a motion-contaminated dataset before rejection. . . . .                | 60 |
| 3.14 | Motion rejection applied to a contaminated dataset for HR estimation. . . . .                      | 61 |
| 3.15 | Experimenting with skeleton tracking for single-camera motion compensation. . . . .                | 62 |

|      |  |    |
|------|--|----|
| 3.16 | Example where periodic gross body motion dominates the physiological signal in the chest depth, causing an overestimation of RR. . . . . | 63 |
| 3.17 | Example of motion compensation using two cameras. . . . .  | 64 |
| 3.18 | Bland-Altman analysis of participant data using two HR estimation techniques. . . . .  | 66 |
| 3.19 | Example of good estimation with a high resting HR. . . . .   | 67 |
| 3.20 | Motion rejection applied to participant dataset for HR estimation. . . . .   | 68 |
| 3.21 | Frequency spectrum of a dataset from subject 3. . . . .  | 69 |
| 4.1  | FOVs of the three cameras in the 360-degree camera configuration. . . . .  | 76 |
| 4.2  | 3-D surfaces used to create a 3-D model of a subject's torso. . . . .  | 77 |
| 4.3  | Diagram of a 4f system [92]. . . . .   | 79 |
| 4.4  | Double-helix PSF showing two rotating lobes that change position according to distance [93]. . . . .                                     | 80 |
| 6.1  | Voltage divider circuit to retrieve ECG voltage. . . . .   | 86 |

# Chapter 1

## Introduction

### 1.1 Motivation

Vital signs monitoring is a cornerstone of health care. Measurement of physiological parameters such as heart rate, respiratory rate, oxygen saturation, blood pressure, and body temperature allow quantification of a person's overall state of health and subsequent planning of treatment. In a clinical setting such as a hospital, vital signs monitoring is ubiquitous in the form of electronic monitors and physical sensors. Modern health care is highly dependent on the efficacy and reliability of these devices. In recent years, several factors have led to a shift of the health care industry to a greater emphasis on non-contact methods of vital signs acquisition, referred to in this thesis as “remote” methods.

With an aging global population, development of health care technologies is more important today than ever before. It is well established that the elderly population is highly susceptible to adverse health events. It is estimated by the United Nations that by 2050, the number of people in the world above the age of 65 is expected to nearly double compared to 2019 [1]. This trend, along with increasing complexity of hospital care and shortage of nursing staff [2], may potentially create a heavy burden on the health care system. Hence, the need for new health monitoring technologies and strategies is likely to increase dramati-

cally in the next few decades. Remote vital signs monitoring is expected to play a significant role in this new age as a promising technology for both clinical and at-home monitoring. This section discusses different motivating factors for its continued development, as well as potential use cases and applications.

### 1.1.1 Detecting Patient Deterioration

One of the benefits of remote monitoring of a continuous nature is the ability to mitigate “failure-to-rescue” mortality. Patients who undergo surgery are at a high risk of postoperative complications, with observed rates of 33-44% [3]. These complications, when managed poorly, result in significant morbidity and mortality rates [4]. As a method of detecting complications in hospital settings, vital signs are often used to form early warning scores (EWS), which serve as an indicator to detect deterioration of a patient. While EWS systems have been shown to be beneficial, they consist of manual observations by clinical staff and thus are subject to human interpretation and error [5]. Additionally, outside of the intensive care unit (ICU), their intermittent nature in the form of “spot checks” leads to inaccurate representations of a patient’s physiological state [6], [7]. These factors contribute to undetected patient deterioration that leads to otherwise preventable complications. Hence, there is a need for a continuous vital signs monitoring system that would ideally take the form of an ambient, remote sensor to minimize discomfort. Such ambient methods have been shown to ease the burden on patients by providing increased comfort, mobility, and unobtrusiveness [8]. Patient attitudes towards continuous remote monitoring have been generally positive as they benefit from increased safety and sense of reassurance [9]. Some have cited concerns with data privacy but were willing to share data if emerging health problems could be identified [10]. In this context, “continuous” refers to constant monitoring to avoid long gaps in time where the patient’s condition is effectively unknown. To facilitate this, vital signs information can be generated on a rolling basis using a specified sample duration. This ensures that adverse events are promptly detected.

Buekers et al. [11] demonstrated the usefulness of continuous monitoring by

providing chronic obstructive pulmonary disease (COPD) patients with wearable pulse oximeters for oxygen saturation measurements. These devices revealed significant fluctuations in oxygen saturation that were not apparent with the low-frequency spot checks that are normally performed in a hospital ward. Based on this discovery, the authors warned that data measured from spot checks should be interpreted with caution.

### 1.1.2 Disadvantages of Contact Methods

Common contact-based methods of vital signs monitoring include the electrocardiogram (ECG) and pulse oximeter. In a clinical setting, these typically take the form of self-adhesive electrodes and a finger clip, respectively. Images of this equipment are shown in Figure 1.1. In addition to patient discomfort, a major disadvantage of these sensors is the occurrence of poor or loose connections, often the result of improper skin preparation or lack of proper changing, which require nursing staff to repeatedly reapply them. Poor connections contribute to false alarms, which in turn contribute to alarm fatigue among nursing staff due to an overwhelming number of devices and sensors per patient. Alarm fatigue has been identified as a negative factor in patient safety and has even been shown to be responsible for cases of patient death [12]. A transition to remote patient monitoring would be a step towards alleviating the issue of alarm fatigue caused by contact-based sensors.

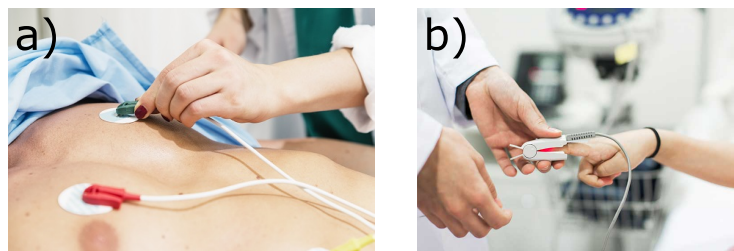


Figure 1.1: Conventional equipment for vital signs monitoring. a) ECG self-adhesive electrodes. b) Pulse oximeter. Taken from [13], [14].

Aside from alarm fatigue, self-adhesive electrodes from bedside monitors

have been shown to cause skin irritation in a small number of patients [15]. Some case studies have shown patients with allergic reactions to material in the electrodes, a condition called contact dermatitis [16]. Furthermore, cables and wires leading from the monitors can cause mobility issues for patients, particularly when they leave their bed to walk around or use the restroom. Skin contact must also be prevented in certain groups of patients, such as neonates or patients with skin damage.

### 1.1.3 Telemonitoring

Along with the development of monitoring technologies in clinical settings, various factors have led to interest in telemonitoring, particularly for the elderly population. Using wearable or remote sensors, adults with chronic health conditions can be safely monitored at home without direct hospital surveillance. Miniaturization of sensors and ever-increasing rates of smartphone usage facilitate the development of both wearable and ambient health monitoring systems that continuously acquire data from the patient. This results in an improved quality of life and reduced health care costs by creating a proactive approach to treatment via early detection and intervention [17]. Detecting deterioration early can reduce or prevent hospitalization, saving time and money. Telemonitoring is commonly used for conditions such as chronic heart failure, chronic obstructive pulmonary disease, and diabetes mellitus [18]. As the patient does not need to be hospitalized, their mental wellness also sees improvement. A portable, low-cost system for health monitoring is ideal for these cases.

Telemonitoring also has uses in isolated and low-income communities where access to hospital care is difficult, expensive, or unavailable. Patients in remote areas can easily communicate with health care professionals via online consultations and share data proactively, potentially reducing the occurrence of medical emergencies. It has been shown that health status is inversely proportional to remoteness [19], which emphasizes the importance of telemedicine and telemonitoring for such communities. Integration of telemonitoring into remote communities should be analyzed on a case-by-case basis, due to unique

logistical and social characteristics of each location. If enough data can be gathered, it may facilitate large-scale epidemiological surveillance for entire regions, providing a strong boost to public health and safety [20].

#### **1.1.4 Public Health Applications**

The usefulness of remote health monitoring is evident in times of public health crisis, most recently seen in the COVID-19 pandemic, as the nature of a contagious disease requires proper physical distancing and protection to avoid transmission. Symptoms of COVID-19 include fever, elevated heart rate, and shortness of breath [21], [22], all of which are detectable via remote monitoring. Once symptoms and characteristics of a disease are identified, they can be programmed into a system database; a subject's vital signs can then be matched with a disease or condition in the database for screening purposes. Such a system, implemented in the form of a kiosk or ambient sensor, can be used to screen patients who enter a hospital for disease symptoms. Other possible locations include airports, factories, construction sites, sports events, and more.

### **1.2 Overview of Vital Signs**

#### **1.2.1 Heart Rate**

Heart rate (HR) is one of the easiest vital signs to measure yet contains an abundance of information about a person's overall health. A normal resting HR lies between 60 to 100 beats per minute (BPM) in adults [23]. A high resting HR is associated with cardiovascular morbidity and mortality, as well as the occurrence of sudden adverse events such as myocardial infarction and long-term heart failure. An increase in HR by 10 BPM has been shown to be associated with a 20% increase in risk of cardiac death [24]. This risk is significant in both young and elderly persons. Elevated resting HR results in higher mechanical stress on blood vessel walls and is shown to be correlated with increased arterial stiffness, one of the risk factors for the development of chronic

cardiovascular complications such as atherosclerosis [25]. Lifestyle factors also associate it with hypertension, obesity, and lower amounts of physical activity. HR is typically measured using ECG or a pulse oximeter, both contact-based methods. The former measures electrical impulses and the latter is an optical method for measuring blood volume changes due to pulse.

Due to the origin of heartbeat as an electrical pulse from the sinoatrial node, monitoring a subject's HR over time can provide insight into the autonomic nervous system (ANS). Heart rate variability (HRV) is the natural variation in HR and cardiac output as regulated by neural and hormonal factors [26]. HRV is typically evaluated in the time-domain as the time between successive R waves of the ECG reading. Increased HRV parameters are seen in healthy subjects, as this reflects the ability of the heart and the nervous system to dynamically adjust cardiac output based on the need of the body [27]. Figure 1.2 shows an example of how HRV changes between healthy and diseased persons. HRV analysis is useful to assess mortality risk in post-infarct patients, as well as a variety of conditions that involve cardiac autonomic dysfunction such as diabetes, stroke, and multiple sclerosis [27].

Furthermore, HRV has been used to assess an individual's stress level. Psychological stress is correlated with physiological changes, namely a significant decrease in the amplitude of HRV [28]. The vagus nerve is responsible for regulating parasympathetic control of the heart and lungs. Under stressful environments, the activity of the vagus nerve decreases, causing parasympathetic deactivation and a subsequent decrease in HRV. Vagus nerve activity is difficult to measure directly, so HRV is used as a reliable indicator of stress. This provides an objective evaluation of an individual's stress level that cannot be obtained with subjective psychometric assessments [29].

### 1.2.2 Respiratory Rate

Respiratory rate (RR), measured in breaths per minute (BrPM), is a vital sign that provides insight into neural and pulmonary functions. It is often cited to be a physiological parameter that is accurately indicative of early deterioration of a



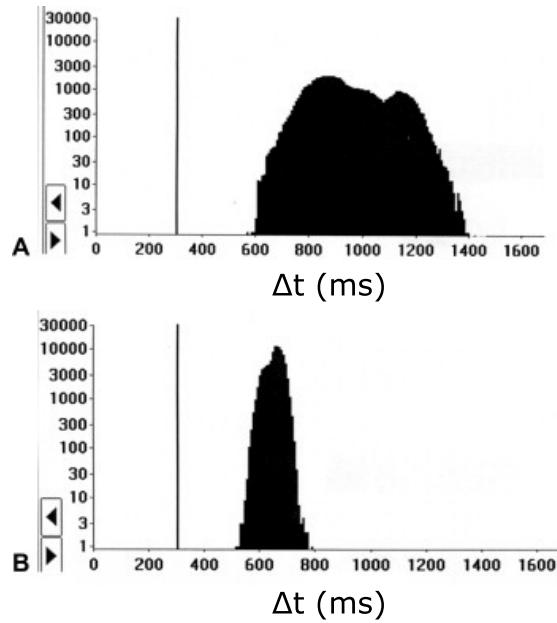


Figure 1.2: Example of heart rate variability in a) healthy subject and b) patient with diabetes and autonomic neuropathy. Horizontal axis represents time in milliseconds between heartbeats. Modified from [27].

patient, as well as a predictor of adverse events and ICU admission [30]–[32]. It has been shown that changes in RR can precede other physiological changes such as oxygen saturation [33]. Normal resting RR lies between 6 and 20 BrPM; lower or higher rates are associated with unwellness and increased patient mortality [34]. Techniques to monitor RR include airflow measurements, acoustic sensing, air temperature/humidity measurements, and monitoring chest wall movement. RR can also be obtained through certain cardiovascular measurements. These techniques are summarized in Figure 1.3.

Respiration in humans can be classified as one of two types: chest/shallow breathing and abdominal/diaphragmatic breathing. Shallow breathing is created by movement of the intercostal muscles rather than the diaphragm and results in an inward-outward movement of the chest. Rapid, shallow breathing is referred to as tachypnea and is associated with respiratory distress [36]. Conversely, diaphragmatic breathing is a deeper form of breathing that involves flattening of the diaphragm during inhalation, which pushes the abdominal wall out. It is shown to reduce stress and improve mental function [37].





| Contact-based Techniques |   |                          |                                       |   |                                 |   |
|--------------------------|---|--------------------------|---------------------------------------|---|---------------------------------|---|
| Respiratory airflow      | Respiratory sounds  | Air Temperature          | Air Humidity                          | Air components  | Chest wall movements            | Modulation cardiac activity   |
| Flow measurements        | Acoustic measurements   | Temperature measurements | Relative humidity measurements        | CO <sub>2</sub> measurements  | Strain measurements             | Biopotential measurements   |
| Differential flowmeters  | Microphones   | Thermistors              | Capacitive sensors                    | Infrared sensors  | Resistive sensors               | ECG sensors   |
| Turbine flowmeters       |  | Thermocouples            | Resistive sensors                     | Fiber optic sensors   | Capacitive sensors              | Light intensity measurements  |
| Hot wire anemometers     |   | Pyroelectric sensors     | Nanocrystal and nanoparticles sensors |  | Inductive sensors               | PPG sensors   |
| Fiber optic sensors      |   | Fiber optic sensors      | Fiber optic sensors                   |   | Fiber optic sensors             | Fiber optic sensors   |
|                          |   |                          |                                       |   | Impedance measurements          |  |
|                          |   |                          |                                       |   | Transthoracic impedance sensors |   |
|                          |   |                          |                                       |   | Movement measurements           |   |
|                          |   |                          |                                       |   | Accelerometers                  |  |
|                          |   |                          |                                       |   | Gyroscopes                      |   |
|                          |   |                          |                                       |   | Magnetometers                   |   |

Figure 1.3: Contact-based techniques for measuring respiratory rate by method. Taken from [35].

Despite its proven usefulness, RR is often measured or interpreted inaccurately in clinical settings, leading some to call it a “neglected” vital sign [34]. Clinical staff have low confidence that RR recordings are reliable, believing that they are often estimated due to perceived lack of time for proper measurement [38]. This attitude has led to (and is further caused by) a belief that collecting accurate RR as part of clinical rounds is not essential, despite organization requirements dictating it as a necessity [31]. Hence RR is often unrecorded, or in some cases, recorded incorrectly, which can lead to misleading clinical care. In cases where it is recorded, manual counting is the most frequent method of measurement, despite capnography being the gold standard method [32]. This culture of neglect prevents clinicians from taking full advantage of a highly useful physiological parameter that is ultimately easy to measure and does not require complex technology. The integration of continuous monitoring would be greatly beneficial to alleviate this problem, as it would require little to no attention from nursing staff to accurately capture respiratory data from the patient.

### 1.2.3 Oxygen Saturation

Oxygen saturation in blood is dependent on hemoglobin and is defined as the ratio of oxygenated hemoglobin (oxyhemoglobin,  $HbO_2$ ) concentration to total hemoglobin concentration, which refers to the sum of  $HbO_2$  concentration and deoxyhemoglobin (Hb) concentration. Oxygen saturation is represented as a percentage with a normal range of 94-100% in healthy adults; values lower than 93% are generally considered hypoxemic [39]. While invasive blood gas analysis remains the gold standard for arterial oxygenation measurement, the development of pulse oximetry in the 1980s provided a non-invasive indirect method that is widely used in clinical settings [40]. Oxygen saturation refers to arterial saturation, but is most commonly measured as peripheral oxygen saturation ( $SpO_2$ ), which is most easily obtained using pulse oximeters. Peripheral in this case refers to measurements at locations such as the fingers or ears. Other variations are arterial ( $SaO_2$ ), venous ( $SvO_2$ ), and tissue ( $StO_2$ ) saturation, which are typically measured invasively.  $SpO_2$  is calculated as shown in Equation 1.1.

$$SpO_2 = \frac{[HbO_2]}{[Hb] + [HbO_2]} \quad (1.1)$$

Unlike HR and RR,  $SpO_2$  is difficult to measure without equipment. The most reliable physical indicator is cyanosis of the hands and feet, which typically begins to occur at around 75% saturation [41]. Since the brain begins to develop visual and cognitive deficits at around 80-85% saturation [42], this is not a reliable indicator. Hence in a clinical setting, measurement of  $SpO_2$  necessitates the application of a pulse oximeter.

## 1.3 Vital Sign Measurement Techniques

Acquiring vital signs through remote methods requires extensive utilization of optical techniques and understanding of human physiology. This section will focus on camera-based approaches to extract physiological information.

### 1.3.1 Light Interaction with Tissue

Photoplethysmography (PPG) is a low-cost optical technique that is used to detect blood volume variations in tissue, which can be used to retrieve vital signs such as heart rate, respiration rate, and oxygen saturation. PPG is believed to measure variations in reflected or transmitted light that are affected by the presence of chromophores in blood; these variations are collectively known as the blood volume pulse (BVP). While this theory for the origin of PPG (known as the volumetric model) is occasionally disputed, it is sufficient to support the use of PPG in clinical applications and therefore is generally accepted [43]. PPG performed without contact is referred to as remote PPG (rPPG).

In order to explain the origin of the PPG signal, a model must first be devised for skin reflectance. Figure 1.4 illustrates a light source illuminating skin tissue and a camera capturing reflected light. It is assumed that the light source has a fixed spectral composition. When incident light reaches the skin, approximately 4-7% is reflected off the skin surface as specular reflection regardless of wavelength and skin tone [44]. This specular reflection contains the same spectral composition as the incident light, but contains no useful physiological information.

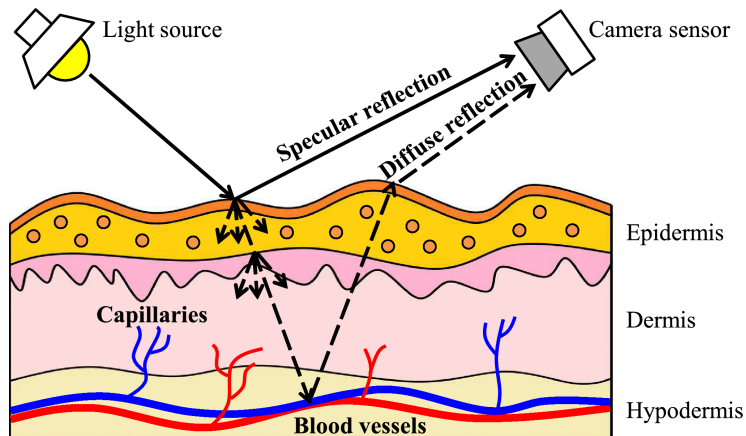


Figure 1.4: Skin reflectance model illustrating specular and diffuse reflection. Taken from [45].

The remaining light that penetrates the skin undergoes two processes: absorption and scattering. The substances in blood that are primarily responsible for absorption are Hb and HbO<sub>2</sub> in the dermis. The wavelength-dependent extinction spectra for both chromophores are shown in Figure 1.5. The extinction coefficient, also known as the attenuation coefficient, is the sum of the absorption coefficient ( $\mu_a$ ) and the scattering coefficient ( $\mu_s$ ). In the visible spectrum (450-700 nm), Hb absorption peaks at approximately 555 nm, whereas HbO<sub>2</sub> peaks at approximately 540 and 575 nm. Higher absorption results in less light transmitted. Note that water is not a significant absorber of light at these wavelengths.

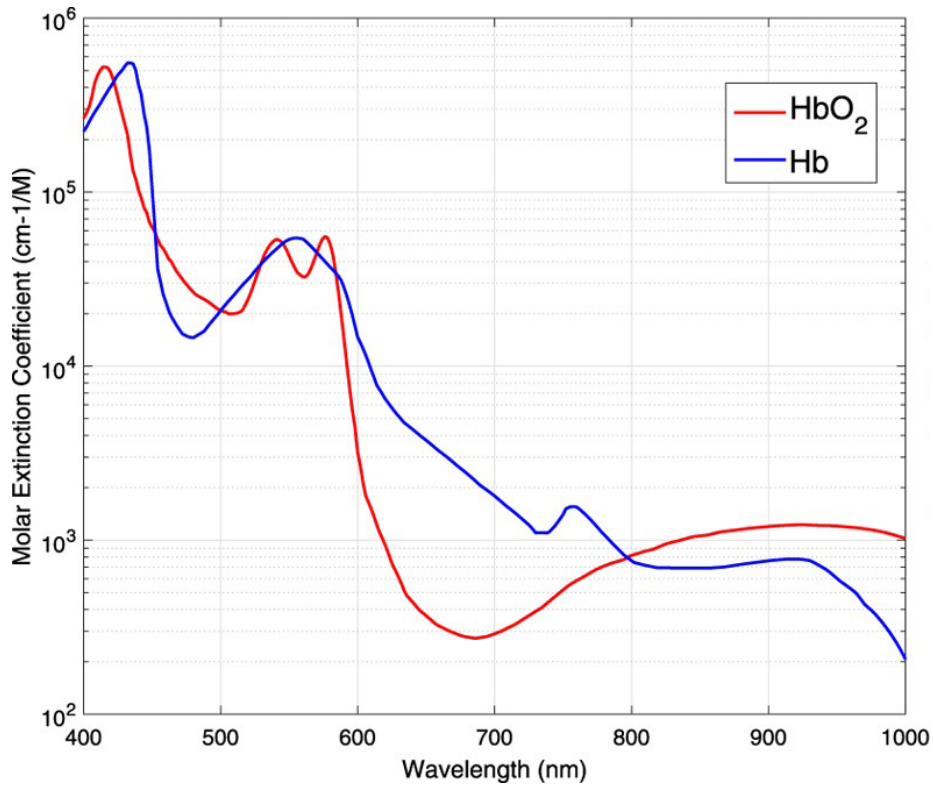


Figure 1.5: Extinction spectra for Hb and HbO<sub>2</sub>. Taken from [46].

Light that is not reflected off the skin surface or absorbed undergoes the process of scattering, which is defined as a change in direction, polarization, or phase of light within a medium [47]. No energy is lost during a scattering

event. Scattering is particularly significant in the ultraviolet (UV), visible, and near-infrared (NIR) spectra due to photon interactions with cellular structures, though less occurs in the NIR spectrum. A photon's path within tissue can be modelled using the mean free path (MFP) approach, which is defined as the average distance that a photon will travel between two scattering events. The MFP is calculated as follows:

$$\text{MFP} = \frac{1}{(\mu_a + \mu_s)} \quad (1.2)$$

In most tissues, it is assumed that  $\mu_s \gg \mu_a$ , hence Equation 1.2 can be simplified to:

$$\text{MFP} \approx \frac{1}{\mu_s} \quad (1.3)$$

The MFP is approximately 100  $\mu\text{m}$  in tissue, though this value will vary slightly depending on the type of tissue. When a photon travels through tissue and undergoes multiple scattering events, the angle at which the photons are scattered must be taken into account. This new parameter, called the transport mean free path (TMFP), is defined as follows:

$$\text{TMFP} = \frac{1}{\mu_s(1 - g)} \quad (1.4)$$

where  $g$  is a function that defines the probability of forward scattering as a Henyey-Greenstein phase function [48].  $g$  is typically 0.8 to 1 for tissue scattering, meaning that the scattering is mostly in the forward direction. A higher  $g$  means deeper penetration distances and a longer travel time before photon diffusion. By the nature of randomness, a percentage of photons will penetrate into deeper skin layers and be reflected, emerging from the skin as diffuse reflection (shown in Figure 1.4). It is this reflected light that contains useful physiological information.

For diffuse reflected light to contain pulsatile information, light must pen-

etrate deeply enough to reach arterioles in the dermis in order to be affected by absorption from Hb and HbO<sub>2</sub>. Skin models typically separate the skin and underlying tissues into three layers as illustrated in Figure 1.6: the epidermis, dermis, and hypodermis. The epidermis consists mostly of dead skin cells and is approximately 800 μm thick on the finger pad [43]. The dermis contains capillary loops and arterioles. The hypodermis contains subcutaneous fat, connective tissue, and arterioles and arteries.

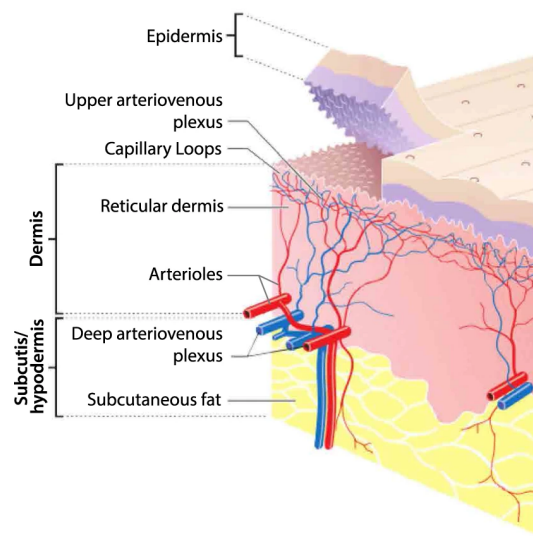


Figure 1.6: Model of the skin as stacked layers. Taken from [43].

