A Machine Learning Method to Improve Non-Contact Heart Rate Monitoring Using an RGB Camera

Hamideh Ghanadian

Thesis submitted in partial fulfillment of the requirements for the
MASTER OF APPLIED SIENNCE
IN ELECTIRCAL AND COMPUTER ENGINEERING

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and engineering
University of Ottawa
Ottawa, Canada

Abstract

Recording and monitoring vital signs is an essential aspect of home-based healthcare. Using contact sensors to record physiological signals can cause discomfort to patients, especially after prolonged use. Hence, remote physiological measurement approaches have attracted considerable attention as they do not require physical contact with the patient's skin. Several studies proposed techniques to measure Heart Rate (HR) and Heart Rate Variability (HRV) by detecting the Blood Volume Pulse (BVP) from human facial video recordings while the subject is in a resting condition. In this thesis, we focus on the measurement of HR.

We adopt an algorithm that uses the Independent Component Analysis (ICA) to separate the source (physiological) signal from noise in the RGB channels of a facial video. We generalize existing methods to support subject movement during video recording. When a subject is moving, the face may be turned away from the camera. We utilize multiple cameras to enable the algorithm to monitor the vital sign continuously, even if the subject leaves the frame or turns away from a subset of the system's cameras. Furthermore, we improve the accuracy of existing methods by implementing a light equalization scheme to reduce the effect of shadows and unequal facial light on the HR estimation, a machine learning method to select the most accurate channel outputted by the ICA module, and a regression technique to adjust the initial HR estimate. We systematically test our method on eleven subjects using four cameras. The proposed method decreases the RMSE by 27% compared to the state of the art in the rest condition. When the subject is in motion, the proposed method achieves a RMSE of 1.12 bpm.

Acknowledgments

I would like to extend my greatest appreciation to Dr. Hussein Al Osman for his tremendous support and encouragement he has shown during the last two years. He has always listened to me patiently and enthusiastically and guided me toward the best solutions by asking deep questions. He taught me how to do academic research. He has read my drafts carefully, and has generously provided me with detailed feedback on my research paper and on my thesis. He has been honest enough to point out my weaknesses to me and to push me to work harder, and caring and supportive enough to see my potential even when I could not.

I am extremely grateful to my mother, Sedigheh, and my father, Ahmad, for believing in me, even when I didn't believe in myself, for always encouraging me to be the best person I can be. They have been pillars of support, guidance and love in my life since the day I was born. I cannot express how important their presence has been throughout my life. I also thank my brothers, Ali, Hossein and Abolfazl for being part of my foundation, for all of the support they've provided me over the last several years, and for all of the incredible strength they've forced me to see in myself.

I would like to thank my friends for their participation and help in my research. I appreciate your cooperation, support and simply being wonderful friends.

I wish to give my heartfelt thanks to my husband and my best friend, Mohammadreza, for his unconditional love, patience, and continual support during the past five years. He has been always my joy, my inspiration, and my guiding light. It would have been impossible for me to complete this thesis without him in my life.

And finally, "All the praises and thanks are to Allah, the Lord of all the worlds." (Quran 1:2)

Dedication

To My Parents

And Mohammadreza

Table of Contents

Table of Figures

Table of Tables

Glossary of Terms

HR: Heart Rate

HRV: Heart Rate Variability

ECG: Electrocardiography

PPG: Photo-plethysmography

EVM: Eulerian video magnification

BVP: Blood volume pulse

RMSE: Root Mean Square of Error

RR Interval: The time between two consecutive Heart beats

LF: Low frequency

HF: High frequency

PSD: Power spectral density

FFT: Fast Fourier Transform

FIR: Finite impulse response

IIR: Infinite impulse response

RGB: Red, Green, Blue

HSL: Hue, Saturation, Lightness

RGBCO: Red, Green, Blue, Cyan, Orange

OpenCV: Open Computer Vision

ROI: Region of Interest

JADE: Joint Approximate Diagonalization of Eigen matrices

FPS: Frame per second

Chapter 1. Introduction

Vital signs, including Heart Rate (HR) and other physiological signs such as respiratory rate and Heart Rate Variability (HRV), are important indicators of patient health. They provide critically needed information to make life-saving decisions. Continuous monitoring of HR and HRV is becoming an important aspect of home-based healthcare. In addition to physical health assessment, HR and HRV measurement can be employed for psychological health monitoring [1]. For instance, mental stress can be effectively assessed using HR and HRV measures [2].