Due to scattering and absorption coefficients being highly dependent on light wavelength, different wavelengths of light will penetrate to different depths, as shown in Figure 1.7. Moço et al. [43] showed that green wavelengths are sufficient to retrieve dermal BVPs, and red-NIR wavelengths can reach as far as subcutaneous arterioles. This is in agreement with the absorption spectra in Figure 1.5; lower absorption corresponds to deeper penetration. Martinez et al. [49] observed that optimal wavelengths for extracting HR were 480-610 nm (visible) and 800-925 nm (NIR), whereas optimal wavelengths for RR were 450-490 nm (visible) and 600-980 nm (NIR). Thus, changes in intensity of reflected light reveal pulsatile information caused by absorption and scattering throughout different skin layers.

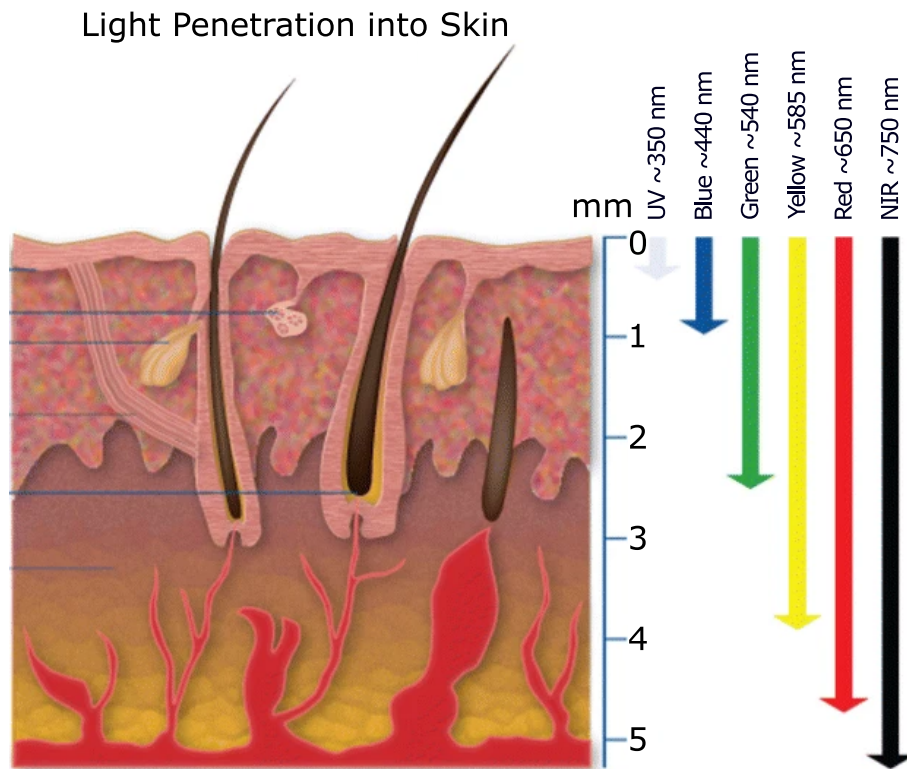


Figure 1.7: Different depths of light penetration depending on wavelength. Modified from [50].

A typical PPG signal can be divided into two components as shown in Figure 1.8: a DC component determined by the medium in which the light passes through (tissue, bone, non-pulsatile components of blood, etc.), and an AC component created by variations in blood volume [51]. The AC component therefore contains pulsatile information that can be used to measure HR. Amplitude of the PPG signal is typically calculated as the ratio of the AC component to the DC component.

### 1.3.2 Retrieving Vital Signs from Camera Images

Various camera channels have been used to retrieve BVP from intensity of reflected light, most commonly RGB (visible) and NIR channels. Some systems utilize a dedicated light source, though ambient light is often sufficient. One of



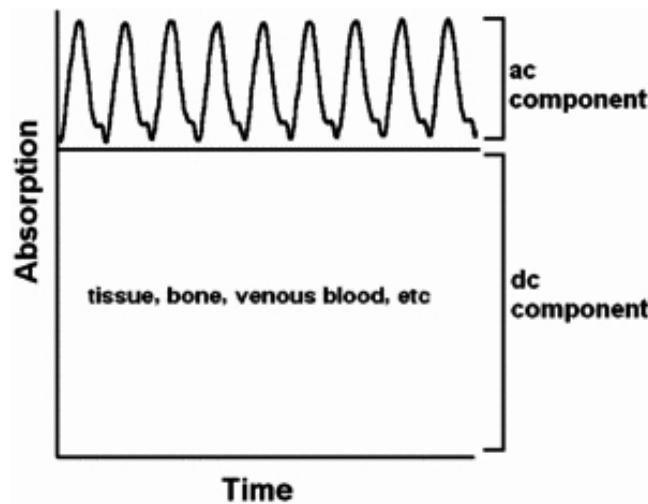


Figure 1.8: AC and DC components of a PPG signal. Taken from [51].

the earliest published systems for video-based non-contact pulse measurements came from Poh et al. [52], [53] in 2010. Their system utilized a built-in laptop RGB webcam with resulting red, green, and blue channels and an intensity-varying signal from each. These three signals, spatially averaged from pixels on the subject’s face, were decomposed into independent source components using blind source separation. Through independent component analysis (ICA), a PPG signal with a strong frequency peak was observed in one of the components, leading to the extraction of HR via frequency analysis. The same group later expanded on this approach by using a five-band camera with red, green, blue, cyan, and orange channels, finding the strongest PPG signal with a combination of cyan, green and orange [54].

In 2015, Kumar et al. analyzed the rPPG signal from different areas of the face, combining signal from various regions to create a weighted average [55]. By dividing the face into a grid of small regions and using frequency analysis, they developed a goodness metric to show that the strongest rPPG signal occurs in the forehead and cheek regions of the face. This approach is shown in Figure 1.9. The final weighted average signal showed close correlation with the ground truth PPG waveform.

While PPG is primarily used for isolating the BVP waveform to gain insight

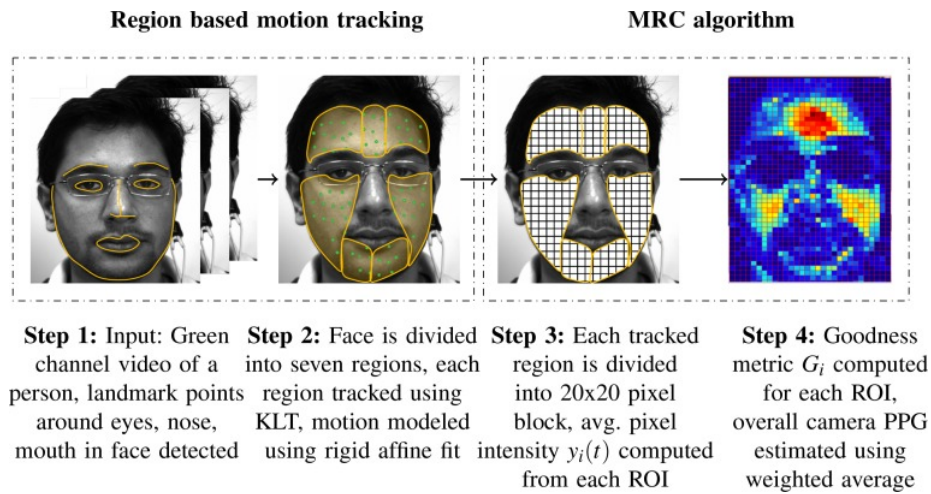


Figure 1.9: Creation of a goodness metric for judging the PPG signal strength for areas of the face. Strongest signal can be observed in the forehead and cheeks regions. Taken from [55].

into pulsatile activity, features linked to respiration have been identified in PPG data [56]. These features, illustrated on an example PPG waveform in Figure 1.10, describe the relationship between respiration and cardiovascular activity as follows:

- Respiratory induced frequency variation (RIFV): Periodic changes in HR are modulated by the ANS, causing HR to synchronize with the respiratory cycle. This synchronization is known as respiratory sinus arrhythmia.
- Respiratory induced intensity variation (RIIV): Changes in the baseline of the PPG signal are caused by intrathoracic pressure variations relating to respiration, creating an intensity-varying signal due to changes in perfusion.
- Respiratory induced amplitude variation (RIAV): Pulse amplitude changes are caused by decreasing cardiac output during inspiration due to reduced ventricular filling.

Thus, respiration presents itself in PPG data, making it possible to extract RR from PPG waveforms.

A common drawback of video-based PPG is high susceptibility to motion

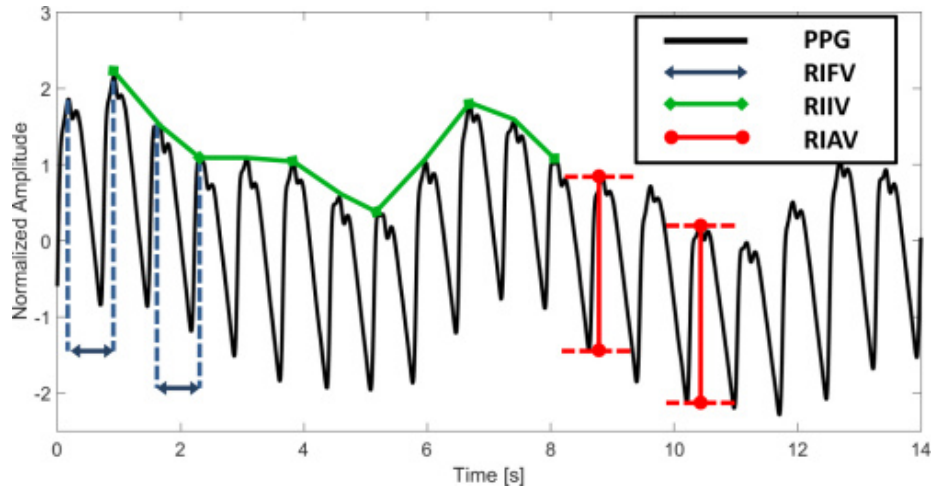


Figure 1.10: Respiratory features illustrated in an example PPG waveform. Taken from [57].

artifacts; a slight shift in subject position can introduce significant noise that overpowers the physiological signal. This is a significant challenge with rPPG due to the relatively low amplitude of the BVP signal. Publications by de Haan et al. [58], [59] improved on motion robustness in RGB video by using a linear combination of red, green, and blue channels, developing so-called “chrominance” methods. Li et al. [60] demonstrated an early method of motion tolerance by taking the image intensity signal of the face, dividing it into smaller segments, and calculating the standard deviation (SD) of each segment as shown in Figure 1.11. Higher SD values indicate a corrupted segment, and the respective segment is removed from the overall signal. The signal is then re-concatenated for frequency analysis. While the authors did not define an exact term for this process, we have termed this kind of motion tolerance as “motion rejection”.

The main challenges of rPPG monitoring can be summarized in four categories: motion artifact susceptibility, ambient illumination variations, image quality, and spectral band selection [23]. Subject motion introduces strong artifacts against the relatively sensitive measurements involved in rPPG monitoring, necessitating compensation in practical scenarios where subjects cannot remain entirely still. Illumination variations are a concern especially in out-

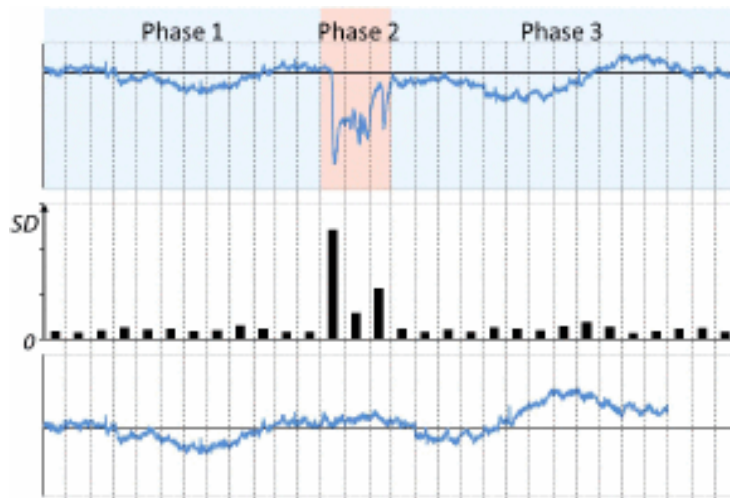


Figure 1.11: Example of motion rejection. Corrupted signal segments are detected by high SD values and removed from the signal. The remaining segments are then re-concatenated to form a continuous signal. Taken from [60].

door environments, but also in common indoor scenarios due to factors such as changes in screen brightness. Changing illumination skews extracted intensity signals, potentially causing under or over-illumination. Image quality is affected by several factors including sensor properties, frame rate, and pixel resolution; the latter two must often be balanced due to bandwidth constraints. Finally, as explained in the previous section, the choice of spectral band (visible, NIR, etc.) affects the depth at which backscattered light is influenced by blood volume variations. As a result, the intensity of diffuse reflections will vary based on camera wavelengths, which lead to certain wavelength bands being superior at extracting vital signs, as shown by [49].

### 1.3.3 Head Ballistocardiography

Pulsatile activity is associated with periodic head oscillations due to the ejection of blood from the heart through the carotid arteries during systole. Measurement of these subtle body movements is known as ballistocardiography (BCG), which has recently gained a resurgence of interest due to improved techniques for motion measurement and signal processing [61]. The use of BCG to measure cardiac activity typically refers to measurements of gross body motion in the

longitudinal direction; a subject would typically stand or lie on a low-friction platform, and displacement of the platform would be tracked over time. Using the head is of interest due to its relatively non-rigid movement relative to the rest of the body, and head motion associated with pulsatile activity has been recorded as artifacts in magnetic resonance imaging [62], showing that such a signal exists and can be recorded. However, there are several types of involuntary movement that complicate the use of head BCG, including oscillating motion that keeps the head in positional equilibrium and head motion associated with respiration [63]. Figure 1.12 shows a measured BCG waveform for one heartbeat, with the three most prominent waves referred to as the “IJK complex” [64].

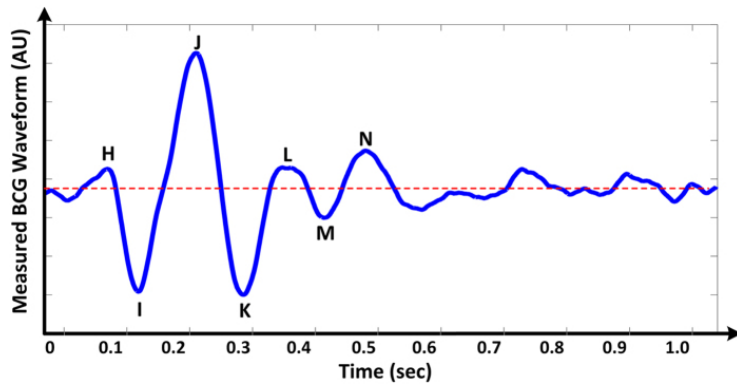


Figure 1.12: A standard BCG waveform measured in the longitudinal direction. Major waves are labelled, with the IJK complex being most prominent. Taken from [64].

Da He et al. [65] demonstrated head BCG measurements using an ear-anchored accelerometer to provide tri-axis acceleration values. Their results showed the most significant head motion in the y-axis (upward-downward direction) when the subject is standing at rest, as shown in Figure 1.13. They were also able to measure the time between the heart’s left ventricular depolarization (the R wave on ECG) and the strongest ballistic force on the body (the J wave on BCG), which is referred to as the RJ interval and gives insight into the heart’s contractility.

Balakrishnan et al. [63] demonstrated one of the first instances of detecting

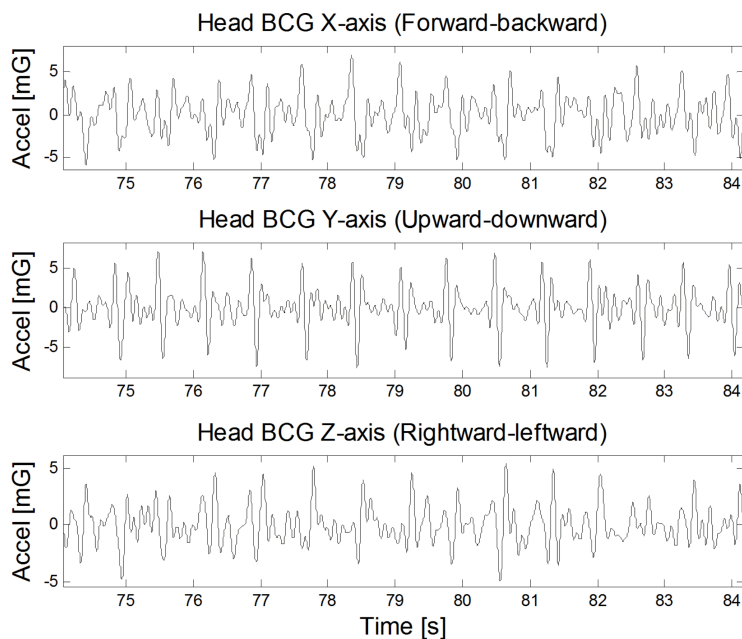


Figure 1.13: Accelerometer signals taken from the head in three axes. The BCG signal appears strongest in the y-axis (upward-downward direction), but can still be seen in the x and z-axes. Taken from [65].

HR using video-based BCG. Their approach identified features on the subject’s face and tracked their movement in the longitudinal direction over time. Principal component analysis was applied on the resulting signals to generate a set of source signals, at least one of which contained a strong frequency peak representative of the HR. Interestingly, this approach does not require direct view of the subject’s face, being able to pick up HR from the back of the head as well as when the face is obscured by a mask. Video-based BCG was also demonstrated by Shao et al. [66], who used the mouth area to identify features to track. Their results agreed strongly with rPPG-derived HR values. Al-Naji et al. [67] used a similar approach, but were able to pick up vital signs at long distances (up to 60 m) by using magnification techniques.

### 1.3.4 Depth Sensing

Depth sensing in this context involves retrieving information about the distance between the camera and the environment in the camera’s frame, often referred to

as the z-axis. There are three major types of approaches towards depth sensing: stereo triangulation, time-of-flight, and structured light imaging. Illustrations of these three methods are shown in Figure 1.14.

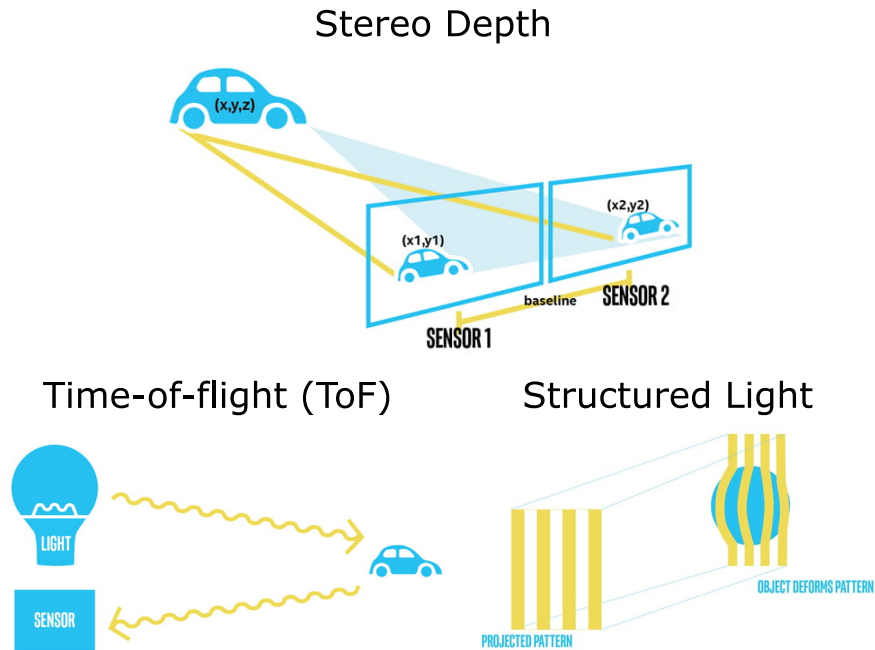


Figure 1.14: Illustrations of depth sensing methods. Images taken from [68].

Stereo triangulation is one of the more popular approaches towards 3-D depth sensing as it is low-complexity and low-cost. It involves using two slightly offset camera sensors to capture images of an environment from two perspectives, then calculating the disparity between the two images by matching points from one image to another. Since no energy is emitted from the cameras, it is considered a passive method. In order to derive physical dimensions from disparity, the intrinsic and extrinsic parameters of the camera system must be known, including the distance and angle between the two sensors. As the horizontal distance between the two imagers (called the baseline) increases, depth resolution increases along with the minimum and maximum ranges of the system. A drawback of stereo triangulation is its requirement to match points from one image to another; surfaces in the environment that lack significant texture, such as flat walls, can introduce “holes” in the depth image where the disparity

algorithm cannot match points.

Time-of-flight (ToF) imaging utilizes the speed of light to determine depth. A laser pulse is emitted from the camera and a precise timer measures the amount of time it takes for the pulse to return after being reflected off the environment. LIDAR (an acronym of "light detection and ranging") is an example of ToF imaging and is commonly used in vehicles and robotics. ToF imaging has depth resolution in the order of millimeters and has exceptional range, but requires specialized equipment and suffers from crosstalk, in which the detector cannot distinguish between different pulses arriving at the same time [69]. This also means that the performance of ToF cameras occasionally suffers in outdoor conditions.

Structured light cameras involve projecting a known pattern of typically infrared light onto an environment and detecting how the pattern appears to the camera. The pattern can vary spatially or temporally. Objects in the environment will naturally warp the shape of the light pattern; the severity of the distortion will vary based on its distance to the camera. This type of imaging shares the same drawback of ToF imaging, namely that it is susceptible to interference from other light sources. Furthermore, structured light imaging is relatively low range, as it relies on accurately visualizing the distortion of the projected pattern.

Several methods of vital signs detection with depth sensing utilize the Microsoft Kinect, a low-cost depth camera originally designed for entertainment. The original Kinect camera used structured light imaging by projecting a NIR pattern, whereas newer versions use ToF imaging for increased frame rate [70]. Yu et al. [71] utilized the Kinect to measure respiratory volume remotely through movement of the chest wall, but encountered some difficulties with rigid body movement, being unable to distinguish it from respiratory motion. Centonze et al. [72] applied the same approach during sleep, finding sufficient correlation with results from a polysomnograph. Yang et al. [73] utilized the Kinect to estimate HR via 3-D motion tracking of the head, using several denoising procedures to overcome the inherent noisiness of the depth images.



Using depth information to monitor vital signs is of interest because numerous physiological processes are associated with some form of physical movement. Intensity-based techniques are also susceptible to artifacts from subject motion, which can be directly measured with depth information. Furthermore, most depth sensing methods are illumination-invariant, which may allow for vital signs monitoring in low-light environments or environments with constantly changing illumination. Ideally, depth information could be integrated with intensity-based techniques such as rPPG. Our system explores this, as described in the next chapter.

## 1.4 Thesis Objectives

Remote monitoring of vital signs is a technology with several motivating factors behind its continued development, including potential applications in the clinic, at home, and in a public environment. Vital sign measurement techniques exploit the physiological processes behind these vital signs, such as changing intensity of reflected light due to blood volume variations and body movement due to respiration. Both intensity and depth-based techniques have been described for extracting information for HR and RR estimation.

To the author’s knowledge, a system that correlates both intensity and 3-D depth information to increase reliability and motion tolerance has not yet been explored in existing literature. Each of these two channels has limitations when used independently. Intensity-based approaches are highly susceptible to artifacts caused by gross subject motion and illumination variations. Conversely, depth-based approaches have inherent depth noise that limits their accuracy, and motion-based measurements such as head BCG are sufficiently small in amplitude that they can be overshadowed by noise. Combining these two channels provides two independent methods of vital signs estimation, which increases redundancy in the case that one method fails.

Using these channels, we have developed a system that makes use of both intensity and depth data to estimate HR and RR from camera recordings. The

system was designed with indoor ambient illumination as a light source, based on the potential use cases in a clinic or home where an additional light source may not be feasible. Data collected using the system consisted of one subject in the camera frame, though analysis of multiple subjects is theoretically possible with more intensive processing. The system requires no calibration for different subjects. Hence, vital sign estimates can be feasibly acquired within a 10-second interval for HR and a 30-second interval for RR, though longer durations improve accuracy. In order for the system to remain minimally intrusive, the system must be able to extract vital signs from a distance of at least 1.5 m away from the subject. In a clinical setting, this would allow the camera(s) to be placed at the foot of a bed or mounted on the ceiling. Finally, the system must be tolerant to gross subject motion. While this is not as much of an issue in clinical settings where the patient is mostly at rest, there are still events that can cause a motion artifact, such as the subject shifting positions. The objective is therefore to apply techniques to extract vital signs during motion when possible, and if no vital signs can be retrieved, the system will identify and discard motion-corrupted events.

This thesis will cover the following objectives:

1. Use intensity and depth information to independently extract vital signs from each channel,
2. Identify motion artifacts and quantify their severity,
3. Apply motion tolerance to extract vital signs even during motion, and
4. Validate accuracy of novel system by assessing agreement with a reference.

System accuracy must be evaluated to ensure that HR and RR estimations are correct. As with any measurement device, there is a level of error that is present in the returned values. For example, the Masimo MightySat Rx pulse oximeter reports an accuracy range of 3 BPM for HR (5 BPM during motion) and 3 BrPM for RR [74]. For the purpose of validating the novel remote monitoring system, the author chose to use these values as a target specification, as the Masimo oximeter is used to monitor patients and thus would be a valid

system for comparison. The device has also received 510(K) clearance by the Food and Drug Administration (FDA) in the United States. It is important to note that specifications for accuracy may differ between environments, with clinical settings requiring more precise measurements than home settings. Patients in clinical settings are generally more susceptible to complications and adverse events, necessitating greater accuracy of measurements for clinical evaluation. Conversely, subjects at home who do not require hospitalization may be more active and exhibit greater subject motion. In these cases, it may be necessary to trade measurement sensitivity for greater tolerance to motion. Furthermore, while the environment in a hospital room remains relatively static, a home environment may consist of varying backgrounds and illumination conditions. It is important that the system is robust to these changing scenarios.

In the following chapter, the architecture of this novel system is discussed, including methods of extracting physiological data from camera recordings and techniques for improved motion tolerance.

## Chapter 2

# Experimental Methods

## 2.1 Equipment

### 2.1.1 Intel RealSense Cameras

The Intel RealSense D400 Series Depth Cameras are stereo depth sensing cameras equipped with NIR and RGB channels. There are four camera models in the D400 series, shown in Figure 2.1: D415, D435, D435i, and D455; the latter two are bundled with an inertial measurement unit (IMU). The D435 and D435i are identical aside from the latter's inclusion of the IMU.



Figure 2.1: The Intel RealSense D400 series. Images taken from [75].

The D400 cameras are lightweight and portable, making them ideal for a remote monitoring setup. Depth sensing is achieved through active IR stereo provided by two identical camera sensors, each of which is sensitive to visible and NIR spectra (400-865 nm). Each module also has a dedicated color sensor for

the RGB channel. The camera modules are bundled with an 850 nm structured light projector, which provides illumination and increased depth accuracy in low-texture environments. The cameras have a pixel resolution of up to 1280x720 and a frame rate of up to 90 frames per second (fps). The D415 uses a rolling shutter, whereas the D435 and D455 use global shutters; for this reason, the D415 requires a longer exposure time to capture images and is more sensitive to fast-moving motion artifacts [76]. The D435 and D455 also have a wider field-of-view (FOV) of 87° horizontal and 58° vertical compared to 65° horizontal and 40° vertical for the D415. D435i and D455 cameras were used for the majority of data collection due to the aforementioned global shutter allowing faster exposure times and higher fps.

The two channels utilized from the RealSense cameras are 8-bit IR and 16-bit depth. Example images from both channels are shown in Figure 2.2. Each of the two IR imagers on the camera captures 8-bit IR images (16-bit IR is available but is unrectified, and Intel does not recommend its usage outside of calibration environments). Using these images and the known extrinsic parameters (i.e. relative position and rotation) of the imagers, depth is calculated by estimating the disparity between matching keypoints on the left and right images. Lower disparity corresponds to a farther object. The resulting output is a 16-bit depth frame that contains depth information for each pixel. The bitwise depth unit is configured to be 0.0001 m, resulting in a theoretical maximum range of approximately 6.5535 m. The infrared structured light projector serves to project texture onto an environment, which improves keypoint matching between left and right images in areas with low texture. Note that depth is measured from the parallel plane of the imagers and not the absolute range between the imager and the object, as explained in Figure 2.3. Hence, the RealSense cameras use a combination of stereo triangulation and structured light approaches to depth sensing.

The nature of depth imaging means that there is a margin of error associated with retrieved depth measurements. This depth error can be reduced with proper exposure settings and pixel resolution. The theoretical limit for root-mean-squared (RMS) depth error is:

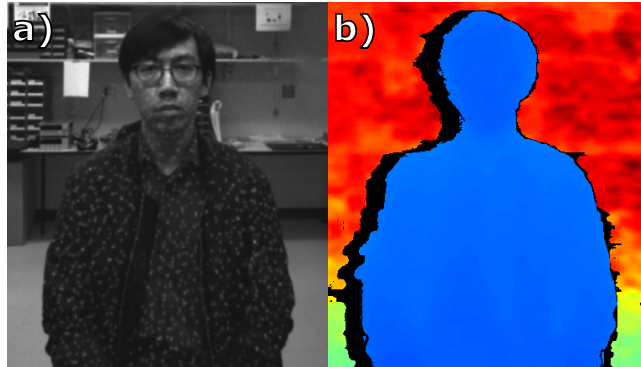


Figure 2.2: Example of images taken from a) 8-bit IR channel and b) 16-bit depth channel. The textured pattern from the structured light projector can be seen on the IR image. A colormap has been applied to the depth image for better visualization. Shadowing artifacts are also visible on the depth image as a byproduct of stereo depth imaging.

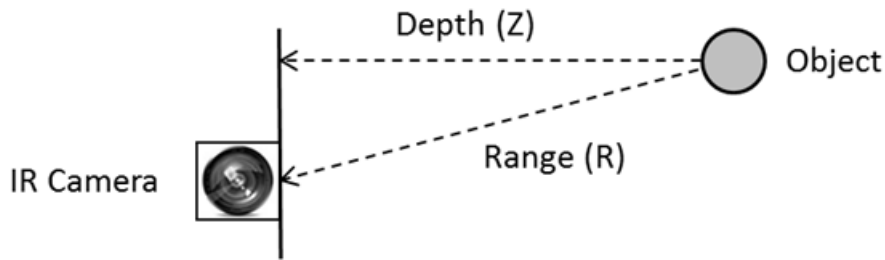


Figure 2.3: Illustrating the difference between depth and range. The RealSense cameras return pixel depth values. Taken from [77].

$$\text{Depth RMS error} = \frac{z^2 \cdot \text{subpixel}}{f \cdot b} \quad (2.1)$$

where  $z$  is the depth distance, *subpixel* refers to the subpixel RMS error and usually falls between 0.05 and 0.1,  $f$  is the focal length in pixels, and  $b$  is the baseline, i.e. the physical distance between the center of the two imagers. Hence, depth error is proportional to the square of the depth distance between the camera and the object. The D455 is notable for having a baseline of 95 mm compared to 55 and 50 mm for the D415 and D435 respectively, which reduces the depth error by a factor of 1.9.

Software interfacing with the cameras is performed using the RealSense Software Development Kit (SDK). A USB 3.0 connection is used to connect each

camera to a PC. RealSense also supports multi-camera configurations for the purpose of simultaneous image capture to either image an object from multiple perspectives (inward-facing) or to create a wider FOV (outward-facing). There is no significant crosstalk between cameras when their FOVs overlap. The cameras must be hardware triggered to capture at the same time; one of the cameras can serve as the master in this scenario to send trigger pulses to the remaining slave cameras. Alternatively, an external triggering pulse can be used. Each camera has a hardware sync port that is used to connect it to other cameras in a network via a triggering cable.

### 2.1.2 Camera Setup and Configurations

The default experimental setup for single-camera data collection is shown in Figure 2.4. The subject is standing a fixed distance away from the camera with the option to either be standing or sitting.

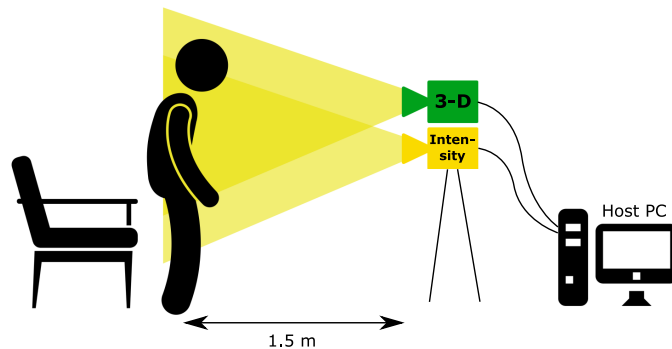


Figure 2.4: Side-view schematic of a typical single-camera acquisition setup. Intensity and 3-D depth channels are represented as different modules connected to the host PC. The subject sits or stands 1.5 m away from the camera, though this distance can be changed.

The use of multiple RealSense cameras allows a variety of camera configurations that serve different purposes. The number of cameras is limited practically by the processing power of the host PC and the number of independent USB hubs due to bandwidth constraints. Our experiments have utilized up to three synchronously triggered cameras positioned around the subject at various an-

gles. Figure 2.5 shows three different configurations that have been used.

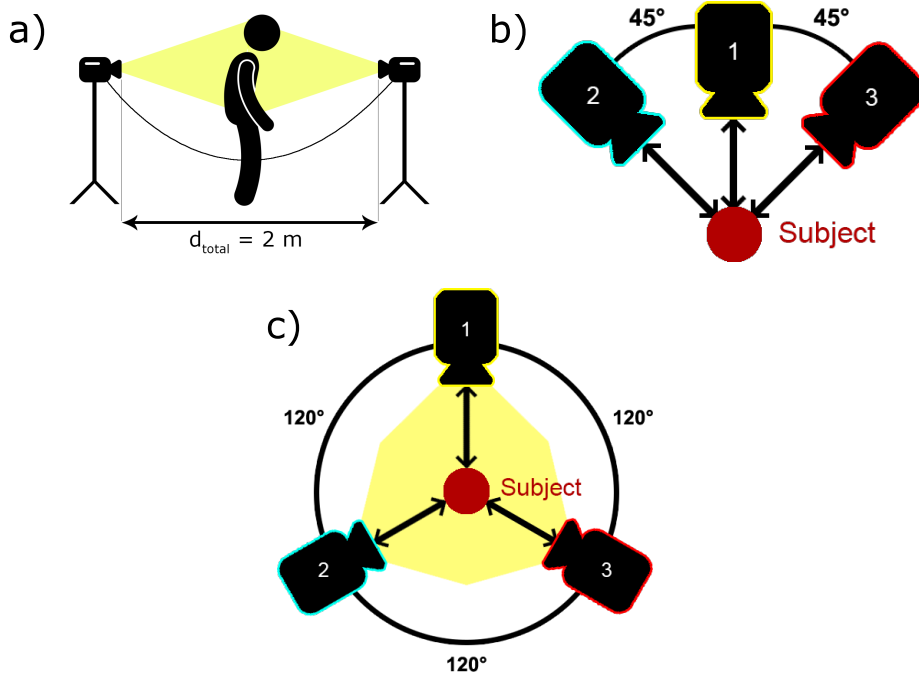


Figure 2.5: Multi-camera configurations. a) “Front-back” configuration, shown from a side view. b) “Three-in-front” configuration, shown from an overhead view. c) “360-degree” configuration, shown from an overhead view.

The “front-back” configuration shown in Figure 2.5a consists of two cameras, one in front of the subject and one behind the subject. Thus, the cameras have view of the subject’s front and back. The distance between the two cameras is known. This configuration was used for an approach to motion tolerance described in the next section.

The “three-in-front” configuration in Figure 2.5b uses three cameras placed in front of the subject at 45-degree angles. The purpose of this configuration is to have three sources of rPPG and depth signal to perform blind source separation, as was done with RGB signals in [52]. Additionally, the use of three angled cameras can potentially allow for construction of a more detailed face model using depth information, which can provide insight into head motion.

The “360-degree” configuration in Figure 2.5c consists of three cameras



placed at 120-degree angles around the subject. This setup ensures a view of the subject's front and back regardless of the angle that they are facing, which is used for motion tolerance.

### 2.1.3 GE Dash Patient Monitor

The GE Dash 3000 Patient Monitor (shown in Figure 2.6) manufactured by General Electric is a portable monitoring system used as a clinical gold standard for the purpose of ground truth vital sign measurements. The Dash includes several vital sign monitoring features, including ECG for HR and RR and a pulse oximeter for SpO<sub>2</sub> data.



Figure 2.6: GE Dash 3000 Patient Monitor.

For the purpose of obtaining physiological data to act as a gold standard during subject data collection, a 3-lead ECG configuration was used along with a pulse oximeter. Four electrodes, with right leg (RL) as a reference, are used to form Einthoven's triangle creating three lead configurations (I, II, and III) using the left arm (LA), right arm (RA), and left leg (LL) as shown in Figure 2.7. The location of these adhesive electrode pads was chosen to be easily accessible for subjects to place by themselves. Methods used to extract data from the Dash monitor are described in Section 6.1.

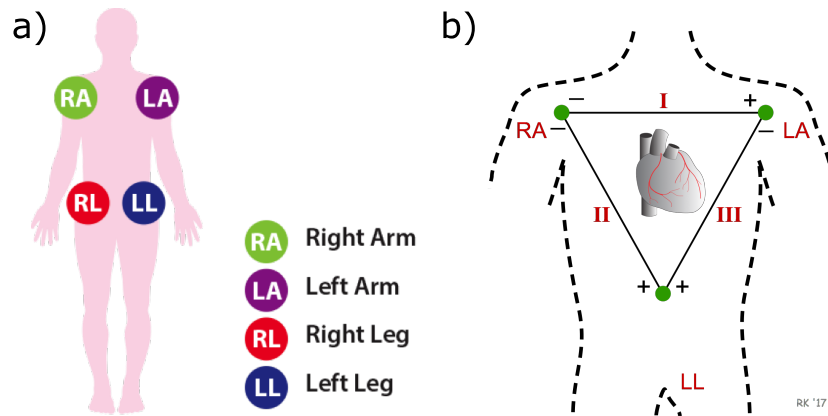


Figure 2.7: ECG adhesive electrode placement. a) Four electrodes are placed on the subject, one on the base of each limb. b) Einthoven's triangle is formed from the placement of electrodes, creating three ECG lead configurations. Images taken from [78], [79].

## 2.2 Extracting Physiological Data

### 2.2.1 Retrieving Images

For each RealSense camera in the system configuration, a pipeline was configured to save incoming frames to the PC's random-access memory. Separate computing threads would subsequently write the frames to the hard drive. Each frame was assigned a number and a hardware timestamp by the camera processor as a method of detecting dropped frames, which could occasionally occur if several cameras were connected and a bandwidth bottleneck was reached. Both IR and depth images were saved individually in the Tagged Image File Format (TIFF), a lossless image format.

### 2.2.2 Regions of Interest

A region of interest (ROI) is defined in this scenario as a selection of pixels with which to extract values for the estimation of vital signs. As the subject can be feasibly be located anywhere in the camera FOV, it is necessary to implement automatic ROI detection.

The cubemos Skeleton Tracking SDK [80] was used to detect skeleton keypoints on a subject within the frame. The skeleton tracking algorithm, when run on an IR image, returns the pixel coordinates of 18 joints on each subject, which can be used for ROI generation as shown in Figure 2.8. To generate a chest ROI, a quadrilateral is defined with vertices consisting of the two shoulder joints (3 and 6) and the two hip joints (9 and 12). As shown, this ROI generation remains robust even when the subject is standing at an angle away from the camera. The quadrilateral is bisected based on its vertical height to create two sub-ROIs for the chest (upper) and the abdomen (lower). For a more detailed analysis, the quadrilateral can be divided into ten horizontal strips of equal height to create ten sub-ROIs.

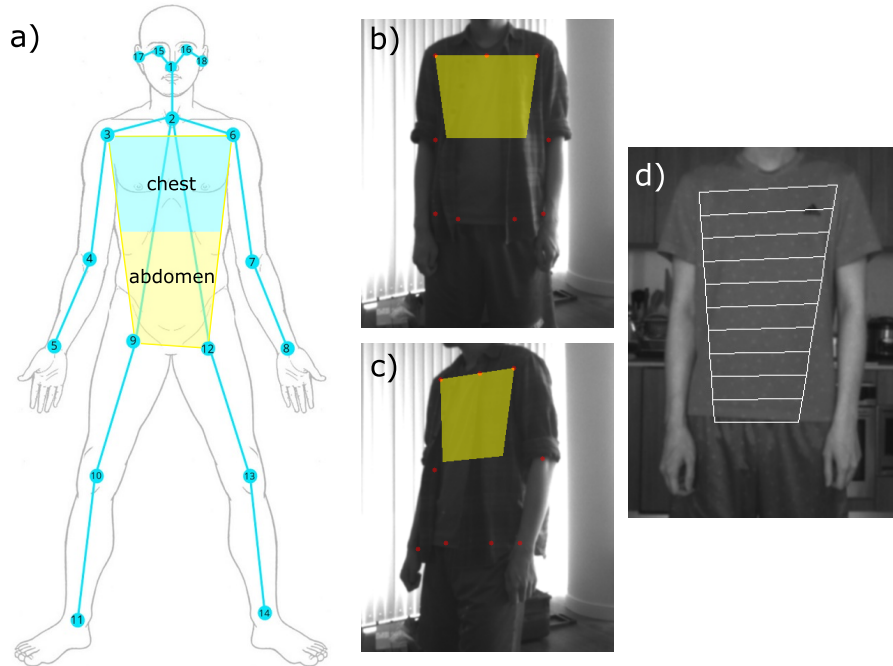


Figure 2.8: Generating ROIs for chest and abdominal regions. a) 18 joints are detected on each subject. Chest and abdominal ROIs can be estimated based on the detected joints. b) Detected chest ROI overlaid on an IR image. c) Chest ROI detection is still possible when subject is standing at an angle away from the camera. d) Chest/abdominal ROI can be divided into horizontal strips for further analysis of motion based on area of the body.

While chest and abdominal ROIs are easily estimated with skeleton tracking,

face ROI detection is more difficult due to the lack of reliable joints on the head. The algorithm includes landmarks on the face (joints 1, 15, 16, 17, and 18), but based on experimentation, the accuracy of these landmarks is generally low and is often negatively affected by eyewear. A separate face detection method was required.

Included in OpenCV 3.3 and later is a “deep neural network” (DNN) face detector module that provides fast and accurate face detection. The detector draws a bounding box with returned coordinates of all faces in an image. Initially, running the face detector on the entire image frame returned poor results, likely due to the low pixel resolution of the face relative to the image size. Therefore, an approach was taken to use skeleton tracking to generate an approximate bounding region of where the face would be located based on the detected joints. Using the two shoulder joints (3 and 6) as the lower vertical bound, a rectangular box was defined with a width of twice the distance between the two joints and a height of three times the distance. The IR image was cropped to this box and used as an input to the face detector, which accurately detected the face with a  $> 95\%$  confidence with front-facing subjects.

While [52] suggested to perform face detection on every individual frame, the generated ROI moves slightly between frames, which introduces an additional source of noise to the signal. To avoid this, a threshold was implemented to avoid small shifts in ROI location, referred to as “jitter tolerance”. This tolerance threshold, which is determined experimentally, causes the previous ROI to be retained if the coordinate shift of the ROI between frames is below the threshold. This jitter tolerance is applied to individual detected skeleton joints and the vertices of the face ROI.

Once a face ROI is established, smaller sub-ROIs can be generated within the face. These sub-ROIs consist of the forehead and the two cheeks. Figure 2.9 shows these ROIs on an IR image of a subject. The location of the forehead and cheeks is calculated according to the overall face coordinates using proportions based on the dimensions of the face ROI. These sub-ROIs were chosen as they contain significant skin surface area that can be used for rPPG analysis.

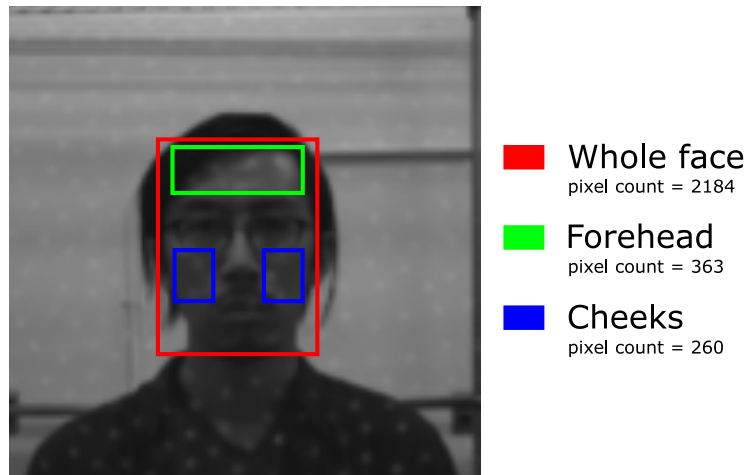


Figure 2.9: Example of sub-ROIs of the face.

### 2.2.3 Physiological Signal Processing

Based on the ROIs described in the previous section, both IR and depth values were extracted from each frame as the mean value of all pixels in each ROI. The result was several “raw” signals that required signal processing to identify and extract physiological data. From experimentation, it was found that HR could be reliably estimated from face IR intensity and face depth movement. Under ideal conditions, a weak HR signal could also be detected from chest depth movements. For RR, chest movement is the primary source of respiratory signal, while face IR Intensity and face depth movement also allow RR estimation under ideal conditions.

MATLAB, developed by MathWorks, was used for the processing of raw IR and depth signals. For HR estimation, a Butterworth bandpass filter was designed with the following parameters:

- Passband edge frequencies: 1 Hz, 2.16 Hz
- Stopband edge frequencies: 0.5 Hz, 3.33 Hz
- Passband ripple: 3 dB
- Stopband ripple: 10 dB

The passband cutoff frequencies represent the approximate bounds of adult resting heart rate: 60 – 130 BPM. These parameters were used as inputs to the MATLAB function *buttord*, which returned the filter order and normalized cutoff frequencies. The function *butter* then generated the filter design with zeros, poles, and gain, which was converted to second-order sections with the *zp2sos* function. Finally, the function *filtfilt* was used to perform zero-phase filtering (also known as forward-backward filtering) on the raw signal, generating a bandpass filtered waveform. Zero-phase filtering is used to prevent phase distortion on filtering.

Once a bandpass filtered waveform is generated, the signal is inverted as per conventions for PPG imaging in literature [81]. The reason for this is to convert the measurement of reflected light into a proportional measurement of absorbed light. Cubic spline interpolation is then applied to effectively increase the sample rate by a factor of 1000. The purpose of this is to generate more data points to increase peak detection accuracy. Peak detection is performed on the interpolated waveform using MATLAB’s *findpeaks* function, with a specified minimum peak distance of  $60/130 = 0.4615sec$ , as 130 BPM is the maximum heart rate defined by the bandpass filter. By averaging the time elapsed between peaks, an estimated HR value can be generated for a given signal.

As an alternate approach, frequency analysis can be applied to HR estimation. This involves performing a Fourier transform on the raw signal to obtain the signal’s frequency spectrum. The frequency bin with the highest power between frequencies 1 Hz and 2.16 Hz (the above method’s passband edge frequencies) is taken as the HR frequency. While this method is relatively straightforward, cases have been observed where the power of the HR frequency bin is not high enough to distinguish from surrounding noise frequencies. A comparison and analysis of these two techniques on participant data can be found in Section 3.5.

For RR estimation, the raw signal is first detrended, removing the linear trend line. In preparation for a Fourier transform, a Hann window is applied to the detrended signal, then MATLAB’s *fft* function is applied, generating

a frequency spectrum. The frequency bin with the highest power between 0.1 and 1.0 Hz (corresponding to 6 – 60 BrPM) is selected as an approximate bandpass center frequency,  $f_{max}$ . This frequency serves as an early estimate of RR. A bandpass filter is then applied with passband edge frequencies  $[0.7 \cdot f_{max}, 1.3 \cdot f_{max}]$  to refine the estimate. Peak detection is performed on the bandpass filtered waveform and the average time between the peaks is taken to calculate an estimation of RR. This approach is used for RR estimation and not HR estimation because respiration tends to present a pronounced peak in the frequency spectrum, whereas the frequency peak for HR may be more difficult to determine.

## 2.3 Motion Tolerance Techniques

A drawback of conventional monitoring techniques, such as PPG, is significant sensitivity to motion artifacts. These artifacts can manifest in the form of distorted amplitudes or false signals. Motion of a periodic nature is particularly corrupting as it appears in the frequency domain and imitates a periodic physiological signal. This means that the subject must remain relatively still during monitoring to prevent artifacts from occurring. Since it has been shown that motion artifacts can be mitigated using accelerometers in wearable devices as a reference [82], a similar approach can be applied to remote monitoring. Using the depth channel of the RealSense cameras, depth information can be used to track subject motion and identify corrupted signal regions, either rejecting this data or developing techniques for compensation. The technique would vary based on the severity of the artifact.

### 2.3.1 Identifying Motion-Corrupted Signal

In order to quantify the severity of a motion artifact, a metric must be developed for the quality of a signal relative to the amount of corrupting noise. Two approaches were identified for this, the first being analysis of depth velocity. Given a depth signal in a particular ROI, the derivative is calculated and plotted as

the depth velocity. Large values in the depth velocity indicate sudden movements and areas where the signal is corrupted. This velocity analysis, shown in Figure 2.10b, can be applied to both depth values and intensity values, as a sudden change in intensity is also indicative of corrupting motion. Alternatively, a moving window can be applied to the depth velocity to calculate the SD in each window. While SD is easy to calculate, it relies on detecting sudden changes in intensity or depth and thus gives no information about the quality of a signal when such changes are not present.

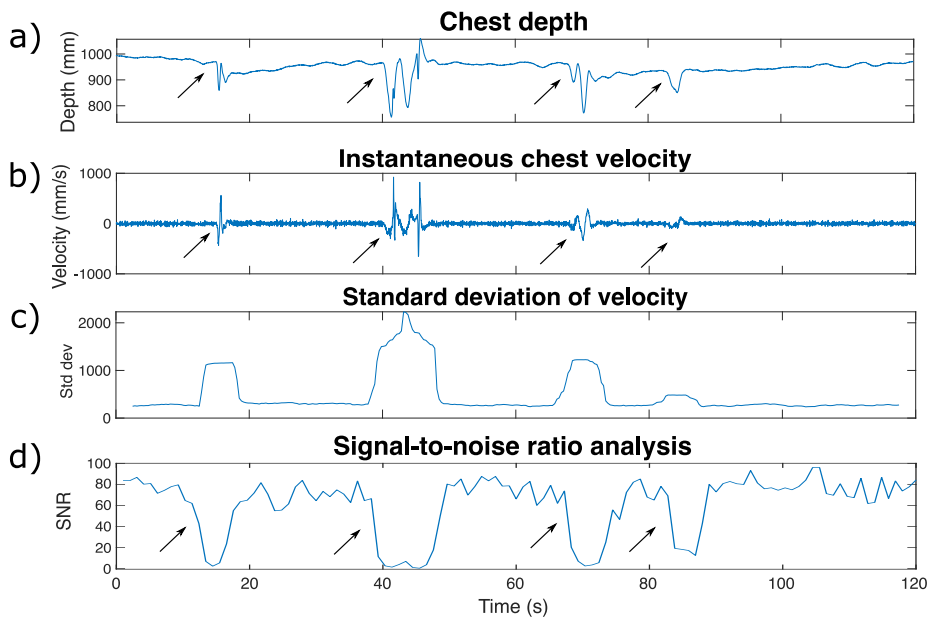


Figure 2.10: Identifying motion artifacts in a depth signal. a) Chest depth signal contaminated by motion events, indicated by arrows. b) Instantaneous chest velocity calculated as the derivative of the signal in (a). High velocity values are indicators of motion events. c) Standard deviation of instantaneous velocity calculated using a moving window. d) SNR analysis of the signal in (a), showing significantly lower SNR values during motion events.