1.1. Problem Statement

Methods to measure HR are classified into two categories: Contact and Non-Contact based. Contact based measurement involves the use of sensors attached onto the patient's skin. There are several limitations in contact based-measurement. These sensors limit the movement of the patient due to their wiring system. Moreover, in some cases, connecting the sensor to the skin is not possible due to the patient's condition. For example, monitoring vital signs in the Neonatal Intensive Care Unit (NICU) requires the use of adhesives to attach sensors to the skin of pre-term infants which might result in pain and skin irritation [3]. However, non-contact measuring methods address effectively theses issues. Optical non-contact methods estimate a photoplethysmography (PPG) signal without directly being placed on the subject's body. A growing body of literature investigates the possibility of using camera-based contactless techniques, better known as camera based non-contact PPG, to measure HR. Several attempts have been made to increase the reliability and accuracy of these measurement techniques. These methods are based on the principle that the light absorption properties of blood differ from the surrounding skin. Hence, when the arterioles'

blood volume increases, the skin color varies. These changes can be recorded by a standard RGB video camera. However, motion artifacts still restrict the use of remote HR measurement schemes during movement.

1.2. Goals and Objectives

We have developed a method to extract HR from a facial video of a subject in movement. Hence, the objective of this study is to assess the accuracy of the latter method by comparing the physiological signals extracted from a facial video with those collected via an electrocardiography (ECG) sensor. We note that previous video-based HR measurement studies suffer from several limitations:

- First, the subjects were permitted only slight movements[4][5][6].
- Second, the measured PPG signal was vulnerable to noise caused by lighting condition changes [7].
- Third, since the implemented ICA module produces three signals as output, mostly unproven heuristic methods were used to select the signal best reflecting the BVP information [4].

1.3. Motivation

Contactless techniques to monitor vital signs have a variety of applications. For example, in long term HR monitoring [8], contact-based techniques such as ECG may lead to skin irritations as the electrodes are replaced daily over roughly the same area. Hence, contactless techniques are well suited for this scenario. In addition, there are situations where the patient might not be interested

in keeping skin-irritating sensors attached. Prisoners on suicide watch who are monitored to ensure that they do not self-harm fall under this category. Hence, we recognize a need for the development of effective mechanisms to monitor vital signs remotely.

1.4. Contributions

We summarize our contributions as follows:

1. We propose a light equalization scheme to minimize the negative effect of noise generated from the fluctuation of light on HR estimation;
2. We develop a ML algorithm to select the PPG information carrying component after performing ICA on the RGB channels
3. We propose a linear regression model to improve the accuracy of the HR estimation obtained through the ICA.
4. We propose the use of multiple cameras to improve the estimation of HR of subjects roaming an environment.

1.5. Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2 discusses relevant background and related work. We summarize previous literature related to the contactless approaches to monitor physiological signals, more specifically the HR signal.

Chapter 3 describes the proposed method to estimate the HR signal with a single and multiple cameras. Furthermore, we describe the proposed Light Equalization, Machine Learning, and Linear Regression schemes.

Chapter 4 presents the results for the validation and lab experiments we conducted to assess the effectiveness of the proposed method. Moreover, we discuss the results and present our conclusions.

Chapter 5 summarizes the thesis findings and provides insights into future work.

Chapter 2. Background and Related Works

In this chapter, we provide a brief explanation of the relevant concepts in image processing, physiological signals, and machine learning. Moreover, we present existing remote HR measurement solutions and summarize them in Table 1 and Table 2.

2.1. HR Measurement

HR, defined as the number of heart beats per unit of time, is a significant measure for assessing cardiac well-being [9].Typically, it is measured as the number of heart contractions per minute (bpm). Its regular measurement is crucial in assessing and monitoring a myriad of health conditions. The regularity of the heart pulse can change during the lifetime, especially as people get older. Monitoring these changes is essential in preventing health issues related to the heart [10][11].