The second approach involves analysis of signal-to-noise ratio (SNR) via the frequency domain. This type of analysis was described by Moço et al. [83], in which the power of the physiological signal is compared to the signal noise power. As a first step, the target frequency range is defined depending on the vital sign of interest; for example, a frequency range of 1.0-2.16 Hz was used



for acquiring HR. The frequency bin with the highest power in this range is defined as the fundamental frequency of the physiological signal. The signal power  $P_{signal}$  is then calculated using the fundamental frequency bin and the two immediate adjacent bins.  $P_{signal}$  is compared to the noise power  $P_{noise}$ , which is the total power of the remaining frequency bins not included in the calculation of  $P_{signal}$ . In some instances, the first two harmonic frequencies of the HR/RR are excluded from the noise, as they are still considered to be components of the underlying signal. The ratio of  $P_{signal}/P_{noise}$  is defined as the SNR. An example is shown in Figure 2.10c, where a moving window was used to calculate SNR on each signal segment using a target frequency range of 0.1-1.0 Hz for respiration. This frequency domain approach is visualized in Figure 2.11, with in-bands and out-bands illustrated.

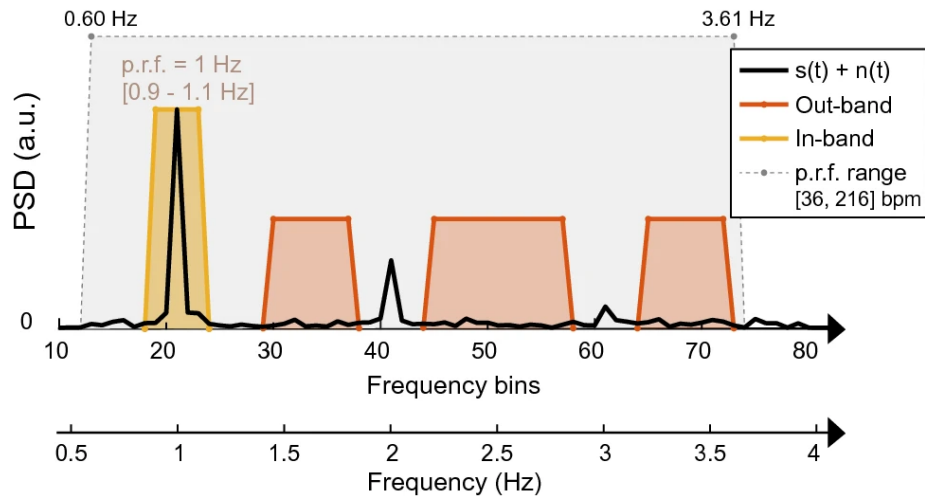


Figure 2.11: Frequency bands used for the calculation of SNR. The signal in this example is an HR-related signal. The in-band is selected as the fundamental frequency of the HR signal and surrounding frequency bins, whereas the remaining bins (excluding 1st and 2nd harmonics of the HR signal) are taken as out-bands. Summing the power of the bins in each band allow calculation of SNR. Taken from [83].

Once motion has been identified in the signal, there are two methods of correction that depend on the severity of signal corruption: rejection and compensation.

### 2.3.2 Motion Rejection

When severe motion artifacts are identified, it is challenging to salvage physiological data from the corrupted signal. In these cases, motion rejection is applied, in which the areas of corrupted signal are identified using either of the aforementioned methods and removed from the overall data. This preserves the remaining signal areas where physiological information can be extracted. This approach was used by Li et al. [60], and an example is shown in Figure 1.11.

A drawback to motion rejection is that the original signal must be cut to remove the corrupted areas, creating signal discontinuities. One approach is to analyze each segment independently and compare the extracted vital signs for each segment to determine the overall estimation. Alternatively, the segments can be “stitched” together to form a continuous signal, as was done in Figure 1.11. This approach relies on the predominance of the physiological signal over potential artifacts introduced by the stitching to produce an accurate vital sign estimate.

### 2.3.3 Motion Compensation

Under less severe motion, it may be possible to extract physiological data from the contaminated signal. The first approach utilized skeletal tracking on each frame to dynamically resize and reposition the ROI as the subject moved. As a more sophisticated method, one or more cameras can be used as a reference to track when and how the subject moves. This allows vital signs to be extracted even during periodic gross body motion. This approach utilized the “front-back” configuration described in the previous section, allowing the subject’s front and back sides to be recorded by the two cameras. ROIs of the subject’s chest and back can either be drawn manually or generated using skeleton tracking. The distance between the front and back cameras  $d_{total}$  was set at 2 m. Based on this, the depth of the subject’s thorax  $d_{thorax}(t)$  could be calculated as follows:

$$d_{thorax}(t) = d_{total} - d_{front}(t) - d_{back}(t) \quad (2.2)$$

where  $d_{front}(t)$  is the distance between the front camera and the subject's chest and  $d_{back}(t)$  is the distance between the back camera and the subject's back.  $d_{thorax}(t)$  changes as the subject breathes, increasing during inspiration and decreasing during expiration, following the natural expansion of the lungs. As the subject moves between the two cameras, calculating  $d_{thorax}(t)$  allows retrieval of the respiratory waveform regardless of intensity or periodicity of body motion. Sections 3.4.2 and 3.4.3 discuss how these methods are applied to experimental datasets.

## 2.4 Summary

In this chapter, experimental methods for retrieving images, capturing signals, and isolating physiological data were described. Various ROIs including the face, forehead, cheeks, and chest have been illustrated. Two motion tolerance techniques were discussed, namely rejection and compensation. In the following chapter, we will discuss how this system is used to analyze experimental datasets to retrieve estimates of HR and RR even when subject motion is present.

## Chapter 3

# Results and Analysis

### 3.1 Overview

Extracting physiological information and estimating vital signs from camera recordings is a multi-step process. Figure 3.1 shows a flowchart that illustrates the processing pipeline of the remote vital signs system. Each numbered step of the flowchart is explained below.

1. The system can consist of a single-camera with one FOV or multiple cameras positioned in a certain configuration to simultaneously record images. Each camera is equipped with an intensity channel and a depth channel. Camera parameters such as resolution, frame rate, and exposure are set.
2. Intensity and depth images are recorded from each camera.
3. ROI generation combines methods such as face detection, skeleton tracking, and segmentation to generate ROIs as detailed in Section 2.2.2.
4. For each acquired frame, the value of the pixels within each ROI are averaged to create a data point. This can be done for both channels. Combining data points from every frame creates a waveform signal that represents changing intensity or depth within an ROI over time.

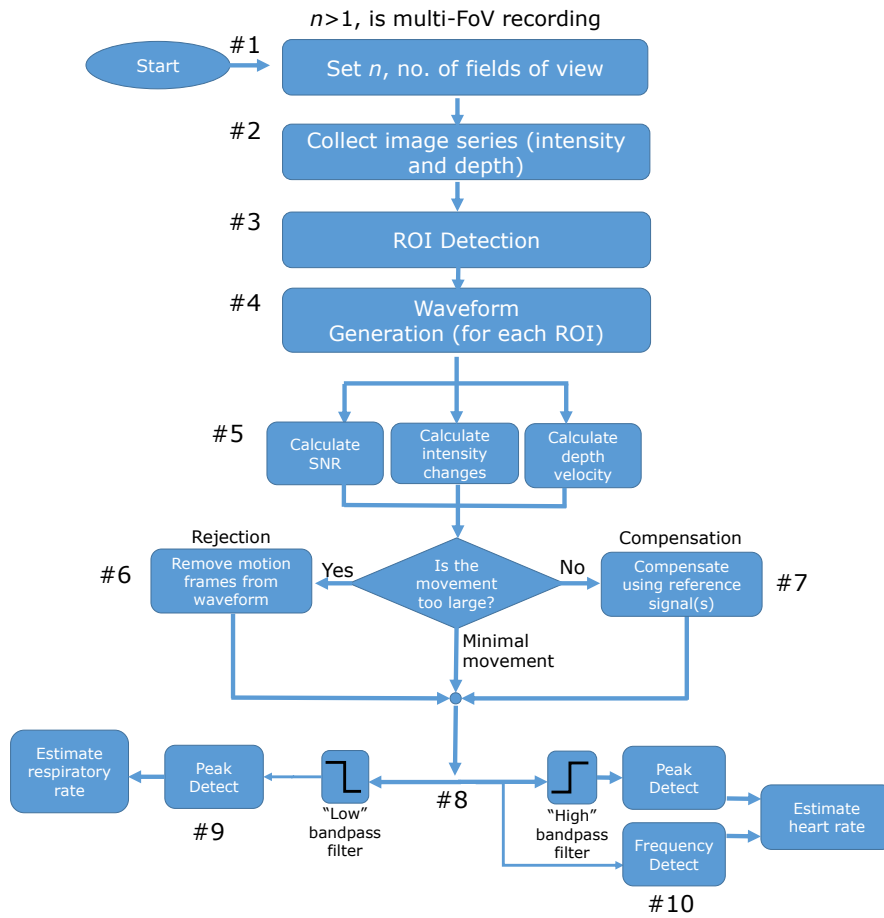


Figure 3.1: Flowchart for processing pipeline of remote vital signs monitoring system.

5. As methods of quantifying the amount of motion contamination in a signal, three approaches are applied: calculation of SNR, analysis of intensity changes, and analysis of depth velocity.
6. If the above methods identify a severe motion artifact in the signal via a drop in SNR or a sudden change in intensity or body velocity, rejection is applied to the signal. Motion-contaminated signal segments are identified and removed.
7. If less severe motion is present, motion compensation can be applied to clean up the signal. Compensation can refer to different methods including

ROI resizing or using a second camera as a reference.

8. When estimating RR, a bandpass filter is applied with a relatively low frequency passband (0.1-1.0 Hz), whereas a higher frequency bandpass filter (1.0-2.16 Hz) is applied for HR estimation.
9. For RR estimation, peak detection is performed on the filtered respiratory waveform. The time intervals between the detected peaks are averaged to obtain an estimate of RR for the sample period.
10. HR can be estimated from peak detection on the filtered respiratory waveform (peak analysis) or by detecting the frequency with the highest power in the frequency spectrum of the signal (frequency analysis). These two approaches are compared on participant data in Section 3.5.

Once estimations for HR and RR are obtained from recorded datasets, it is necessary to evaluate their agreement with the gold standard bedside monitor. To do this, Bland-Altman analysis [84] was used to plot the agreement of the two measurements. Bland-Altman plots are used to quantify agreement by constructing statistical limits using the mean and SD of the measurements; it is often used in medical environments to evaluate a new method or instrument [85]. These plots provide a graphical representation of both the mean bias and the SD of the measurement errors. The measurement difference between the gold standard and the remote system is plotted against the average/mean of the two values in a scatter plot. While some studies involving a gold standard will plot the difference against the gold standard measurement itself, Bland and Altman have stated this to be misleading, as it will always imply a correlation between the difference and the measurement magnitude even when there is none [86]. Therefore, it was decided to plot the difference against the mean. Data was collected from volunteer members of the lab group, as external participants could not be recruited due to the COVID-19 pandemic.

Bland-Altman plots were generated using MATLAB code from [87]. For analyzing the agreement of the remote monitoring system, nonparametric Bland-Altman analysis was used. This was chosen due to the small number of subjects

that were able to be tested; retrieved data was assumed to be non-Gaussian as a result. This analysis was performed on a wide variety of data scenarios.

## 3.2 Heart Rate Results

### 3.2.1 Extracting Blood Volume Pulse

Figure 3.2 shows an example of results from analyzing single-camera datasets to estimate HR. For simplicity, these datasets did not contain excessive subject motion, meaning that motion correction methods were not used. The subject was seated approximately 1 m away from a D435 camera. The subject’s face was taken as an ROI and raw signals from both intensity and depth channels were retrieved. These signals were processed through a bandpass filter to generate BVP waveforms. Peak analysis was used as the primary approach for HR estimation and an estimated HR value was generated from both channels. Note that two different datasets are presented in Figure 3.2, which is why the HR values do not necessarily agree with each other.

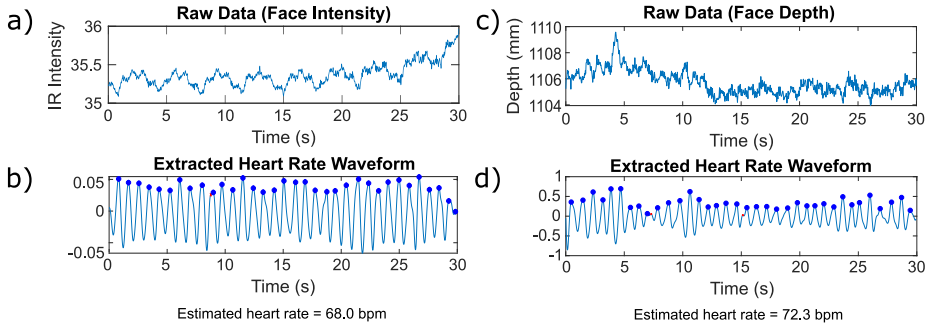


Figure 3.2: Examples of HR estimation from IR and depth data. a) Raw signal taken from IR values of the face over 30 s. b) Bandpass filtered waveform of (a) with peak detection and estimated HR. c) Raw signal taken from depth values of the face over 30 s. d) Bandpass filtered waveform of (c) with peak detection and estimated HR. Note that (a-b) and (c-d) are taken from different datasets.

Hence, it is demonstrated that HR can be estimated from both intensity and depth data. In the following section, more ROIs are analyzed and their results

compared.

### 3.2.2 Comparing Results of Different ROIs

As described in Section 2.2.2, multiple different ROIs can be used for retrieving vital sign estimates, each of which have unique characteristics. This is particularly evident in the face, where subtle movements can significantly skew the agreement of the estimate with the true value. For example, it may be feasible to use the entire face as an ROI for HR evaluation with the rPPG method, but areas such as the eyes and mouth can create artifacts due to blinking or talking. These areas also do not contain significant physiological signal as shown in Figure 1.9, providing another reason for their exclusion from the ROI. However, a reduction in the number of overall pixels in the ROI results in less pixel values for averaging, which results in a drop in SNR on a per-frame basis. This section therefore focuses on determining which ROI(s) contain the strongest physiological signal, as well as discussing some factors that may influence which ROI to use for vital signs estimation. Peak analysis was used as the primary method for HR estimation.

Figure 3.3 shows correlation and Bland-Altman analysis of HR estimation agreement with various ROIs. Both IR and depth methods are used to analyze four datasets of a single subject. The subject was sitting 1.5 m away from a D455 camera, which was recording at 90 fps with a 10 ms exposure time. The structured light projector was enabled at 150 mW power. Each consecutive set of three frames was averaged to reduce depth noise, resulting in an effective 30 fps.

The mean difference (solid line) and 95% limits of agreement (dotted lines) are shown on the Bland-Altman plots. Agreement of the remote HR estimation with the value obtained from the Dash monitor is represented by the reproducibility coefficient (RPC), which is defined as 1.96 times the SD of the differences. A lower RPC value corresponds to a tighter spread of differences. Using the RPC value and the mean difference, the agreement is determined. Pearson's  $r$ -value squared and the root-mean-square error (RMSE) are also given.



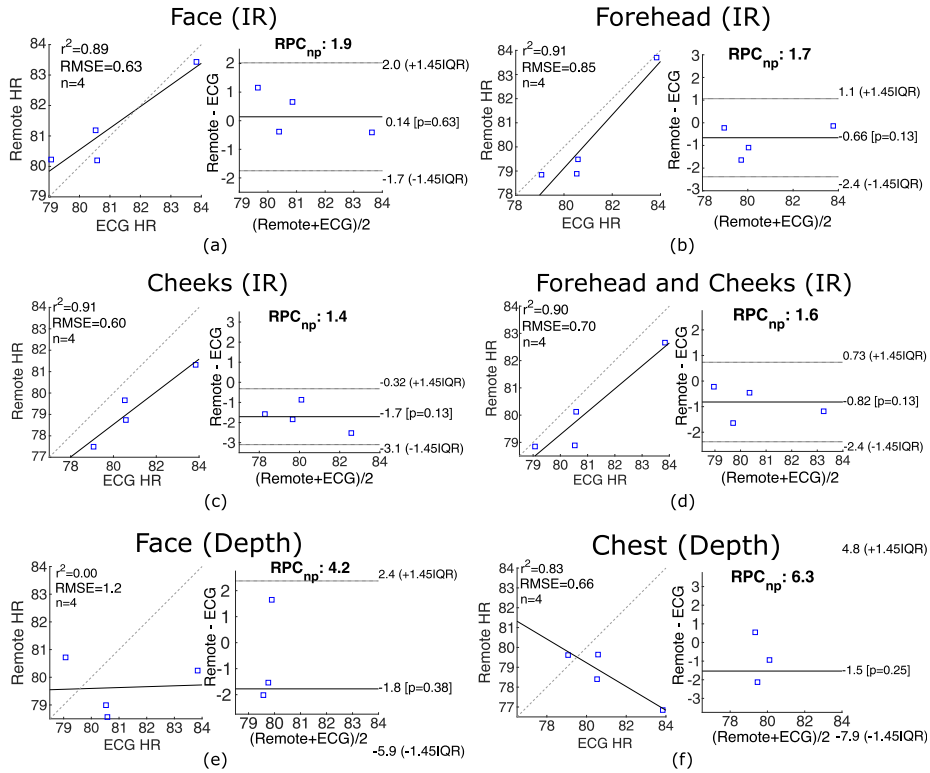


Figure 3.3: Bland-Altman analysis of different ROIs for HR estimation. Data was recorded from a single subject. All values are in BPM. Each ROI has a left and right plot: the left plots display correlation of the remote HR measurement with the reference ECG value, with the line of best fit (solid). The right plots show the mean of the two measurements plotted against the difference (remote HR - ECG HR). The average difference is drawn as a solid line, with 95% limits of agreement drawn as dashed lines.

While all six methods shown above are seen to make a decent estimation of HR within 5 BPM, using the IR intensity channel with an ROI of the forehead and cheeks appear to give the best agreement with the ground truth, with RPC of 1.6 and a mean difference of -0.82 BPM. The forehead ROI by itself has a mean difference closer to zero but a poorer RPC, whereas the cheeks ROI by itself has poorer a mean difference but an RPC closer to zero. Utilizing both ROIs together gives a higher number of pixels, which contributes to better agreement.

The first four estimation approaches (a-d) utilize the intensity channel of the cameras to perform rPPG analysis. Subtle changes in reflected intensity of

light on different regions of the face allow extraction of blood volume variations from which a value of HR can be estimated. These approaches using intensity data are well correlated with the ECG reference value. The approach that analyzes the depth of the face (lower left) estimates HR from ballistocardiographic head motion in the forward-backward direction. As this approach is more easily contaminated with artifacts from gross body motion or other head motion not associated with pulse, its accuracy is lower than the rPPG approaches. Finally, it is difficult but possible to retrieve HR from chest movements associated with cardiac motion (lower right). This approach is difficult due to the very low amplitude of motion, which is often overshadowed by the dominant respiratory motion of the chest wall [88]. Obstructive or baggy clothing renders this technique almost impossible to use practically; however, this subject was only wearing a thin shirt during recording, which allows HR values to be acquired from chest motion. Overall, the ability to extract physiological information from both intensity and depth channels creates a highly redundant system and the ability to switch from one method to another when certain ROIs or channels are not available. For example, in environments with low illumination such as night driving where intensity data may not be readily available or reliable, depth information – assisted by the structured light projector – can still be retrieved for estimation of vital signs.

It is important to note that some ROIs have inherent drawbacks in their usage. Examples of ROI obstructions that appeared in recorded datasets are shown in Figure 3.4. As mentioned before, using the whole face as an ROI will include regions that may contaminate the physiological signal with motion, most notably the eyes and mouth areas. Furthermore, strong reflections of infrared light have been observed from the structured light projector in subjects wearing glasses (Figure 3.4a), which significantly skews the IR intensity values. These issues do not occur with the forehead and cheeks ROIs. However, there are cases where the forehead ROI may be partially or totally obscured, such as cases involving subjects with longer hair (Figure 3.4b) or wearing headwear (Figure 3.4c). Due to these factors, the ROI in which the skin is most reliably exposed is the cheeks.

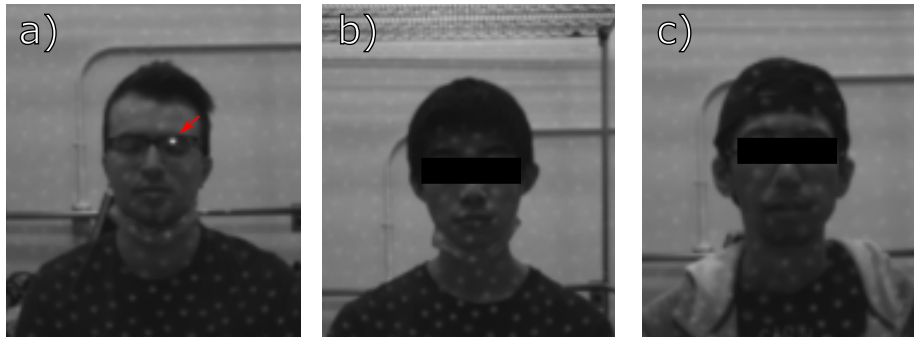


Figure 3.4: Examples of artifacts or obscurations within ROIs. a) Bright dot associated with reflection from glasses. b) Forehead partially obscured by hair. c) Forehead partially obscured by headwear.

HR analysis of participant data using the cheeks ROI, which has been decided as the ROI with the lowest possibility of artifacts unrelated to motion, is shown in Section 3.5. This section also compares the two techniques for estimating HR: peak analysis and frequency analysis.

### 3.3 Respiratory Rate Results

#### 3.3.1 Extracting Respiratory Waveform

Figure 3.5 shows an example of RR estimation using both intensity and depth. For simplicity, these datasets did not contain excessive subject motion, meaning that motion correction methods were not used. The subject was seated approximately 1 m away from a D435 camera. For the intensity channel, the subject's face was taken as an ROI, whereas the subject's chest was used for the depth channel. The frequency spectrum of these signals was plotted and the highest frequency bin was recorded as an early RR estimate. A bandpass filter was then applied to the raw signals with cutoff frequencies determined by the early estimate. Peak detection was then applied to the filtered waveforms to generate a final RR estimate based on the time intervals between the peaks. It is shown that results obtained from face intensity and chest depth are in agreement.

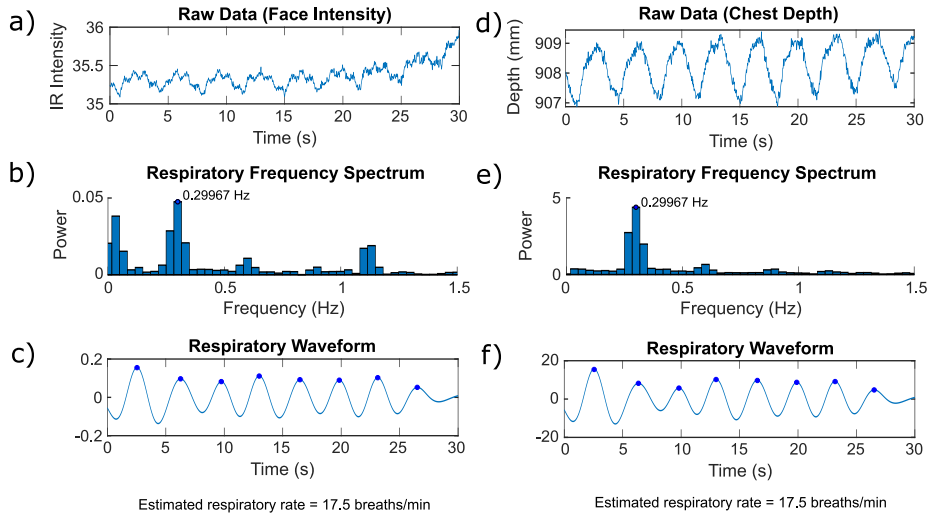


Figure 3.5: Examples of RR estimation from IR and depth data. a) Raw signal taken from IR values of the face over 30 s. b) Frequency spectrum of (a), with peak frequency bin labelled. c) Bandpass filtered waveform of (a) with peak detection and estimated RR. d) Raw signal taken from depth values of the chest over 30 s. e) Frequency spectrum of (d), with peak frequency bin labelled. f) Bandpass filtered waveform of (d) with peak detection and estimated RR.

### 3.3.2 Comparing Results of Different ROIs

Similar to the reference HR values, reference RR values were obtained from the patient monitor, which calculates respiration based on the impedance between ECG leads. Impedance varies based on the amount of air in the lungs, which allows respiration to be detected. During analysis of data obtained from subjects, it was evident that RR values recorded from the patient monitor were inaccurate and could not be used as a trustworthy reference. An example of this occurred in a dataset where the RR reported by the ECG was 56.86 BrPM, which far exceeded the normal range of RR values, casting doubt on the accuracy of the ECG for respiration. Hence, the experiment was revised and additional datasets were recorded with a single subject for the purpose of evaluating RR agreement. Rather than using the patient monitor as a reference, a digital metronome app was used to guide the subject to inhale and exhale at certain frequencies. These frequencies represented the following RR values: 8, 10, 12, 20, 30, and 45 BrPM. The subject was seated 1.5 m away from a D455 camera, which was recording at 90 fps with a 10 ms exposure time. Each consecutive set of three frames was

averaged to reduce depth noise, resulting in an effective 30 fps. Bland-Altman analysis of the resulting datasets is shown in Figure 3.6.

As evidenced by the mean difference of 0.05 BrPM and the RPC of 0.18, taking the average depth value of the chest ROI as the respiratory signal (blue) gave the most accurate estimation of RR. This is expected, as movement of the chest wall is associated with expansion and contraction of the lungs during respiration. However, respiration is also shown to be correlated with head motion, as shown by the analysis of face depth (red). To further examine this correlation, the raw depth waveforms and bandpass-filtered respiratory waveforms were plotted in Figure 3.7 for both face and chest ROIs. The plot reveals that the two waveforms are approximately antiphase to each other; as the chest expands outward during inspiration, the head appears to move backwards. Similarly, when the chest contracts during expiration, the head moves forward and returns to its original position. This relationship can be useful in scenarios where the movement of the chest cannot be detected, such as in the case of a subject wearing thick clothing. In such a scenario, respiration could still be detected through analysis of head movement.

Additionally, as shown by the Bland-Altman plot of RR estimations retrieved from the intensity channel (green), RR can also be estimated from face intensity data. Respiratory features have been shown to be present in the rPPG signal (see Section 1.3.2), allowing estimation of RR from the respiratory induced intensity variations in the IR signal of the face. The ROI consisting of the forehead and cheeks was used as opposed to only the cheeks ROI, as the subject's forehead was reliably exposed in the recordings. The estimations generated from this approach, while not as accurate as the depth-based methods, still produced results correlated with the target RR.

### 3.3.3 Strip Analysis for Respiration Classification

The differences between shallow and diaphragmatic breathing were discussed in Section 1.2.2. Rapid, shallow breathing is often exhibited by patients in respiratory distress. Being able to classify the type of breathing would allow

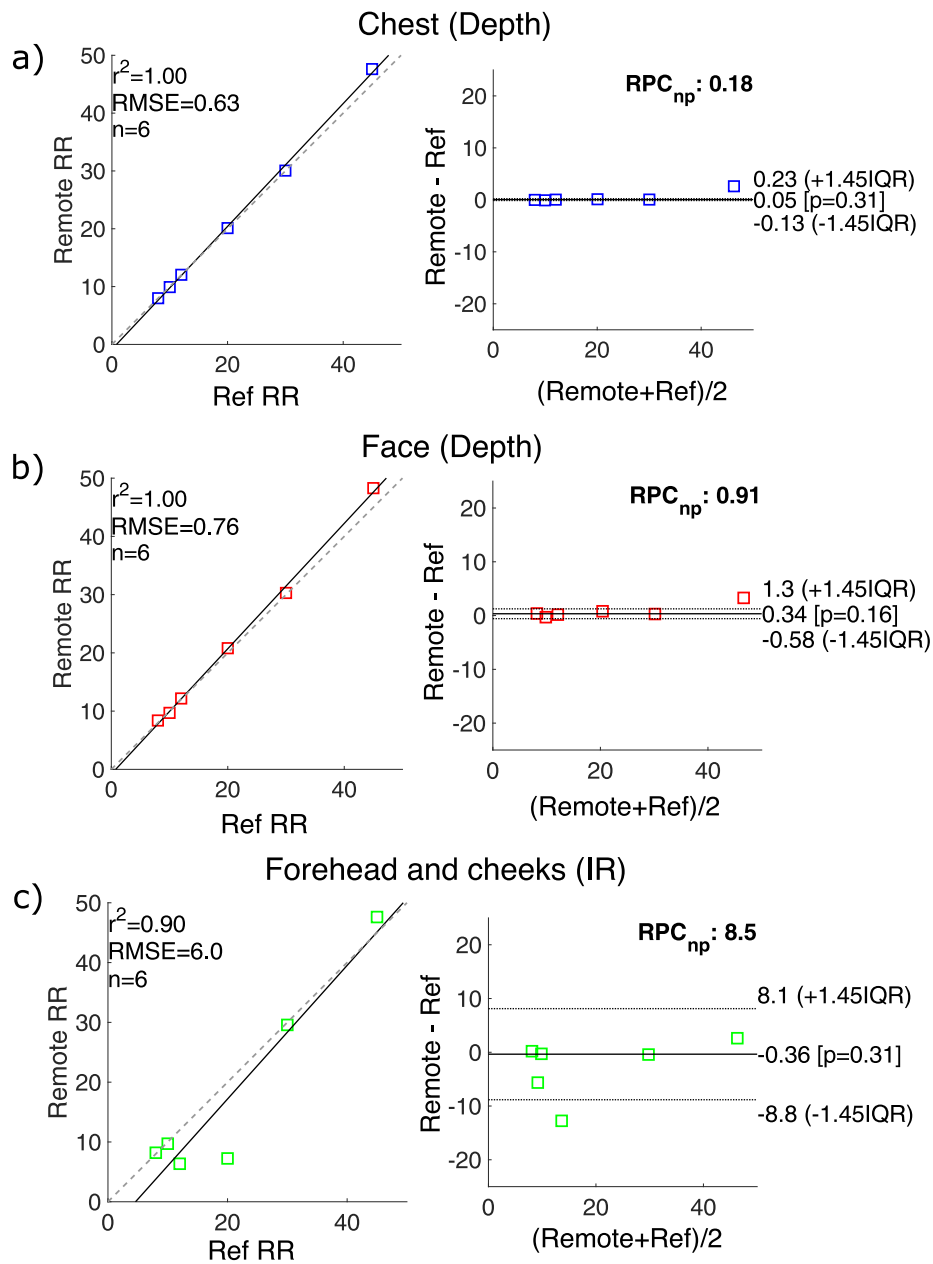


Figure 3.6: Bland-Altman analysis of RR agreement using different ROIs/channels. Each ROI has a left and right plot. The left plots display correlation of the remote RR measurement with the ground truth, with the line of best fit (solid). The right plots show the mean of the two measurements plotted against the difference (remote RR - reference RR). The average difference is drawn as a solid line, with 95% limits of agreement drawn as dashed lines.

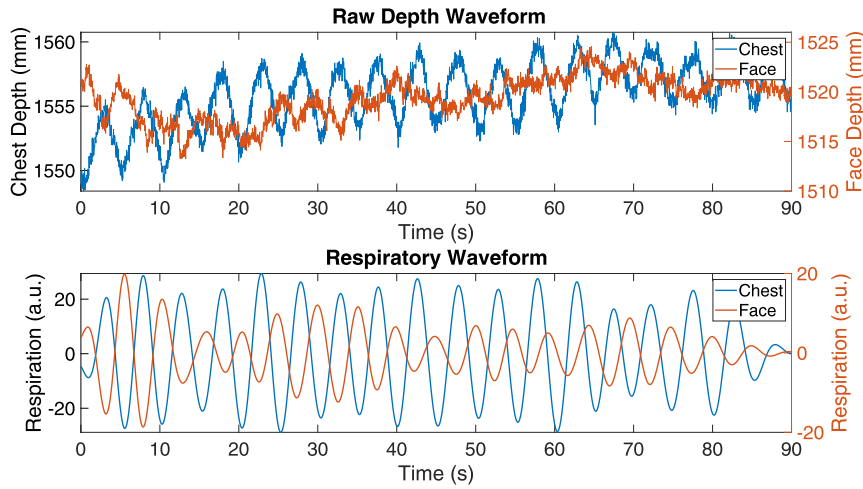


Figure 3.7: Negative correlation of chest and face movement during respiration.

identification of such individuals in times of emergency or during infectious disease screening. It was hypothesized that analyzing motion magnitude of chest and abdominal areas would reveal differences that could identify distressed breathing.

In this experiment, the subject simulated two scenarios: one in which they were breathing diaphragmatically at an approximate rate of 12 BrPM (normal scenario), and one in which they simulated distressed breathing by taking shallow breaths at approximately 30 BrPM (distressed scenario). A digital metronome app was used for timing the breaths. The subject stood 1.5 m away from a D455 camera that was recording at 30 fps with exposure time of 30 ms. Ceiling lights in the lab provided ambient lighting. Once both datasets were recorded, the subject’s frontal area was divided into ten horizontal strips of equal height; each strip was treated as a separate ROI and respiratory waveform analysis was performed on each. Figure 3.8 shows the results of waveform analysis for each strip.

Three metrics were used to quantify the quality of signal from each of the strip ROIs. The first metric was to simply compare the estimated RR from each strip to the known ground truth. This provided an indication if the system was unable to properly detect respiration in a certain strip ROI, but in the event

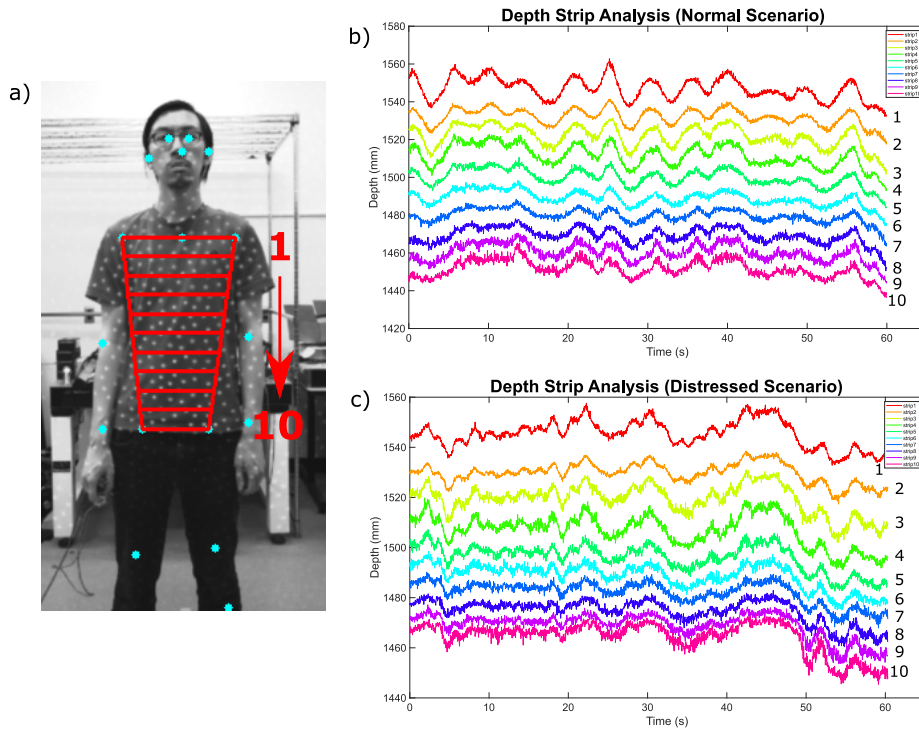


Figure 3.8: Raw depth waveforms for frontal body strip analysis. a) Strip ROIs shown on the subject, labelled from 1 to 10. Joints from skeleton tracking are also shown in blue. b) Depth data from each strip under a normal breathing scenario. c) Depth data from each strip under a distressed breathing scenario.

that respiration could be detected, it gave no information about the quality of the signal. The second metric utilized the detected peaks in the respiratory waveform for each strip. The SD of the times between successive peaks indicated how reliably the system was able to retrieve the true respiratory waveform. For this purpose, it was assumed that the subject’s respiration remained consistent throughout the recording. As the third metric, the SNR of each respiratory signal was determined through analysis in the frequency domain. The power in the frequency bins associated with the detected RR was summed as the signal power  $P_{signal}$ , and the power in the remaining frequency bins, with the exception of low-frequency bins below 0.1 Hz, was summed as the noise power  $P_{noise}$ . SNR is subsequently calculated as  $P_{signal}/P_{noise}$ . Figure 3.9 shows the results of this analysis.

In the normal breathing scenario, RR can be reliably estimated in eight out



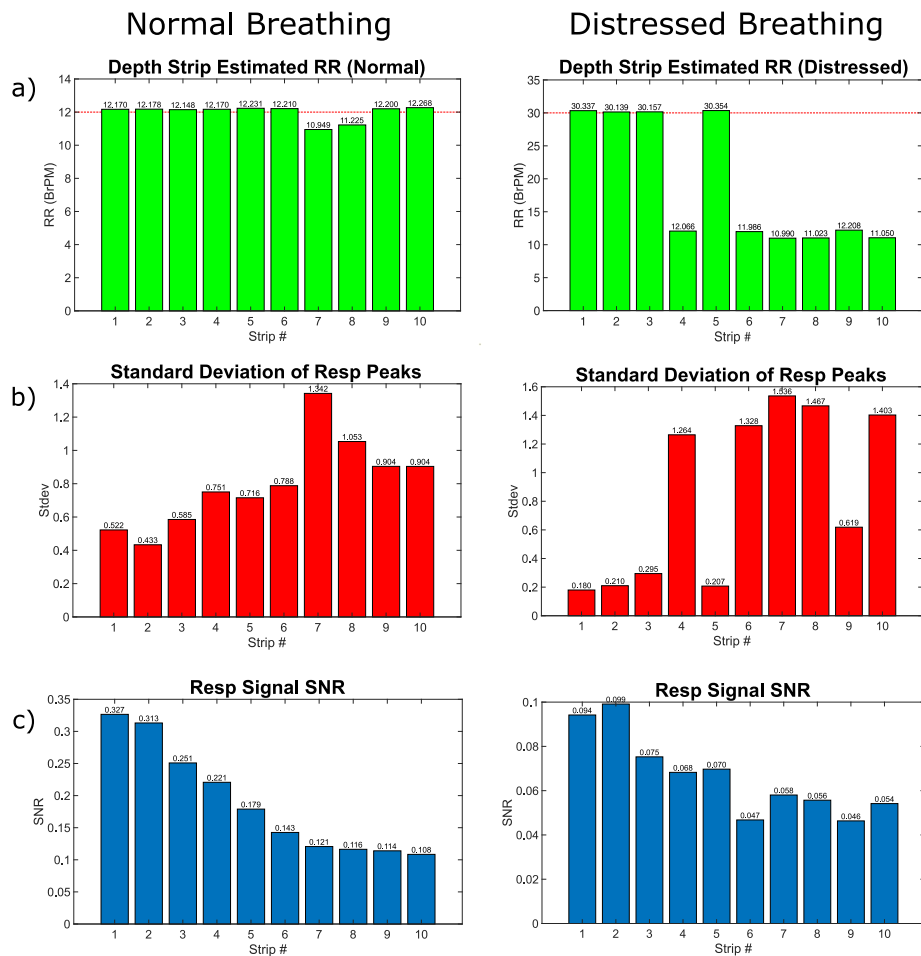


Figure 3.9: Evaluation of strip respiratory signal for both normal and distressed breathing scenarios. a) Estimated RR calculated from each strip. The dashed line indicates the "ground truth" RR as dictated by the metronome. b) Standard deviation of respiratory waveform peak locations. c) SNR of respiratory signal calculated from frequency domain.

of ten strips, whereas only four strips contain an accurate estimation in the distressed breathing scenario. Regarding the SD graphs, SD tends to be higher in the lower abdominal strips (6-10) compared to the higher chest strips (1-5). A higher SD means less consistency in the timing between the detected respiratory waveform peaks and thus a poorer estimation of the respiratory waveform. It is expected that the magnitude of breathing-related motion decreases in the abdomen relative to the chest, which naturally results in higher SD of the peak intervals. This is also reflected in the SNR graphs, where SNR is generally lower

in strips 6-10 compared to strips 1-5. It is noteworthy, however, that the SNR values for the distressed breathing scenario are significantly lower than the SNR values for the normal breathing scenario. This reflects the lower magnitude of motion caused by shallow breathing and associated with respiratory distress. Thus, higher RR values and lower overall SNR values are indicative of distressed breathing.

### 3.4 Motion Analysis

In order to emulate practical scenarios where the subject may not remain entirely still during recording, datasets were taken in which the subject simulated motion events. Using both the intensity channel and depth channel provided by the RealSense camera, various motion correction techniques were applied to these datasets. The depth channel is useful as a direct method of quantifying motion by visualizing the subject's position in 3-D space.

#### 3.4.1 Rejecting Severe Motion

In these experiments, the subject simulated sporadic motion events with the goal of intentionally corrupting the retrieved intensity and depth signals. This included actions such as stretching and moving in a way that obscured ROI areas, examples of which are shown in Figure 3.10. The subject was standing approximately 1 m away from the camera during recording.



Figure 3.10: Examples of simulating severe motion events in a dataset.

Figure 3.11 shows respiratory analysis of an uncorrected depth signal from the subject’s chest ROI. Motion events are indicated by arrows on the depth waveform. In order to better visualize which signal areas are contaminated by motion, the instantaneous chest velocity is plotted as the first derivative of the chest depth. To more clearly identify artifact segments, the SD of the chest velocity is calculated using a 5-second moving window. High SD reflects significant velocity changes within the window, which directly indicates motion.

As a second metric of identifying corrupted signal, the SNR of the chest depth signal is also calculated using a moving window as described in Section 2.3.1. It is shown that SNR drops in signal segments where motion artifacts are present; the drop is significant in this example, implying a strong artifact. When the uncorrected depth signal is processed to obtain a respiratory waveform, it is shown that the motion artifacts overpower the periodic respiratory signal, causing a poor estimation of RR relative to the reference.

Calculating instantaneous depth velocity and SNR creates metrics of identifying corrupted signal. This allows a threshold to be set at which a signal segment is considered to be corrupted. Figure 3.12 shows the process of motion rejection applied to the same dataset as before. In this dataset, SNR was used as the metric for artifact identification, setting the experimental threshold to be 30 upon observing that artifacts are present in segments with SNR below this value. Time values are identified in the SNR waveform where the SNR crosses this threshold. The depth signal is then segmented according to these time values and the motion artifacts are excised from the signal. The remaining signal segments are then “stitched” together to form one continuous signal that retains respiratory information. Performing analysis on this corrected chest depth signal yields an estimated RR that has a much higher agreement with the true RR than the uncorrected waveform.

This same rejection approach can also be applied to HR estimation. Figure 3.13 shows another dataset with simulated motion events affecting the intensity signal from the face ROI. Instantaneous intensity change is calculated in the same fashion as instantaneous chest velocity, and the SD of the intensity change

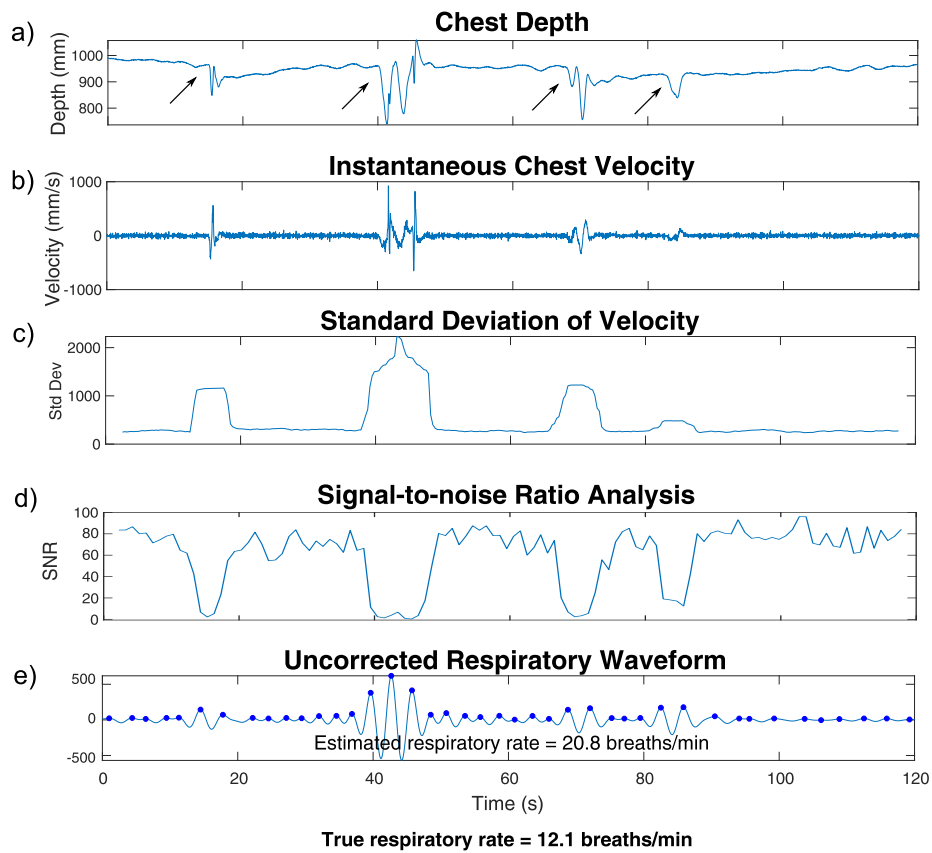


Figure 3.11: Results of RR estimation on a motion-contaminated dataset before rejection. a) Chest depth signal contaminated by motion events, indicated by arrows. b) Instantaneous chest velocity calculated as derivative of (a). c) Standard deviation of chest velocity over time. d) SNR of the chest depth signal over time. e) Generated respiratory waveform without any motion correction, resulting in a poor estimation of RR.

is calculated using a 5-second moving window. While performing HR analysis on the uncorrected face intensity waveform gives a decent estimate due to the prevalence of the rPPG signal, corruption due to motion events is still present in the filtered signal. Rejection this motion also has the potential to improve the estimate further.

In this example, rather than using the SNR to identify motion events as was done in Figure 3.12, the SD of intensity change is used. Motion-corrupted signal segments were chosen by identifying segments with an SD higher than 0.6, a value which was determined by observing the SD of normal segments and

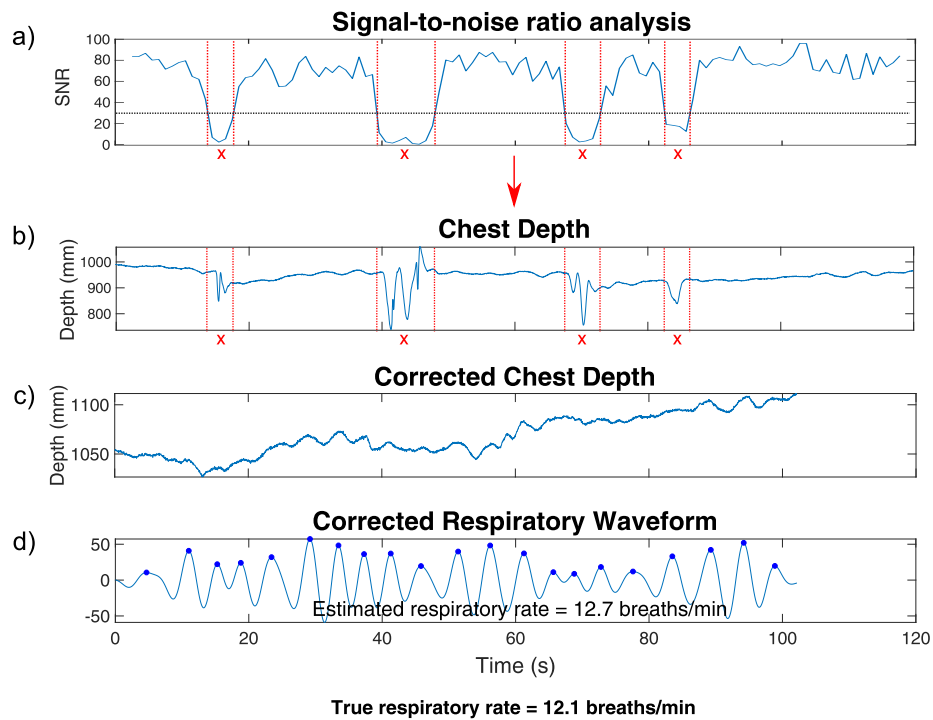


Figure 3.12: Motion rejection applied to a contaminated dataset for RR estimation. a) SNR threshold is set to 30 and time values where the SNR crosses this threshold are identified. b) Chest depth signal is segmented to remove motion artifacts. c) A corrected depth signal is formed by concatenating remaining signal segments. d) The corrected respiratory waveform with a more accurate RR estimate.

creating a threshold value. Removing these corrupted segments and concatenating the remaining segments creates a continuous signal on which HR analysis is performed. The resulting estimated HR of 78.1 BPM is much closer to the reference HR of 78.3 BPM.

This method of motion rejection should only be used when severe motion artifacts are identified and it is unlikely that any physiological information can be derived from the signal. In cases where less severe motion occurs, it may be possible to salvage a physiological signal that can lead to a vital sign estimation. This would involve filtering out the motion-induced signal or using an additional reference to separate it from the desired waveform. These techniques are referred to as “motion compensation”.

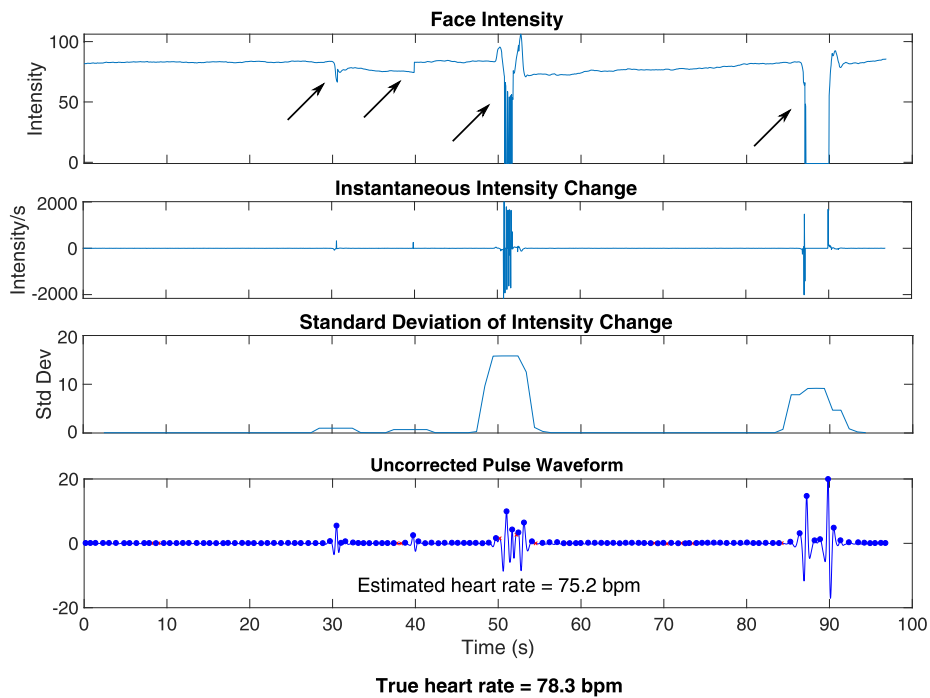


Figure 3.13: Results of RR estimation on a motion-contaminated dataset before rejection. a) Face intensity signal contaminated by motion events, indicated by arrows. Note that an intensity value of zero means that the face detection algorithm was unable to find an ROI in that frame due to obscuration. b) Instantaneous intensity change calculated as derivative of (a). c) Standard deviation of change in intensity over time. d) Generated blood pulse waveform without any motion correction, resulting in a suboptimal estimation of HR.