HR can be measured by finding the pulse of the heart. The process of measuring HR can be classified into either contact or non-contact method.

2.1.1 Contact Measurement

Contact measurement methods are the most widely used for extracting HR signals, and often involve the use of sensors placed directly on the subject's body. ECG and PPG sensors are two examples of non-invasive sensors to measure cardiovascular signs of health. The ECG sensor records the electrical activity of the heart muscle using small electrodes attached to several parts of body [12]. PPG is a low-cost and non-invasive means of sensing the blood volume pulse (BVP)

from variations in the reflected light through the tissue [13]. It is usually obtained through a pulse oximeter that is attached to the finger to measure the oxygen saturation level in the blood.

2.1.2. Non-Contact Measurement

To the best of our knowledge, among the various attempts towards measuring human vital signs remotely, Chen et al. [14] were the first to introduce remote measurement of HR. They use a Doppler radar as an active sensor in a Microwave system [15][16]. The radar detects heart and breathing rates from almost a 30 meters distance [17][18]. The radar projects a low-intensity microwave signal. The signal is reflected onto the subject and received back by the radar. By analyzing the reflected beam, they can measure the motion of the chest wall and abdomen and, consequently, the breathing rate of the subject [19]. Greneker [20] presents another method that uses a microwave Doppler system equipped with a charge-coupled device camera and a 0.6 meter dish to measure HR and breathing rate. The system requires an operator to aim a beam onto the thorax area of the subject's chest to make the measurement.

Bakhtiari et al. [21] develop a millimetre wave radar (MMW) sensor for the remote monitoring of human vital signs over a long distance. The waveband they select provides low propagation losses, which enables them to monitor vital signs over a long distance. They extract and measure the modulation induced by a signal reflected off a moving object's surface. For instance, breathing and heartbeats cause slight motion on the chest's surface that is detectable by the MMW sensors. They effectively extract vital sign signals using the MMW sensor for stationary subjects. However, the method is ineffective in the presence of mere subtle movements.

Huang et al. [22] measure ECG using electrodes attached to the backrest of a sofa. The authors classify this method as a non-contact technique as the electrodes are separated from the body. This system can measure HR and HRV parameters under various living conditions.

Using thermal cameras creates new opportunities to realize non-contact vital signs measurement techniques. Nanfei et al. [23] demonstrate the first remote passive cardiac pulse measurement through a thermal imaging system. This method is based on the computation of the frequency of pulses derived from the temperature changes on the vessel modulated by pulsatile blood flow. Furthermore, Garbey et al. [24] use a thermal imaging system to acquire the thermal signals emitted from major superficial vessels. They calculate the frequency of pulsative blood flow to extract the HR signal. They assume that the heart pulse frequency is dominant in the thermal field of a superficial vessel. Hence, recovering the frequency of the component signal with the highest energy component leads to acquiring the HR signal. However, this approach has some limitations. It requires the use of an expensive thermal imaging camera. In addition, it can yield inaccurate results due to the variation in room temperature [25]

As opposed to the above described methods, RGB video-based methods measure variations in the subject's skin color to extract a PPG signal. PPG is a non-invasive technique to sense the BVP through the light reflected from the skin. As blood absorbs more light compared to the surrounding tissue, changes in the blood volume affects the reflectance of the light [4].Takano and Ohta [26] measure HR using a digital camera in an environment with ambient light. They calculate the average brightness of time-lapse skin images recorded by a charge-coupled device camera. They then apply auto-regressive spectral analysis on the images to calculate the HR.

Balakrishnan et al. [27] estimate the HR from video recordings based on the human head motion caused by Newtonian reaction to the induction of the blood at each beat. They use PCA to decompose the extracted video signals into a set of component motions. Then, they select the component whose temporal power spectrum best matches to the pulse shape. One of the difficulties with this method is discriminating between the movement of the subjects and the motion of the head during the video recording.

Sun et al. [28] investigate the feasibility of remote monitoring of HRV and other physiological signals. They use palm images to extract the HRV. They employ a CMOS camera-based imaging system for capturing the PPG through palm images.