### 3.4.2 Single-Camera Motion Compensation

Before moving to more complex system configurations, experiments were performed with potential motion compensation using a single camera and FOV. Prior to implementation of skeleton tracking for ROI generation, it was impossible to accurately track an ROI as a subject moved within the frame, and manually drawn ROIs would quickly become inaccurate once the subject moved. Performing skeleton tracking to obtain joint coordinates on each frame of the dataset allows an ROI to be generated for every frame, allowing a continuous signal to be obtained by dynamically resizing and repositioning the ROI according to where the subject is located within the frame. Using this approach, an experiment was performed with the goal of retrieving a respiratory waveform

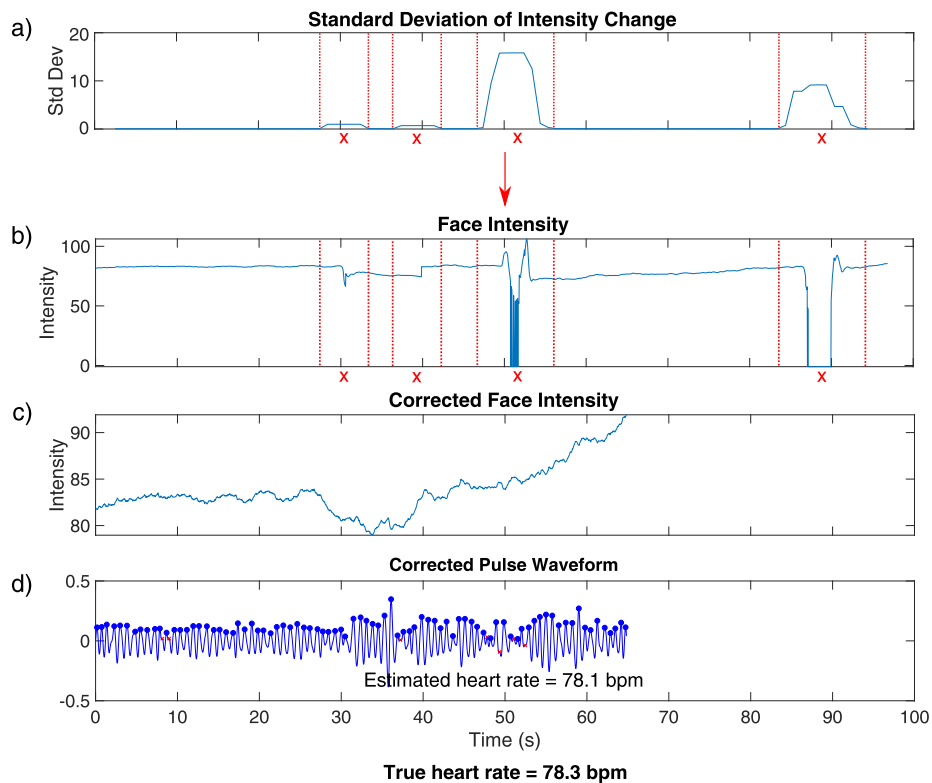


Figure 3.14: Motion rejection applied to a contaminated dataset for HR estimation. a) SD threshold is set to 0.6 and time values where the SNR crosses this threshold are identified. b) Face intensity signal is segmented to remove motion artifacts. c) A corrected intensity signal is formed by concatenating remaining signal segments. d) The corrected blood pulse waveform with a more accurate HR estimate.

under simple subject motion. A dataset was taken of a subject moving around a room and pausing at various locations while still facing the camera. The skeleton tracking algorithm was then applied to the dataset, chest ROIs were generated from the skeleton joints, and a respiratory waveform was obtained. Results are shown in Figure 3.15.

The first 15 seconds of the recording are taken with the subject stationary in order to retrieve their RR to be used as a reference. During movement, the subject kept their breathing as constant as possible. It was found that even during motion, respiration could be extracted from chest movement, regardless of where the subject was located. The estimated RR extracted under motion was in agreement with the reference RR obtained during the stationary period.

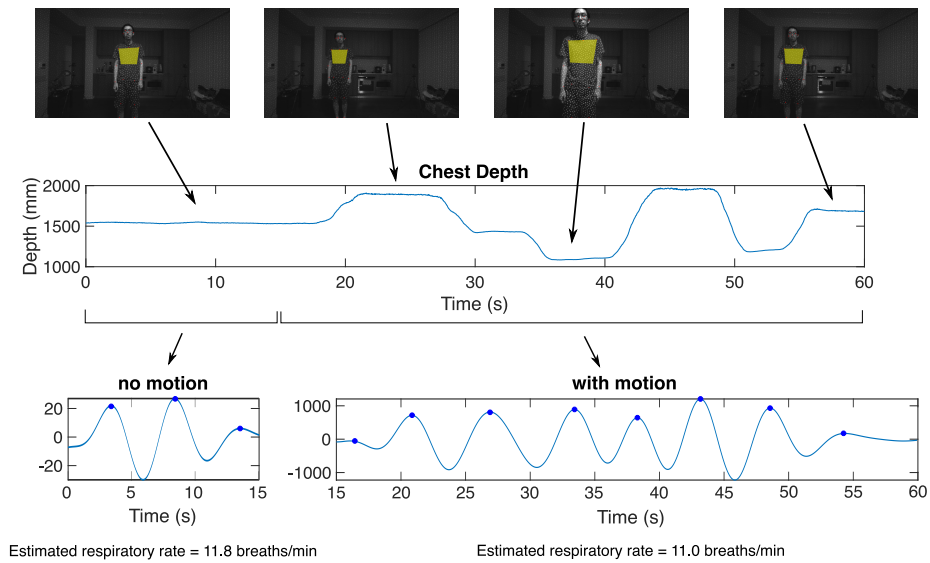


Figure 3.15: Experimenting with skeleton tracking for single-camera motion compensation. Chest ROI is indicated in yellow on sample images from the dataset. RR obtained under motion is in agreement with reference RR obtained when stationary.

While this experiment demonstrated the effectiveness of ROI tracking, it became evident that it was not a universal solution. Specifically, motion compensation using skeleton tracking fails when the movement is periodic in nature. This is one of the major limitations of a single-camera system. To demonstrate this, another experiment was performed with the same setup as previously, in which the subject moved back and forth at a periodic rate. Results are shown in Figure 3.16.

Periodic gross body motion appears in the frequency spectrum at a significantly higher power than respiration-related movement due to its much higher magnitude. Hence, the system mistakenly identifies it as a physiological signal and filters the raw depth waveform as such. The results are a respiratory waveform and RR estimation that are completely inaccurate compared to the reference RR. A more robust approach was required for better motion compensation.



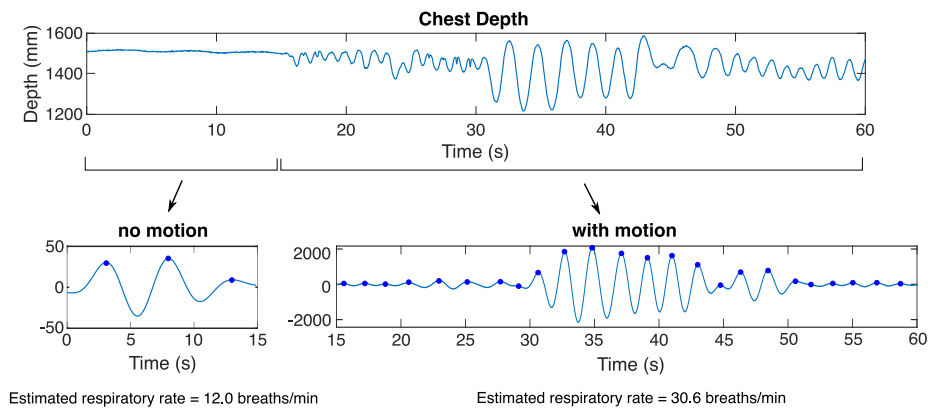


Figure 3.16: Example where periodic gross body motion dominates the physiological signal in the chest depth, causing an overestimation of RR.

### 3.4.3 Multi-Camera Motion Compensation

A single-camera system has no secondary reference to compare measurements with, leaving it susceptible to a variety of artifacts. As demonstrated in the previous section, periodic gross body motion cannot be distinguished from physiological data with a single camera. The introduction of a second camera is necessary to provide the system with an additional source of information that it can compare against the original measurement to isolate certain signal characteristics.

With the goal of improving motion compensation, a method was devised that made use of the expansion and retraction of the chest during respiration. Two cameras were used to create the “front-back” configuration illustrated in Figure 2.5a. The cameras were diametrically opposed, facing inward, and positioned 2 m away from each other. During data acquisition, the subject stood in between the two cameras; the cameras were subsequently designated as the front and back cameras based on which side of the subject they were imaging. Figure 3.17 shows the data and results of this experiment. To simulate motion, the subject moved throughout the camera frame in a periodic fashion. Skeleton tracking was used to dynamically resize and reposition the chest ROI as the subject moved. Depth data provided by both cameras represents distance from each camera to the subject. When these signals are plotted, it shows an antiphase correlation

as expected. Motion compensation analysis involves subtracting the front and back distances from the total distance of 2 m in order to acquire the depth of the chest as a time domain signal. This is further described in Section 2.3.3. Obtaining chest depth over time allows isolation of respiratory chest movement, which can then be analyzed as usual to obtain a bandpass filtered respiratory waveform. The resulting estimated RR is in strong agreement with the true RR.

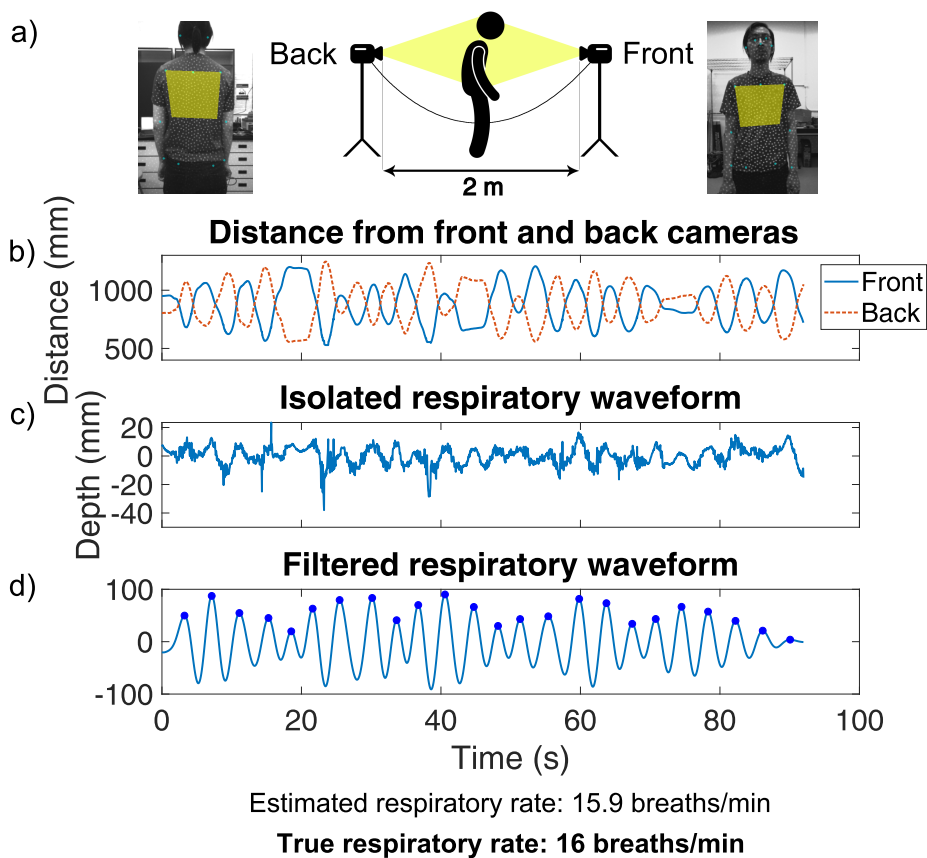


Figure 3.17: Example of motion compensation using two cameras. a) Experimental setup. b) Depth data from front and back cameras represents distance from the camera to the subject and shows periodic body movement. c) Raw respiratory waveform is isolated from the front and back signals. d) Filtered waveform generated and estimated RR is in agreement with the true RR.

Using this two-camera configuration, the system becomes robust to subject movement in two dimensions (front-to-back and side-to-side) when detecting

respiration, even if the movement is periodic in nature. While it is an improvement over the single-camera approach, this method is still limited in certain scenarios. It is contingent upon the skeleton tracking algorithm being able to identify a suitable ROI, which limits the angle at which the subject can be facing. For example, a subject facing at a 90-degree angle to the camera would not present a feasible ROI, as the chest and back are not visible. Hence, the method is not yet robust to rotational motion.

As a further expansion to the system that would allow the subject's chest and back to always be visible, a three-camera system is suggested, as shown in Figure 2.5c. With three inward-facing cameras angled at 120 degrees to each other, the chest and back will always be visible to at least one camera, allowing vital signs estimation using skeleton tracking regardless of the subject's position or angle. While the system was assembled and sample datasets were taken, proper software has not yet been developed at the time of writing this thesis to process the data. As such, it is left for future work.

### 3.5 Participant Data Analysis

For a more in-depth statistical analysis of HR agreement, members of the lab volunteered to be recorded by the remote monitoring system while also connected to the bedside patient monitor. These datasets were captured at 90 fps, but in order to reduce the depth noise, every consecutive set of three frames was averaged to achieve an effective 30 fps. Averaging three frames reduces the depth noise by a factor of  $\sqrt{3} \approx 1.732$ . During recording, subjects were positioned 1.5 m away from the camera and told to remain still. Since remote RR values were in strong agreement with the reference values as described in Section 3.3.2, analysis of participant data was focused on HR estimation.

This section is devoted to assessing the agreement of the remote system using the two methods of HR estimation discussed in Section 2.2.3, namely peak analysis and frequency analysis. The former involves estimating HR based on the time intervals between successive peaks in the filtered waveform, whereas

the latter estimates HR from the frequency bin with the highest power within the specified HR bandwidth. Figure 3.18 shows a comparison of these techniques with datasets taken from six subjects at rest, all volunteers from the lab. The cheeks ROI was used for retrieving intensity values that were used to estimate HR, as it was assessed as being the most reliable ROI in Section 3.2.2.

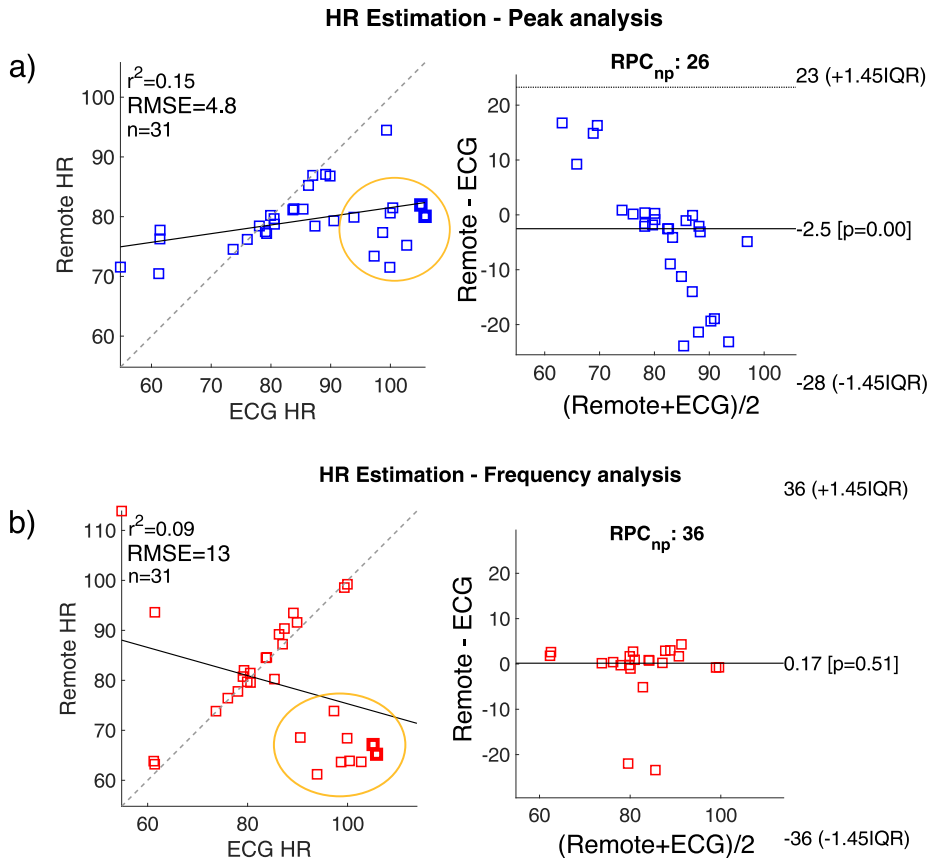


Figure 3.18: Bland-Altman analysis of participant data using two HR estimation techniques. All values are in BPM. Note the cluster of datasets with poor estimations circled in orange, see below for possible explanations. a) Peak analysis. b) Frequency analysis.

The frequency analysis approach has a mean difference of 0.17 BPM compared to the peak analysis approach with -2.5 BPM mean difference. However, the frequency approach also has significantly greater RPC, meaning that there was a wider spread of estimates with frequency analysis.

Agreement between the remote measurement and the ECG value is generally high for mid-range HR values. When observing the correlation plots however, it is evident that the estimation agreement of both methods appears to be lower for abnormally low ( $< 60$  BPM) or high ( $> 90$  BPM) HR values compared to mid-range values. Previous analysis conducted on datasets with a high HR value returned accurate estimations, as shown in Figure 3.19, establishing that the system is capable of detecting high HR values. Therefore, it is important to examine the participant datasets in which poor estimations were made. These particular datasets were linked to three subjects, who will be referred to as subjects 1, 2, and 3.

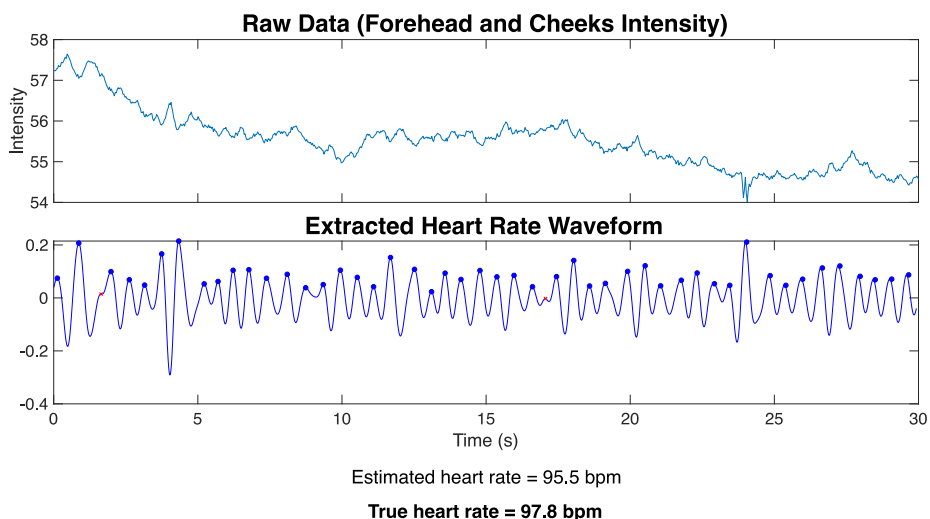


Figure 3.19: Example of good estimation with a high resting HR.

Subject 1 had an average resting HR of 102.1 BPM. However, HR agreement among their datasets was generally poor; estimations averaged 89.9 BPM for the peak analysis method and 72.6 BPM for the frequency analysis method. Upon inspecting this subject's data, significant subject motion was observed in some datasets that corrupted the physiological signal. Hence, motion rejection was applied in an attempt to isolate a BVP signal. Figure 3.20 shows an example with one of subject 1's motion-corrupted datasets.

The HR estimate after application of motion rejection is much more ac-

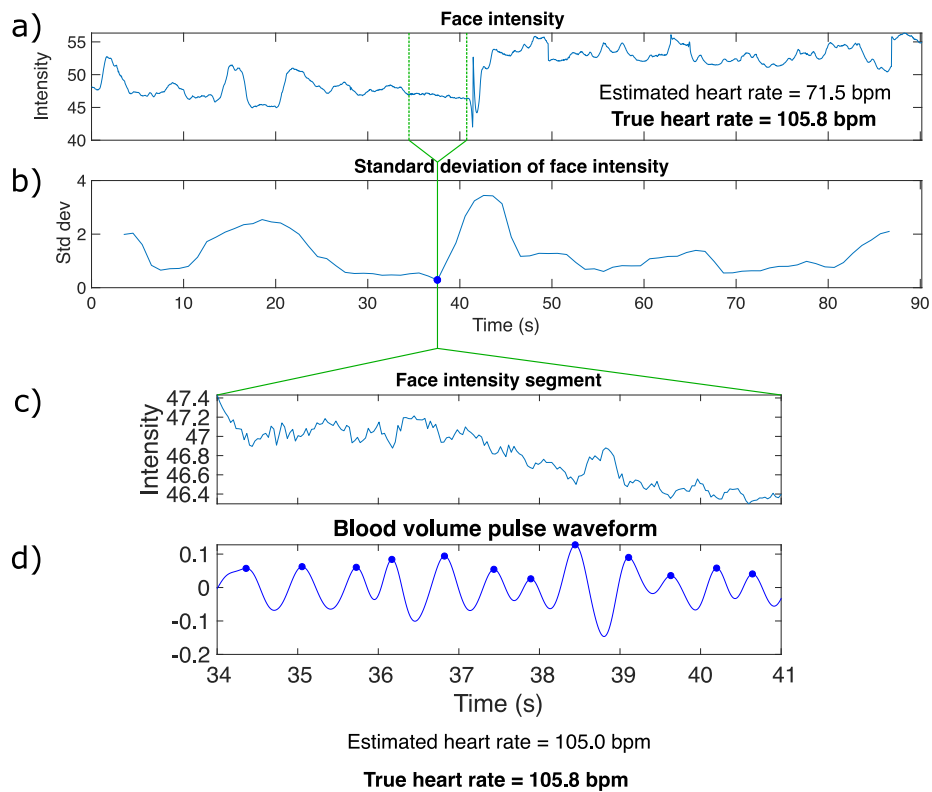


Figure 3.20: Motion rejection applied to participant dataset for HR estimation. a) Face intensity data with significant motion corruption. Estimated HR is in very poor agreement with the reference HR. b) SD of face intensity. c) A signal segment with low SD is identified and used for further analysis. This segment has notably less motion corruption. d) HR analysis performed on the signal segment produces an HR estimate that is much more accurate.

curate compared to analysis of the raw signal. However, some datasets of this subject did not have excessive motion, but the BVP could still not be accurately extracted. Similar cases were observed for subject 2, described below.

Subject 2 had an average resting HR of 98.5 BPM. Like subject 1, HR agreement in their datasets was poor, with estimations averaging 77.2 BPM for peak analysis and 72.3 BPM for frequency analysis. No excessive motion was observed in subject 2's data. It is hypothesized that the inaccurate estimations from subjects 1 and 2 are due to subject-specific characteristics, such as skin pigmentation, which has been shown to adversely affect the SNR of PPG signal [58], or poor blood circulation. This assumption cannot be validated on such a small sample size, necessitating the future recruitment of a larger cohort (see

Section 4.1).

Subject 3 had an average resting HR of 59.72 BPM. This is noteworthy because the bandpass filter used in the physiological signal processing has a minimum passband frequency of 60 Hz. Therefore, HR values lower than 60 BPM may not be accurately detected by the system. When a Fourier transform was performed on the IR signals to retrieve their frequency spectra, a peak was present at the frequency of the ECG HR value (Figure 3.21). Thus, it can be established that the lower bandpass cutoff frequency prevented a proper estimation of HR. While lowering this cutoff frequency would allow detection of lower HR values such as those seen in this subject, doing so may potentially affect the estimation accuracy of other datasets. This would need to be evaluated if changes are made to the filter parameters.

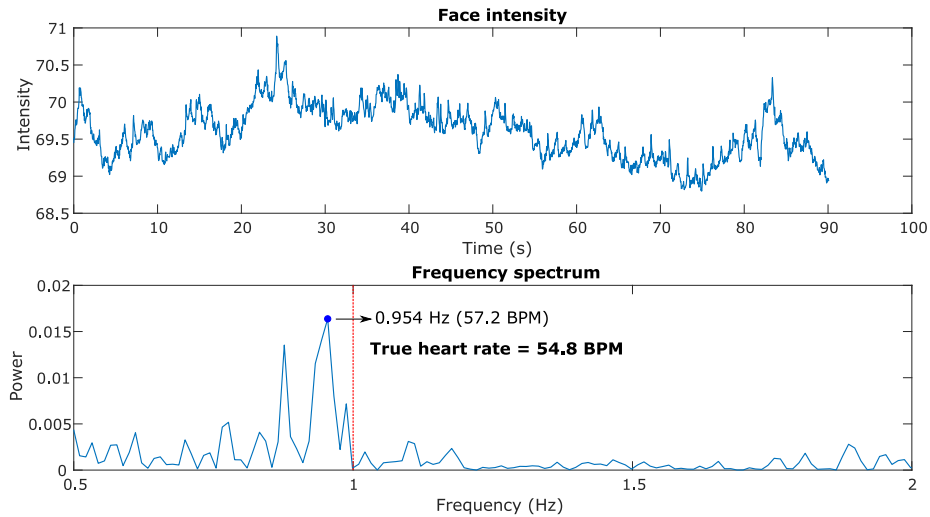


Figure 3.21: Frequency spectrum of a dataset from subject 3. The HR frequency peak (blue dot) is visible but is below the cutoff frequency (red line) for HR estimation. Hence, the correct peak was not picked up by the system.

### 3.6 Summary

Successful isolation of BVP and respiration were demonstrated first in stationary subjects, allowing accurate HR and RR estimations, respectively. The estima-

tions retrieved from different ROIs and channels were compared. Most ROIs examined in this chapter are feasible for vital signs estimation; results from Bland-Altman analysis are summarized in Table 3.1 for HR estimation (Section 3.2.2) and Table 3.2 for RR estimation (Section 3.3.2). The strongest agreement, determined by comparing RPC and RMSE values, is observed for the intensity of the cheeks ROI for HR, and the depth of the chest ROI for RR.

Table 3.1: Summary of Bland-Altman analysis for HR estimation in various ROIs. Values are in BPM.

|                                    | Intensity |          |        |                 | Depth |       |
|------------------------------------|-----------|----------|--------|-----------------|-------|-------|
|                                    | Face      | Forehead | Cheeks | Forehead+Cheeks | Face  | Chest |
| <b>Mean difference</b>             | 0.14      | 0.66     | -1.7   | 0.82            | -1.8  | -1.5  |
| <b>Upper limit of agreement</b>    | 2.0       | 1.1      | -0.32  | 0.73            | 2.4   | 4.8   |
| <b>Lower limit of agreement</b>    | -1.7      | -2.4     | -3.1   | -2.4            | -1.8  | -7.9  |
| <b>Pearson r-value squared</b>     | 0.89      | 0.91     | 0.91   | 0.90            | 0     | 0.83  |
| <b>Reproducibility coefficient</b> | 1.9       | 1.7      | 1.4    | 1.6             | 4.2   | 6.3   |
| <b>RMSE</b>                        | 0.63      | 0.85     | 0.60   | 0.70            | 1.2   | 0.66  |

Table 3.2: Summary of Bland-Altman analysis for RR estimation in various ROIs. Values are in BrPM.

|                                    | Chest Depth | Face Depth | Forehead+Cheeks IR |
|------------------------------------|-------------|------------|--------------------|
| <b>Mean difference</b>             | 0.05        | 0.34       | -0.36              |
| <b>Upper limit of agreement</b>    | 0.23        | 1.3        | 8.1                |
| <b>Lower limit of agreement</b>    | -0.13       | -0.58      | -8.8               |
| <b>Pearson r-value squared</b>     | 1.0         | 1.0        | 0.90               |
| <b>Reproducibility coefficient</b> | 0.18        | 0.91       | 8.5                |
| <b>RMSE</b>                        | 0.63        | 0.76       | 6.0                |

Motion analysis was successfully performed on datasets simulating corrupting motion, demonstrating the usefulness of motion rejection and motion compensation, the latter of which has been expanded to a multi-camera configuration. Analysis of participant data for HR estimation revealed overall agreement, with the exception of a few datasets that were linked to certain subjects. These datasets were investigated in detail, and it was speculated that skin pigmentation may have played a role in lower estimation agreement. It is essential that a larger cohort of participants is recruited to test this theory.



## Chapter 4

# Future Work

Despite the system discussed in this thesis representing a step forward in remote monitoring of vital signs, it requires further development in order to realize its full capabilities. Several potential future advancements have been identified and will be discussed in this chapter; they are ordered by highest to lowest priority. Future students may be able to use these suggestions as guidance for interesting research directions.

### 4.1 Recruitment of External Participants

Due to circumstances surrounding the ongoing COVID-19 pandemic and subsequent lab shutdown affecting the ability to recruit external participants, the analysis and results discussed in this thesis were performed on a low number of subjects. From a statistical perspective, the amount of meaningful information that can be retrieved from this data is limited. In order to gain further insight into the performance of the system, it is highly recommended that the analysis is conducted on a higher number of subjects when possible. These subjects should ideally have a wide range of HR and RR values at rest to fully test the range of the system.

Recruitment of external participants would also permit analysis of the effects of different skin tones on the accuracy of the monitoring system. It has been shown that the SNR of rPPG signals decreases with increasing skin pigmentation [58]. While it was theorized that skin pigmentation played a role in at least one of the subjects studied in this thesis, conclusive statements cannot be made without a larger cohort of participants. The use of depth information in this system is likely to contribute to overcoming the issue posed by darker skin tones. Furthermore, the shape of the PPG waveform has been shown to change according to the amount of blood circulation in the subject [89]. While this can be a particular concern in hospitalized patients whose blood flow may be reduced, the effect may also be present across a larger group of participants with varying states of overall health. In these cases, it may be more difficult to acquire an accurate estimate of vital signs.

An ethics protocol for recruiting external participants was first submitted to the Research Ethics Board at the University of Toronto in November 2019 and approved in January 2020. An amendment to the protocol was approved in September 2020. This protocol permits recruitment of 20 healthy student volunteers from which data can be recorded for analysis. While current university policies do not allow unaffiliated persons to enter the lab, participant recruitment can commence as soon as these restrictions are lifted.

After evaluation is performed on healthy participants and the accuracy of the system is validated, it would be ideal as a next step to also record and analyze data from (1) elderly persons in a home environment and (2) patients in a clinical environment. These are two major application scenarios that have motivated the project thus far, and it is important to assess whether or not the system can derive meaningful insight into the overall health of these individuals. Even further steps would be to implement warnings or alarms if an abnormal health status is detected.

## 4.2 Improved Scenarios and Analysis Methods

Participant data analysis (Section 3.5) was limited to scenarios where the subject was standing still. Additionally, these datasets were taken at resting conditions, which limited the potential range of HR and RR that were tested. For a more thorough analysis of the system’s capabilities, it is recommended that data is analyzed under a wider variety of scenarios. One such example is during exercise, as both HR and RR are elevated. However, motion artifacts are significant during physical activity, making PPG measurements difficult, especially for remote implementations. For this reason, an exercise routine should be chosen that limits upper body movement as much as possible, such as the use of an exercise bike.

Another scenario to be investigated is a subject lying in bed, in either a supine or slightly inclined position. Subjects in a clinical environment are likely to be in bed, rather than standing or sitting upright. However, it has been shown that PPG amplitude is slightly lower in supine subjects [90], likely due to venous pooling caused by gravity. Hence, it may be more difficult to estimate vital signs from rPPG. Conversely, chest movement due to respiration may be easier to detect via the depth channel in supine subjects, as the subject’s back is against the bed surface, potentially making chest movements more prominent. Gross body motion is less concerning with subjects in bed, which may result in motion compensation techniques being unnecessary.

In addition to a wider range of scenarios, the signal processing techniques used in this system were limited in scope compared to what has been done in literature. More complex algorithms have been described [23] that may potentially allow vital signs to be extracted with higher accuracy and under more challenging scenarios. Examples include wavelet decomposition [67] and blind source separation, though the latter requires multiple channels. Some studies have utilized information from the background of the camera images to eliminate corrupting factors such as illumination [73]. Our datasets contained examples where motion or illumination prevented vital signs from being estimated. It is believed that a more sophisticated method of analysis would permit estimates

to be made in these scenarios.

As a method of improving estimates that does not require additional algorithms, it may be feasible to use one vital sign to estimate the other. The current system implementation detects HR and RR within relatively wide frequency bandwidths (1.0-2.16 Hz for HR, 0.1-1.0 Hz for RR). Our RR detection method improves on this by selecting an early estimate from the frequency spectrum to perform bandpass filtering. In a similar way, it is speculated that if either HR or RR are known, the other vital sign can be better estimated using this known information. For example, a subject typically has high HR and RR during exercise. If a high RR is detected, the system can attempt to estimate HR in a narrower bandwidth corresponding to the upper half of the 1.0-2.16 Hz range. This may result in more accurate estimations, but is dependent on the relationship between the two vital signs. Also, this correlation is not representative of all scenarios, such as a subject intentionally slowing their breathing to calm down. This approach should therefore be implemented dynamically and evaluated.

Improvements to equipment would also be beneficial. Intensity-based measurements in this system were taken using the 8-bit IR channel of the RealSense cameras, which provides 256 grey levels. The amplitude of intensity variations observed due to blood volume changes were fractions of a grey level, only obtainable from averaging several pixels in an ROI. An increased number of grey levels could make the system more sensitive to intensity variations, which would allow vital signs to be measured under more challenging conditions. While the RealSense camera is capable of capturing 16-bit IR, it is unrectified and Intel does not recommend its usage. It may be necessary to look for new hardware in this case. The RealSense cameras are simply a combination of intensity and depth channels in a certain configuration; it would be ideal to assemble a system that is better suited to this particular application. For example, the intensity channel of the RealSense camera is sensitive to the entire visible spectrum and the NIR spectrum up to 865 nm, but it has been shown that certain wavelengths (e.g. green) are better at capturing rPPG signals [52]. Therefore, an ideal intensity channel would only be sensitive to these wavelengths to retrieve

a stronger rPPG signal. A better depth channel can be accomplished with a next-generation 4f imaging system, which is discussed in Section 4.6.

The maximum frame rate used for recording data in this thesis was 90 fps. This value is more than sufficient for capturing physiological information. Higher frame rates may allow for more detailed analysis of subtle body motion that contributes to motion artifacts. Calculating ROIs for each frame would result in smoother ROI tracking as the subject moves. However, increased frame rate is limited by processing bandwidth constraints and exposure times; higher frame rates require shorter exposure times, resulting in darker images.

While motion is most often considered detrimental to vital signs monitoring, it may be worthwhile to consider motion as a “vital sign” itself. For example, hospitalized patients who exhibit frequent motion may be more alert and functional than those who move less. Conversely, patients who are incapacitated would naturally exhibit less motion. This may be worth exploring in a future project.

### 4.3 Multi-Camera System Expansion

As mentioned in Section 3.4.3, progress into a three-camera setup has been made for motion compensation that would provide increased robustness to subject position and angle. This setup utilizes the “360-degree” camera configuration shown in Figure 2.5c. An example of the FOVs given by the three cameras is shown in Figure 4.1.

In order to be robust to subject rotation, the system must be capable of ranking its three cameras in terms of which has the better view of the subject. The two highest ranked cameras would have an adequate front and back view of the subject with which to perform motion compensation. As the subject moves, however, the cameras with the best view are likely to change, meaning that the system must constantly keep track of this ranking and dynamically switch FOVs if an ROI becomes invalid in one camera’s FOV.

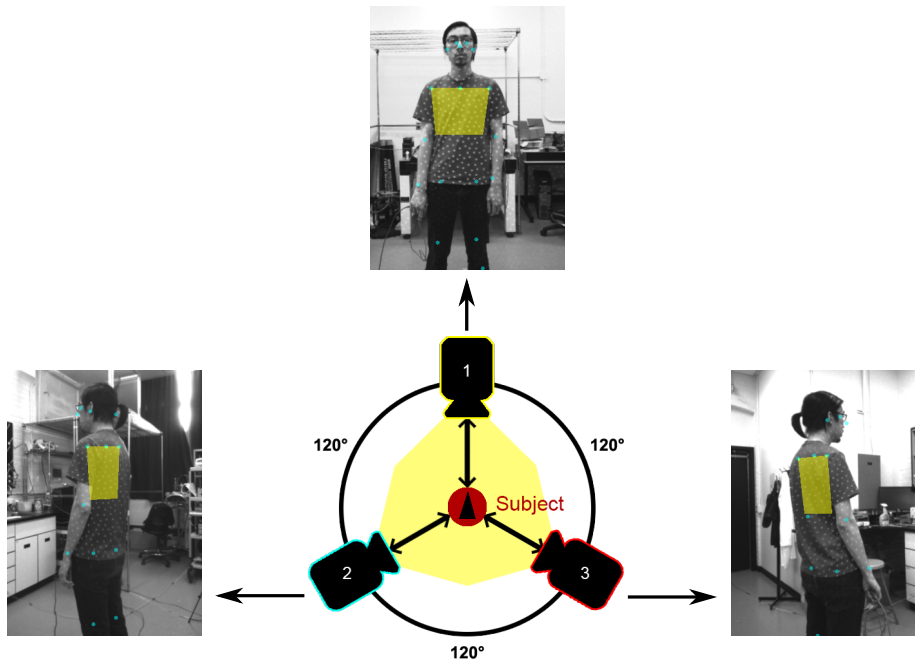


Figure 4.1: FOVs of the three cameras in the 360-degree camera configuration. Skeleton tracking is applied to show an approximate chest ROI.

Another task that can be performed with the 360-degree camera configuration is generation of a 3-D model of the subject using depth images. In this approach, depth images from each of the three cameras are converted to 3-D surfaces. Knowledge of the cameras' relative locations allows their three coordinate spaces to be converted into a single coordinate space. The 3-D surfaces can then be aligned with each other to generate a 3-D model of a body part, as shown in Figure 4.2.

Creation of a 3-D model allows much more detailed analysis of motion and how certain body parts deform with physiological activity. The head and torso are prime ROIs for this reason. Ballistocardiographic head motion and chest motion due to respiration could be easily visualized with a 3-D model. With this approach, it is vital that depth noise is minimized in order to obtain stable 3-D models and visualize small movements associated with cardiac and respiratory activity.

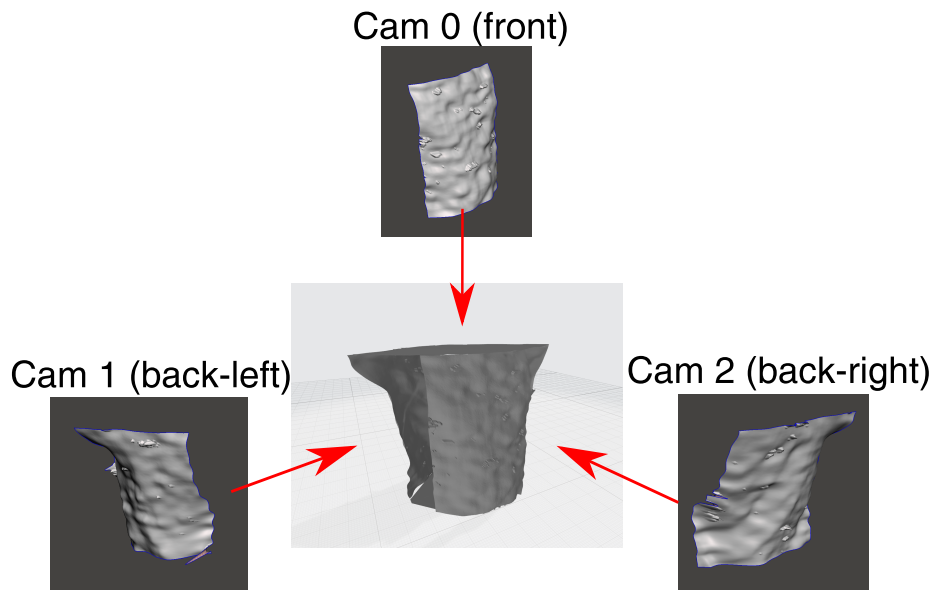


Figure 4.2: 3-D surfaces used to create a 3-D model of a subject's torso. Surfaces must first be converted to a single coordinate space to be aligned with each other. Images generated with Autodesk Meshmixer.

#### 4.4 Motion Compensation Analysis for Heart Rate

The motion compensation techniques described in Sections 3.4.2 and 3.4.3 were used to robustly estimate RR. Attempts were made to perform the same processing for HR estimation, but the amplitude of intensity and depth changes corresponding to HR were found to be too low to be salvaged from even minor motion. Changing illumination from the subject moving throughout a room created noisy intensity signals, whereas ballistocardiographic head motion was too small to be detected after compensation techniques were applied. However, motion rejection can still be applied for HR measurements to remove corrupted signal, as demonstrated in Figure 3.14. Ideally, a method should be developed to extract physiological information relating to HR even under constant motion. This may potentially involve the addition of one or more channels, such as RGB, to apply source separation on face intensity data gathered from each channel. As described in the previous section, more complex signal processing methods could also be applied.

## 4.5 Remote Measurement of Oxygen Saturation

While HR and RR are important physiological parameters, they are only two of a larger number of vital signs that are considered essential for health monitoring. Oxygen saturation in blood is an important vital sign to monitor as cells in the body require a constant supply of oxygen. In order to measure oxygen saturation using PPG, it is necessary to take measurements at two wavelengths, typically red (660-700 nm) and NIR (800-950 nm) wavelengths. Red and NIR are used because of different absorption of HbO<sub>2</sub> and Hb at these wavelengths; red wavelengths are more sensitive to arterial blood saturation levels [83]. The ratio of PPG amplitude from red light over PPG amplitude from NIR light is used to estimate oxygen saturation [51]. Since PPG amplitude is itself calculated as the ratio between its pulsatile AC component and its non-pulsatile DC component, this method of measuring oxygen saturation is sometimes referred to as the “ratio-of-ratios” method.

Remote measurement of oxygen saturation is complicated because different wavelengths have different penetration depths. Hence, issues arise with calibrating the oxygen saturation measurement. Additionally, contact-based PPG involves a deeper light penetration depth than rPPG, meaning that rPPG measurements are much lower in amplitude than contact PPG, complicating calibration even further [91]. Some success has been found in literature with red and NIR wavelengths, as well as red and green wavelengths [83], suggesting that remote oxygen saturation measurements are feasible. However, low signal strength remains an issue, leaving remote measurement of oxygen saturation as an open-ended research direction that is worth exploring. It is important to note that this is likely to become a long-term project without easy solutions, which is why it is placed on lower priority than previously mentioned tasks in this chapter.



## 4.6 Next Generation Imaging Systems

The stereo triangulation method used by the RealSense cameras is capable of depth imaging at a minimum depth unit of 0.1 mm. However, stereo imaging has inherent drawbacks, such as depth error being proportional to the distance squared. Another project that has been worked on concurrently by other members of the lab is development of a  $4f$  lens system for superior depth resolution. A  $4f$  system is defined as an optical system consisting of two lenses: Lens 1, placed one focal length away from the object of interest, and Lens 2, placed one focal length away from the image plane. The distance between the two lenses is equivalent to two times the focal length. It then follows that the total distance between the object and image plane is four times the focal length, hence a “ $4f$ ” system. Figure 4.3 shows a diagram of the lens placement.

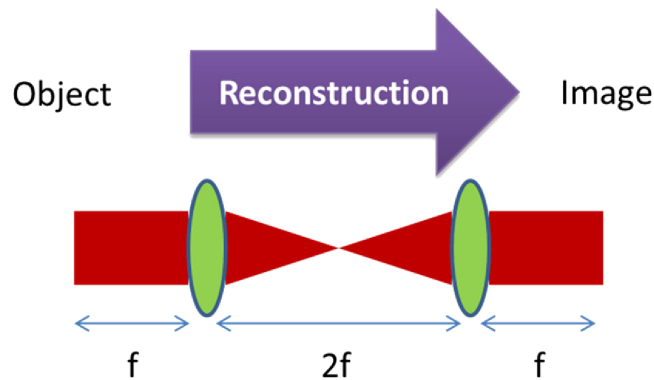


Figure 4.3: Diagram of a  $4f$  system. “ $f$ ” represents one focal length. Taken from [92].

A property of lenses is that if an object is placed one focal length in front of the lens, then the object’s Fourier transform is generated one focal length behind the lens; this point is referred to as the Fourier plane. By using a spatial light modulator (SLM) to place a phase or amplitude mask at the Fourier plane, an optical transformation can be applied to the object. Lens 2 will then apply an inverse Fourier transform to generate a modified image of the object. The response of the system to a point source is referred to as the point spread function (PSF). This technique can be used to create structured light patterns

with a changing appearance based on depth [93]. An example is a double-helix, in which two lobes appear in the PSF that rotate according to distance, as shown in Figure 4.4. This allows sub-micron depth resolution to be achieved.

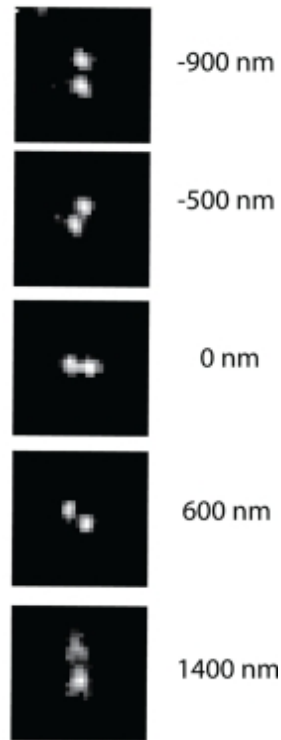


Figure 4.4: Double-helix PSF showing two rotating lobes that change position according to distance. Modified from [93].

While the depth-based measurements described in this thesis does not necessitate the use of sub-micron resolution, better accuracy is desirable for tracking subtle physiological movements such as ballistocardiographic head motion. Furthermore, highly sensitive depth measurements may potentially allow remote visualization of pulsation in arteries as another method of pulse estimation. A future student with an interest in optics and signal processing may be interested in this research direction.

## Chapter 5

# Conclusions

The remote monitoring system described in this thesis presents its novelty in the following areas: (1) utilizing depth information to visualize physiological movements and supplement intensity data, creating a highly redundant system, (2) classifying motion using instantaneous change, SD, and SNR, (3) applying rejection to remove signal segments corrupted by motion, and (4) using single-camera and multi-camera configurations to retrieve physiological information by compensating for subject motion. The existing system is capable of extracting HR and RR from intensity and depth channels, as well as identifying motion-corrupted signal, quantifying the severity of the artifact, and utilizing correction methods to salvage the physiological signal.

In this thesis, it was demonstrated that HR and RR can be reliably and independently estimated from several ROIs using both depth and intensity channels. This creates a redundant system that has numerous independent methods of estimating vital signs in the event that one or more of these methods fails. Bland-Altman analysis was performed to assess the agreement of the remote monitoring system with either a clinical gold standard (HR) or a known ground truth (RR). Results of this analysis are summarized in Table 3.1 and Table 3.2. Best agreement scores were observed when using the cheeks ROI and intensity channel for HR estimation, and the chest ROI and depth channel for RR estima-

tion. These results can be compared to results from other groups in literature that only used intensity measurements, shown in Table 5.1.

Table 5.1: Comparison of the novel system using both intensity and depth to other intensity-based systems in literature. For the novel system, face intensity was used for HR evaluation, whereas chest depth was used for RR evaluation.

|                                 | HR (BPM)               |       |      | RR (BrPM)              |      |       |
|---------------------------------|------------------------|-------|------|------------------------|------|-------|
|                                 | Mean difference (bias) | RPC   | RMSE | Mean difference (bias) | RPC  | RMSE  |
| <b>Novel system</b>             | -1.7                   | 1.4   | 0.6  | 0.05                   | 0.18 | 0.63  |
| <b>2011 Poh et al. [53]</b>     | 0.95                   | 1.62  | 1.24 | 0.12                   | 2.6  | 1.28  |
| <b>2015 Kumar et al. [55]</b>   | -0.02                  | 0.735 | -    | -                      | -    | -     |
| <b>2017 Al-Naji et al. [67]</b> | 1.28                   | 0.615 | 1.32 | 0.195                  | 1.2  | 1.433 |

Studies using depth-based methods for HR and RR estimation are rarer than those using intensity-based methods, and even fewer utilize Bland-Altman analysis. Most implementations use the Microsoft Kinect for depth measurements. Hence, the results of the novel system were compared to Kinect-based systems, as shown in Table 5.2.

Table 5.2: Comparison of the novel system to depth-based systems in literature. Both studies utilized the Microsoft Kinect, which is the most common depth sensor used for vital sign estimations in prior art.

|                                    | HR (BPM)               |      |      | RR (BrPM)              |      |      |
|------------------------------------|------------------------|------|------|------------------------|------|------|
|                                    | Mean difference (bias) | RPC  | RMSE | Mean difference (bias) | RPC  | RMSE |
| <b>Novel system</b>                | -1.7                   | 1.4  | 0.6  | 0.05                   | 0.18 | 0.63 |
| <b>2014 Bernacchia et al. [94]</b> | -0.68                  | 3.03 | -    | 0                      | 0.01 | -    |
| <b>2020 Addison et al. [95]</b>    | -                      | -    | -    | 0.04                   | 1.32 | 0.66 |

Based on these values, the design specification discussed in Section 1.4, namely 3 BPM accuracy for HR and 3 BrPM accuracy for RR, has been met. The accuracy of this system is comparable to that of other systems in literature, but with the addition of a depth channel for added robustness. RR estimated from depth is more accurate compared to intensity-based measurements. However, it is important to note that the analysis performed on our novel system was evaluated only on a single subject. Hence, statistical values shown in Table 5.1 may not fully represent the performance of the system. It is vital that data from a larger cohort of participants is evaluated for a better statistical analysis as future work.

Analysis of data retrieved from volunteer lab members revealed high agreement of HR for the majority of datasets. However, dataset results taken from certain subjects tended to be less accurate. These inaccuracies could occasionally be traced to excessive gross body motion, but there were datasets without excessive motion in which a stable BVP waveform could still not be obtained. It was hypothesized that subject-dependent factors such as skin pigmentation were responsible, as inaccurate estimations were linked to certain subjects, but this could not be validated given the small sample size. This issue must be explored in a larger cohort of participants.

Motion correction was successfully performed on datasets contaminated by motion artifacts. Rejection was applied when severe artifacts were identified via SNR or instantaneous depth/intensity changes, removing signal segments that were corrupted by motion. When motion was present but less severe, compensation was applied. Single-camera compensation using skeleton tracking was successful with non-periodic body motion, but failed when the overall motion was periodic in nature. This led to the development of a two-camera configuration in which a second camera was introduced to act as a reference. This configuration was robust to periodic movement in two dimensions, but failed when the subject was standing at angles where the chest and back were not visible by either camera. As an planned system expansion, a three-camera configuration would have the chest and back always visible by at least one camera. These configurations represent a gradual improvement to motion robustness that is provided primarily due to the combination of intensity and depth channels.

This remote monitoring system has applications in several potential use cases, each with different requirements to better suit the scenario. For example, a patient in a clinical setting exhibits minimal movement, which may allow for increased sensitivity. Conversely, in the case of infectious disease screening in a public environment such as an airport, subjects are more likely to be moving and thus there is a bigger emphasis on motion correction. Future development of the system may allow these aspects to be configured.