MacDuff et al. [29] utilize a five band camera (Red, Orange, Green, Cyan, and Blue) instead of RGB camera to measure HR and HRV from facial images. Non-standard camera such as the five-band camera is more expensive than the conventional RGB camera. They use the LEAR facial landmark detector [30] to find the coordinates of the face landmarks in each frame. Color channels spatially averaged and then presented to the ICA. Their results indicate that the combination of Cyan, Green, and Orange color channels boost the performance of the measurements of vital signs.

Bousefsaf et al. [31] capture the RGB face images and convert them to the LVU color space. They assert that the motion and light artifacts have a smaller impact on the U component. Hence, the HR and HRV signals can be extracted from the U component. To do so, first, they use a skin detection algorithm to mask the face from the background. Second, they extract the raw signal out of the facial images by combining the U component of the LUV images and the masked face

images. Third they perform a wavelet transformation to filter the signal and calculate HR from the final extracted signal.

Kumar et al. [32] extract the HRV using the Maximum Ratio Combining (MRC) method. This method combines the average pixel intensity of signals from different regions of the face. They achieved −5.32 dB for overall Signal to Noise Ratio (SNR) of the PPG signal and 8.09 ± 22.44 bpm for RMSE.

Antink et al. [33] present a multimodal method to monitor HR signal. They extract skin color variation and head motion from the video stream. Moreover, they use Ballistocardiographic sensor (BCG) to measure the ballistic forces on the heart. They infuse the extracted signal from each source to estimate the beat-beat HR signal. A Bayesian approach is chosen in order to remove the different noise characteristics of the three independent sources. They performed PCA and select the first five principal components using FFT to find the most relevant component to the HR signal. In addition, they compared the performance of ICA, MCR and spatial averaging on their dataset. They achieved better results by spatially averaging over the green channels, although the number of subjects (4 subjects) in the experiment is not sufficient to reach strong conclusions.

Park et al. [34] presents a non-contact measurement method to monitor cardiac responses using infrared images of the patient's pupil. Any change in major organs, such as the heart, can cause change in the pupil's contraction rhythm. In this study, they analyze the harmonic frequency of pupillary rhythm to find the response of the heart.

Rubinstein et al. [35] proposed a method called Eulerian Video Magnification (EVM) to reveal and magnify the small changes in skin color. The input of this system is a standard video sequence

from face. EVM uses localized spatial pooling and temporal filtering to detect and magnify subtle changes in the skin color that reflect BVP.

Alghoul et al. [36] investigate two methods to extract HR and HRV from a facial video. They compare the EVM and ICA approaches. Their results show that ICA-based methods are generally more accurate than EVM-based techniques.

Poh et al. [37] propose a method to recover the HR in non-laboratory conditions. They use a face tracker algorithm to track the small motions of the subject during measurement [4]. However, the subject remains in a fixed position relative to the camera. The subject's movements are restricted to tilting the head sideways, nodding, looking up/down, and leaning forward/backward. Then, they select a rectangular area of the face, which includes 60% of the width and the full height of the detected face as a Region of Interest (ROI). They decompose the ROI into the RGB channels by averaging the Red, Green and Blue amplitude values over all the pixels for each frame. They present the RGB signals to the ICA module to separate them into three independent components. One of the independent components produced by the ICA module is most reflective of BVP. However, it is not always clear which one. Poh et al. [4] heuristically choose the second component. Conversely, in another study by the same authors, they choose the signal with the maximum integrated power spectrum [37].

Aarts et al. [38] conduct a pilot study to investigate the feasibility of employing a camera-based PPG measurement system for remotely monitoring the HR of premature babies in the NICU. They use an RGB camera to record a facial video and manually select the ROI from the uncovered parts of the baby such as face and hands. They learned through their experiments that the green channel shows the strongest PPG signal among the other channels, hence, they apply further analysis on

this channel to acquire the proper HR signal. They successfully extracted the HR for all subjects except the ones that moved during the experiment. Moreover, they state that noise caused by continuous motion and light fluctuations negatively affected the performance of their system.