This thesis has shown that motion tolerance can be improved through the use

of multiple cameras. However, there are some environments where introducing additional cameras is not desirable, such as in a tightly-packed hospital room, as they can become intrusive and can raise concerns about privacy. In these cases, a single camera may be preferred. While this means that motion compensation techniques cannot be applied, there may be potential for other algorithms to be explored that can correct for motion in a single-camera scenario.

Some limitations of the system that could be improved in future work were discussed in Chapter 4. The system has not yet been evaluated in participants on scenarios such as exercise and supine position. Furthermore, several potential hardware and software improvements were discussed. While a two-camera system provides robustness to gross body movement in two axes, it is not robust to rotational motion, which necessitates an expansion to a three-camera system. Each of these limitations can and should be addressed in future research projects.

The progress described in this thesis paves the way for several potential advancements and projects. It is the author's hope that future students continue to build upon and improve the remote monitoring system to reach its full potential.

# Chapter 6

## Appendix

### 6.1 Dash Monitor Data Extraction

#### 6.1.1 Numerical Values

An auxiliary (aux) RJ-45 port at the back of the monitor provides a serial connection that enables numerical vital values to be exported to the PC. A custom cable was wired to connect the Dash’s aux port to the RS-232 serial port on the PC. A third-party program, “VitalSignsCapture” developed by xeonfusion on SourceForge [96], was used to parse the Dash’s serial output into a timestamped text file. Using the highest polling frequency, an entry for numerical HR, RR, and SpO<sub>2</sub> could be generated every 2 seconds. The data is logged to a CSV file.

#### 6.1.2 ECG Waveform

ECG waveform is extracted using the Defib Sync port on the back of the Dash monitor; pins 3 and 7 of the port provide an analog voltage that represents the ECG waveform. An Arduino Nano was configured to be an analog-to-digital converter (ADC), enabling ECG to be recorded using a voltage divider as shown in Figure 6.1. As the Arduino is only capable of reading positive voltages, the

voltage divider serves to boost the ECG analog voltages above zero to prevent the waveform from being clipped. The resolution of the Arduino ADC is 10 bits, and the serial baud rate was set to 115,200 bits per second (bps). The application PuTTY was used as a serial monitor with automatic logging to a text file.

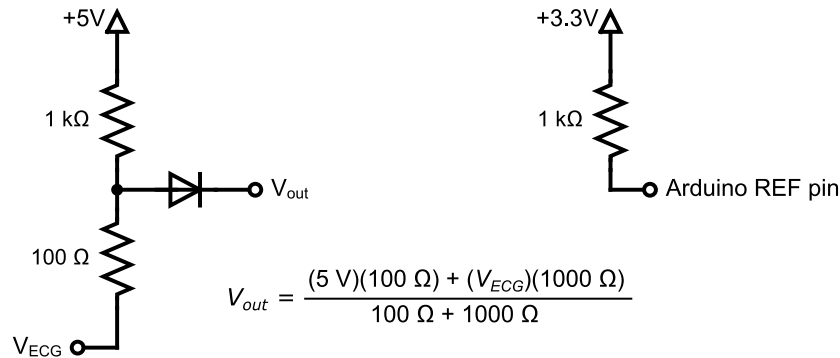


Figure 6.1: Voltage divider from the Dash ECG analog ( $V_{ECG}$ ) to the Arduino's analog input pin ( $V_{out}$ ).

## 6.2 USB Bandwidth Limitations

Connecting multiple RealSense cameras may create issues with USB bandwidth. Each camera was connected to the PC using its own dedicated bus channel with up to 5 Gbps of bandwidth. A dedicated PCIe card was installed to facilitate a three-camera setup.

## 6.3 RealSense Camera Hardware Triggering

In a multi-camera configuration, it is necessary for each camera to capture images synchronously. RealSense cameras have a master-slave feature where one camera can send a triggering pulse to the others. A port in each camera allows them to be connected together. Pin 5 is the sync port and pin 9 is the ground. Cameras were connected in a multi-drop configuration using jumper



wires for easy connections. The connector part number is ASSHSSH28K152 and the housing part number is SHR-09V-S-B.

# Bibliography

- [1] United Nations Department of Economic and Social Affairs, “World population ageing 2019”, New York NY, Tech. Rep., 2020.
- [2] J. P. L. Leenen, C. Leerentveld, J. D. van Dijk, H. L. van Westreenen, L. Schoonhoven, and G. A. Patijn, “Current evidence for continuous vital signs monitoring by wearable wireless devices in hospitalized adults: Systematic review”, *J Med Internet Res*, vol. 22, no. 6, e18636, Jun. 2020.
- [3] T. Jakobson, J. Karjagin, L. Vipp, M. Padar, A. H. Parik, L. Starkopf, H. Kern, O. Tammik, and J. Starkopf, “Postoperative complications and mortality after major gastrointestinal surgery”, *Medicina (Kaunas)*, vol. 50, no. 2, pp. 111–117, 2014.
- [4] A. A. Ghaferi, J. D. Birkmeyer, and J. B. Dimick, “Variation in hospital mortality associated with inpatient surgery”, *N. Engl. J. Med.*, vol. 361, no. 14, pp. 1368–1375, Oct. 2009.
- [5] C. L. Downey, W. Tahir, R. Randell, J. M. Brown, and D. G. Jayne, “Strengths and limitations of early warning scores: A systematic review and narrative synthesis”, *Int J Nurs Stud*, vol. 76, pp. 106–119, Nov. 2017.
- [6] A. H. Taenzer, J. Pyke, M. D. Herrick, T. M. Dodds, and S. P. McGrath, “A comparison of oxygen saturation data in inpatients with low oxygen saturation using automated continuous monitoring and intermittent manual data charting”, *Anesthesia and analgesia*, vol. 118, no. 2, pp. 326–331, Feb. 2014, ISSN: 0003-2999.
- [7] D. G. Mason and J. Stewart, “Emergency admissions: A journey in the right direction?”, in. 2007.

- [8] F. Zhao, M. Li, and J. Z. Tsien, “Technology platforms for remote monitoring of vital signs in the new era of telemedicine”, *Expert Review of Medical Devices*, vol. 12, no. 4, pp. 411–429, 2015.
- [9] C. L. Downey, J. M. Brown, D. G. Jayne, and R. Randell, “Patient attitudes towards remote continuous vital signs monitoring on general surgery wards: An interview study”, *Int J Med Inform*, vol. 114, pp. 52–56, Jun. 2018.
- [10] S. T. Peek, E. J. Wouters, J. van Hoof, K. G. Luijkx, H. R. Boeije, and H. J. Vrijhoef, “Factors influencing acceptance of technology for aging in place: A systematic review”, *International Journal of Medical Informatics*, vol. 83, no. 4, pp. 235–248, 2014.
- [11] J. Buekers, J. Theunis, P. De Boever, A. W. Vaes, M. Koopman, E. V. Janssen, E. F. Wouters, M. A. Spruit, and J. M. Aerts, “Wearable Finger Pulse Oximetry for Continuous Oxygen Saturation Measurements During Daily Home Routines of Patients With Chronic Obstructive Pulmonary Disease (COPD) Over One Week: Observational Study”, *JMIR Mhealth Uhealth*, vol. 7, no. 6, e12866, Jun. 2019.
- [12] S. Sendelbach and M. Funk, “Alarm fatigue: a patient safety concern”, *AACN Adv Crit Care*, vol. 24, no. 4, pp. 378–386, 2013.
- [13] *How to place ecg electrodes*, Oct. 2018. [Online]. Available: <https://www.usamedicalsurgical.com/blog/how-to-place-ecg-electrodes/>.
- [14] A. Gotter, *Pulse oximetry: Uses, readings, and how it works*, Aug. 2017. [Online]. Available: <https://www.healthline.com/health/pulse-oximetry>.
- [15] M. R. Karaoğuz, E. Yurtseven, G. Aslan, B. G. Deliormanlı, Ö. Adıgüzel, M. Gönen, K. M. Li, and E. N. Yılmaz, “The quality of ECG data acquisition, and diagnostic performance of a novel adhesive patch for ambulatory cardiac rhythm monitoring in arrhythmia detection”, *J Electrocardiol*, vol. 54, pp. 28–35, 2019.
- [16] A. C. Deswysen, E. Zimerson, A. Goossens, M. Bruze, and M. Baeck, “Allergic contact dermatitis caused by self-adhesive electrocardiography

- electrodes in an infant”, *Contact Derm.*, vol. 69, no. 6, pp. 379–381, Dec. 2013.
- [17] J. Evans, A. Papadopoulos, C. T. Silvers, N. Charness, W. R. Boot, L. Schlachta-Fairchild, C. Crump, M. Martinez, and C. B. Ent, “Remote Health Monitoring for Older Adults and Those with Heart Failure: Adherence and System Usability”, *Telemed J E Health*, vol. 22, no. 6, pp. 480–488, Jun. 2016.
- [18] V. Nangalia, D. R. Prytherch, and G. B. Smith, “Health technology assessment review: remote monitoring of vital signs—current status and future challenges”, *Crit Care*, vol. 14, no. 5, p. 233, 2010.
- [19] A. Phillips, “Health status differentials across rural and remote Australia”, *Aust J Rural Health*, vol. 17, no. 1, pp. 2–9, Feb. 2009.
- [20] D. Goodridge and D. Marciniuk, “Rural and remote care: Overcoming the challenges of distance”, *Chron Respir Dis*, vol. 13, no. 2, pp. 192–203, May 2016.
- [21] Centers for Disease Control and Prevention (CDC). (2020). Symptoms of coronavirus — cdc, [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html> (visited on 05/13/2020).
- [22] GE Healthcare. (2020). Cardiac complications of covid-19: Signs to watch for on the ecg, [Online]. Available: <https://www.gehealthcare.ca/en-CA/article/cardiac-complications-of-covid-19-signs-to-watch-for-on-the-ecg> (visited on 05/19/2020).
- [23] R. Sinhal, K. Singh, and M. M. Raghuwanshi, “An overview of remote photoplethysmography methods for vital sign monitoring”, in *Computer Vision and Machine Intelligence in Medical Image Analysis*, M. Gupta, D. Konar, S. Bhattacharyya, and S. Biswas, Eds., Singapore: Springer Singapore, 2020, pp. 21–31.
- [24] C. Perret-Guillaume, L. Joly, and A. Benetos, “Heart rate as a risk factor for cardiovascular disease”, *Prog Cardiovasc Dis*, vol. 52, no. 1, pp. 6–10, 2009.

- [25] S. P. Whelton, R. Blankstein, M. H. Al-Mallah, J. A. Lima, D. A. Bluemke, W. G. Hundley, J. F. Polak, R. S. Blumenthal, K. Nasir, and M. J. Blaha, “Association of resting heart rate with carotid and aortic arterial stiffness: multi-ethnic study of atherosclerosis”, *Hypertension*, vol. 62, no. 3, pp. 477–484, Sep. 2013.
- [26] R. Gordan, J. K. Gwathmey, and L. H. Xie, “Autonomic and endocrine control of cardiovascular function”, *World J Cardiol*, vol. 7, no. 4, pp. 204–214, Apr. 2015.
- [27] I. Cygankiewicz and W. Zareba, “Heart rate variability”, *Handb Clin Neurol*, vol. 117, pp. 379–393, 2013.
- [28] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, “Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis”, *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 2015.
- [29] J. Wei, H. Luo, S. J. Wu, P. P. Zheng, G. Fu, and K. Lee, “Transdermal Optical Imaging Reveal Basal Stress via Heart Rate Variability Analysis: A Novel Methodology Comparable to Electrocardiography”, *Front Psychol*, vol. 9, p. 98, 2018.
- [30] B. H. Cuthbertson, M. Boroujerdi, L. McKie, L. Aucott, and G. Prescott, “Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient?”, *Crit. Care Med.*, vol. 35, no. 2, pp. 402–409, Feb. 2007.
- [31] T. Flenady, T. Dwyer, and J. Applegarth, “Explaining transgression in respiratory rate observation methods in the emergency department: A classic grounded theory analysis”, *Int J Nurs Stud*, vol. 74, pp. 67–75, Sep. 2017.
- [32] C. P. Subbe and S. Kinsella, “Continuous Monitoring of Respiratory Rate in Emergency Admissions: Evaluation of the RespiraSense Sensor in Acute Care Compared to the Industry Standard and Gold Standard”, *Sensors (Basel)*, vol. 18, no. 8, Aug. 2018.

- [33] S. J. Barker, K. K. Tremper, and D. M. Gamel, “A clinical comparison of transcutaneous PO<sub>2</sub> and pulse oximetry in the operating room”, *Anesth. Analg.*, vol. 65, no. 7, pp. 805–808, Jul. 1986.
- [34] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, “Respiratory rate: the neglected vital sign”, *Med. J. Aust.*, vol. 188, no. 11, pp. 657–659, Jun. 2008.
- [35] C. Massaroni, A. Nicol?, D. Lo Presti, M. Sacchetti, S. Silvestri, and E. Schena, “Contact-Based Methods for Measuring Respiratory Rate”, *Sensors (Basel)*, vol. 19, no. 4, Feb. 2019.
- [36] E. Fan, D. Brodie, and A. S. Slutsky, “Acute Respiratory Distress Syndrome: Advances in Diagnosis and Treatment”, *JAMA*, vol. 319, no. 7, pp. 698–710, Feb. 2018. eprint: [https://jamanetwork.com/journals/jama/articlepdf/2673154/jama\\_fan\\_2018\\_rv\\_170010.pdf](https://jamanetwork.com/journals/jama/articlepdf/2673154/jama_fan_2018_rv_170010.pdf).
- [37] X. Ma, Z. Q. Yue, Z. Q. Gong, H. Zhang, N. Y. Duan, Y. T. Shi, G. X. Wei, and Y. F. Li, “The Effect of Diaphragmatic Breathing on Attention, Negative Affect and Stress in Healthy Adults”, *Front Psychol*, vol. 8, p. 874, 2017.
- [38] K. Philip, R. Richardson, and M. Cohen, “Staff perceptions of respiratory rate measurement in a general hospital”, *British Journal of Nursing*, vol. 22, no. 10, pp. 570–574, 2013.
- [39] C. Rotariu and V. Manta, “Wireless system for remote monitoring of oxygen saturation and heart rate”, in *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2012, pp. 193–196.
- [40] A. R. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. C. Frise, P. A. Robbins, and L. Tarassenko, “Non-contact measurement of oxygen saturation with an RGB camera”, *Biomed Opt Express*, vol. 6, no. 9, pp. 3320–3338, Sep. 2015.
- [41] C. D. Hanning and J. M. Alexander-Williams, “Pulse oximetry: a practical review”, *BMJ*, vol. 311, no. 7001, pp. 367–370, Aug. 1995.
- [42] B. B. Hafen and S. Sharma, “Oxygen saturation”, in. StatPearls Publishing, 2020.

- [43] A. V. Moço, S. Stuijk, and G. de Haan, “New insights into the origin of remote ppg signals in visible light and infrared”, *Scientific Reports*, vol. 8, no. 1, p. 8501, May 2018.
- [44] R. R. Anderson and J. A. Parrish, “The optics of human skin”, *J. Invest. Dermatol.*, vol. 77, no. 1, pp. 13–19, Jul. 1981.
- [45] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic Principles of Remote PPG”, *IEEE Trans Biomed Eng*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [46] M. van Gastel, S. Stuijk, and G. de Haan, “New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring”, *Sci Rep*, vol. 6, p. 38 609, Dec. 2016.
- [47] T. Lister, P. A. Wright, and P. H. Chappell, “Optical properties of human skin”, *Journal of Biomedical Optics*, vol. 17, no. 9, pp. 1–15, 2012.
- [48] V. Ntziachristos, “Going deeper than microscopy: The optical imaging frontier in biology”, *Nature Methods*, vol. 7, no. 8, pp. 603–614, Aug. 2010.
- [49] L. F. C. Martinez, G. Paez, and M. Strojnik, “Optimal wavelength selection for noncontact reflection photoplethysmography”, in *22nd Congress of the International Commission for Optics: Light for the Development of the World*, International Society for Optics and Photonics, vol. 8011, SPIE, 2011, pp. 2388–2394.
- [50] C. Ash, M. Dubec, K. Donne, and T. Bashford, “Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods”, *Lasers Med Sci*, vol. 32, no. 8, pp. 1909–1918, Nov. 2017.
- [51] P. A. Kyriacou, “Pulse oximetry in the oesophagus”, *Physiological Measurement*, vol. 27, no. 1, R1–R35, Nov. 2005.
- [52] M. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.”, *Opt. Express*, vol. 18, no. 10, pp. 10 762–10 774, May 2010.
- [53] —, “Advancements in noncontact, multiparameter physiological measurements using a webcam”, *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7–11, 2011.

- [54] D. McDuff, S. Gontarek, and R. W. Picard, “Improvements in remote cardiopulmonary measurement using a five band digital camera”, *IEEE Trans Biomed Eng*, vol. 61, no. 10, pp. 2593–2601, Oct. 2014.
- [55] M. Kumar, A. Veeraraghavan, and A. Sabharwal, “DistancePPG: Robust non-contact vital signs monitoring using a camera”, *Biomed Opt Express*, vol. 6, no. 5, pp. 1565–1588, May 2015.
- [56] W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino, and G. A. Dumont, “Estimation of respiratory rate from photoplethysmographic imaging videos compared to pulse oximetry”, *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1331–1338, 2015.
- [57] M. van Gastel, S. Stuijk, and G. de Haan, “Robust respiration detection from remote photoplethysmography”, *Biomed Opt Express*, vol. 7, no. 12, pp. 4941–4957, Dec. 2016.
- [58] G. de Haan and V. Jeanne, “Robust pulse rate from chrominance-based rppg”, *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878–2886, 2013.
- [59] G. de Haan and A. van Leest, “Improved motion robustness of remote-ppg by using the blood volume pulse signature”, *Physiological Measurement*, vol. 35, pp. 1913–1926, 2014.
- [60] X. Li, J. Chen, G. Zhao, and M. Pietikäinen, “Remote heart rate measurement from face videos under realistic situations”, in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 4264–4271.
- [61] L. Giovangrandi, O. T. Inan, R. M. Wiard, M. Etemadi, and G. T. Kovacs, “Ballistocardiography—a method worth revisiting”, *Conf Proc IEEE Eng Med Biol Soc*, vol. 2011, pp. 4279–4282, 2011.
- [62] G. Bonmassar, P. L. Purdon, I. P. J?skel?inen, K. Chiappa, V. Solo, E. N. Brown, and J. W. Belliveau, “Motion and ballistocardiogram artifact removal for interleaved recording of EEG and EPs during MRI”, *Neuroimage*, vol. 16, no. 4, pp. 1127–1141, Aug. 2002.
- [63] G. Balakrishnan, F. Durand, and J. Guttag, “Detecting pulse from head motions in video”, in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3430–3437.



- [64] C. S. Kim, S. L. Ober, M. S. McMurtry, B. A. Finegan, O. T. Inan, R. Mukkamala, and J. O. Hahn, “Ballistocardiogram: Mechanism and Potential for Unobtrusive Cardiovascular Health Monitoring”, *Sci Rep*, vol. 6, p. 31 297, Aug. 2016.
- [65] D. Da He, E. S. Winokur, and C. G. Sodini, “A continuous, wearable, and wireless heart monitor using head ballistocardiogram (bcg) and head electrocardiogram (ecg)”, in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 4729–4732.
- [66] D. Shao, F. Tsow, C. Liu, Y. Yang, and N. Tao, “Simultaneous Monitoring of Ballistocardiogram and Photoplethysmogram Using a Camera”, *IEEE Trans Biomed Eng*, vol. 64, no. 5, pp. 1003–1010, May 2017.
- [67] A. Al-Naji, K. Gibson, and J. Chahl, “Remote sensing of physiological signs using a machine vision system”, *J Med Eng Technol*, vol. 41, no. 5, pp. 396–405, Jul. 2017.
- [68] Intel Corporation. (2019). Beginner’s guide to depth (updated), [Online]. Available: <https://www.intelrealsense.com/beginners-guide-to-depth/>.
- [69] A. L. Diehm, M. Hammer, M. Hebel, and M. Arens, “Mitigation of crosstalk effects in multi-LiDAR configurations”, in *Electro-Optical Remote Sensing XII*, G. Kamerman and O. Steinvall, Eds., International Society for Optics and Photonics, vol. 10796, SPIE, 2018, pp. 13–24.
- [70] H. Sarbolandi, D. Lefloch, and A. Kolb, “Kinect range sensing: Structured-light versus time-of-flight kinect”, *Computer Vision and Image Understanding*, vol. 139, pp. 1–20, 2015, ISSN: 1077-3142.
- [71] M. C. Yu, J. L. Liou, S. W. Kuo, M. S. Lee, and Y. P. Hung, “Noncontact respiratory measurement of volume change using depth camera”, *Conf Proc IEEE Eng Med Biol Soc*, vol. 2012, pp. 2371–2374, 2012.
- [72] F. Centonze, M. Schätz, A. Procházka, J. Kuchyňka, O. Vyšata, P. Cejnar, and M. Vališ, “Feature extraction using ms kinect and data fusion in analysis of sleep disorders”, in *2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, 2015, pp. 1–5.

- [73] C. Yang, G. Cheung, and V. Stankovic, “Estimating heart rate and rhythm via 3d motion tracking in depth video”, *IEEE Transactions on Multimedia*, vol. 19, no. 7, pp. 1625–1636, 2017.
- [74] Masimo, *Mightysat rx fingertip pulse oximeter*, Brochure, 2019. [Online]. Available: [https://www.masimo.com/siteassets/us/documents/pdf/plm-11294e\\_brochure\\_mightysat\\_rx\\_us.pdf](https://www.masimo.com/siteassets/us/documents/pdf/plm-11294e_brochure_mightysat_rx_us.pdf).
- [75] Intel Corporation. (2020). Stereo depth - intel realsense depth and tracking cameras, [Online]. Available: <https://www.intelrealsense.com/stereo-depth/>.
- [76] Baumer. (2020). Rolling shutter, global shutter – two principles of exposure [an201906], [Online]. Available: <https://www.baumer.com/ca/en/service-support/know-how/technical-information-industrial-cameras/rolling-shutter-global-shutter-two-principles-of-exposure-/a/rolling-shutter-global-shutter>.
- [77] *Intel realsense d400 series product family datasheet*, Rev. 7, Intel Corporation, Oct. 2019.
- [78] J. Dames. (2018). 12-lead ecg placement – a basic guide, [Online]. Available: <https://www.aedsuperstore.com/resources/12-lead-ecg-placement/>.
- [79] R. Klabunde. (2017). Electrocardiogram standard limb leads (bipolar), [Online]. Available: <https://www.cvphysiology.com/Arrhythmias/A013a>.
- [80] cubemos, *Skeleton tracking sdk*. [Online]. Available: <https://www.cubemos.com/skeleton-tracking-sdk>.
- [81] A. A. Kamshilin, E. Nippolainen, I. S. Sidorov, P. V. Vasilev, N. P. Erofeev, N. P. Podolian, and R. V. Romashko, “A new look at the essence of the imaging photoplethysmography”, *Sci Rep*, vol. 5, p. 10 494, May 2015.
- [82] D. Pollreisz and N. TaheriNejad, “Detection and removal of motion artifacts in ppg signals”, *Mobile Networks and Applications*, Aug. 2019.
- [83] A. Moço and W. Verkruyse, “Pulse oximetry based on photoplethysmography imaging with red and green light : Calibratability and challenges”, *J Clin Monit Comput*, Jan. 2020.

- [84] J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement”, *Lancet*, vol. 1, no. 8476, pp. 307–310, Feb. 1986.
- [85] D. Giavarina, “Understanding Bland Altman analysis”, *Biochem Med (Zagreb)*, vol. 25, no. 2, pp. 141–151, 2015.
- [86] J. M. Bland and D. G. Altman, “Comparing methods of measurement: why plotting difference against standard method is misleading”, *Lancet*, vol. 346, no. 8982, pp. 1085–1087, Oct. 1995.
- [87] R. Klein, *Bland-altman and correlation plot*, Retrieved September 19, 2020, 2019. [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot>.
- [88] G. Shafiq and K. C. Veluvolu, “Surface chest motion decomposition for cardiovascular monitoring”, *Sci Rep*, vol. 4, p. 5093, May 2014.
- [89] M. Elgendi, “On the analysis of fingertip photoplethysmogram signals”, *Curr Cardiol Rev*, vol. 8, no. 1, pp. 14–25, Feb. 2012.
- [90] A. Moço, S. Stuijk, and G. de Haan, “Posture effects on the calibratability of remote pulse oximetry in visible light”, *Physiol Meas*, vol. 40, no. 3, p. 035005, Apr. 2019.
- [91] W. Verkruysse, M. Bartula, E. Bresch, M. Rocque, M. Meftah, and I. Kirenko, “Calibration of Contactless Pulse Oximetry”, *Anesth. Analg.*, vol. 124, no. 1, pp. 136–145, Jan. 2017.
- [92] W. Han, W. Cheng, and Q. Zhan, “Design and alignment strategies of 4f systems used in the vectorial optical field generator”, *Appl. Opt.*, vol. 54, no. 9, pp. 2275–2278, Mar. 2015.
- [93] S. R. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. E. Moerner, “Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function”, *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 9, pp. 2995–2999, Mar. 2009.

- [94] N. Bernacchia, L. Scalise, L. Casacanditella, I. Ercoli, P. Marchionni, and E. P. Tomasini, “Non contact measurement of heart and respiration rates based on kinect<sup>TM</sup>”, in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2014, pp. 1–5.
- [95] P. S. Addison, P. Smit, D. Jacquiel, and U. R. Borg, “Continuous respiratory rate monitoring during an acute hypoxic challenge using a depth sensing camera”, *J Clin Monit Comput*, vol. 34, no. 5, pp. 1025–1033, Oct. 2020.
- [96] J. George, *Vitalsigncapture*, <https://sourceforge.net/projects/vscapture/>, 2020.