Monkaresi et al. [5] followed the Poh et al.[37] approach to extract the three independent components post ICA. However, they use ML techniques to select the best extracted component. They indicate that the output of their ML algorithm chose the third component as the most probable component to carry the HR information among all other components. They achieved lower RMSE compare to Poh et al. [37][4], however they did not present the result of their ML model in their published paper.

Some of the pervious work study extracting biological signals from multiple images. Gupta [39] presents a real-time non-contact imaging PPG method which uses synchronized multiple cameras to extract HR and HRV. They utilize a thermal, an RGB, and a Monochrome camera. In the experiment, the participants were seated in front of the three cameras. The extracted RGBMT (Red, Green, Blue, Monochrome, and Thermal) signals are spatially averaged over all the pixels in each frame. Furthermore, they use ICA to extract the independent signals and band pass filtering to remove the possible noise from the light and motion artifacts. Their results show that the RGT channel combination gives the lowest error (4.62%) in comparison with the other channel combinations (e.g. BGRM: 4.72% and RG: 4.77%).

Table 1 compares the existing algorithms with the proposed method in terms of the activity the subjects are engaged in, number of subjects, and techniques employed for face detection. The activity includes sitting without motion, sitting with natural movement, and movement without restrictions. The Number of Subjects column also reports the gender of the participants. We report

on the variety of facial detection algorithms utilized by the referenced studies. Table 2 presents the signal acquisition methods, including color spaces in use and source separation schemes are presented separately. The filters and their related cut-off frequencies are presented in the Noise Reduction column. The light modification algorithm responsible for reducing the negative effects of the light fluctuation is presented under the Light Equalization Method column. The last column describes the selected component to extract the final HR.

Table 1: Summary of the measurement set up of the reference works and proposed method

Ref.	Activity	# of Subjects	Human Detection Algorithm
[4]	1: Seated Stationary 2: Seated with Natural Movement	12 (2 F, 10M)	Face: OpenCV Library (a boosted cascade classifier)[40]
[5]	1: Seated Stationary 2: Natural Movement During Gaming 3: Indoor Cycling	10 (2 F, 8 M)	Face: OpenCV Library (a boosted cascade classifier)[40]
[27]	Stationary and seated upright	18 (7 F,11 M)	Face and Head (: OpenCV Library (a boosted cascade classifier)[40]
[28]	Seated stationary	10 (3 F, 7 M)	Palm Images
[29]	1: Seated Stationary 2: Seated under Cognitive Stress	10 (7 F, 3M)	Face: LEAR Face detector [30]
[31]	1 = Sated stationary and calm 2 = Seated with pre-defined head movements	12 (2 F, 8 M)	Face: OpenCV Library (a boosted cascade classifier)[40] Skin: Detection Using YC_bC_r color space [41]
[32]	1 = Seated Stationary 2 = Seated Reading 3 = Seated Watching video 4 = Seated Talking	12 (5 F, 7M)	Face: Deformable Face fitting algorithm [42] Feature Tracker [43]
[33]	1 = Seated stationary 2 = Seated Reading without motion 3 = Seated Reading without further instructions	4	Face: OpenCV Library (a boosted cascade classifier)[40]
[36]	Seated Stationary	12 (3 F, 9 M)	Face: OpenCV Library (a boosted cascade classifier)[40]
[37]	Seated Stationary	12 (4 F, 8 M)	Face: OpenCV Library (a boosted cascade classifier)[40]
[44]	Seated Stationary	4	Face: OpenCV Library (a boosted cascade classifier)[40]
[39]	Seated Stationary	9	Face: conditional regression forests (CRF) [45]
proposed Method	1: Seated Stationary 2: Walking freely (Distance 7 meter from camera)	11 (4 F, 7M)	OpenFace[46] KLT feature Tracker[43][47][48]

Table 2: Summary of the algorithm of the reference works and proposed method

Ref.	Color Space	Blind Source Separation Method	Noise Reduction	Light Equalization	Component selection
[4]	RGB	ICA	A Band pass Filter	N/A	Second Component
[5]	RGB	ICA	Moving Average Filter	N/A	ML* technique= Third component
[27]	RGB	PCA	Butterworth filter	NA	Component with highest peak in PSD
[28]	RGB	N/A	Butterworth Filter	N/A	First Component
[29]	RGB	ICA	Hamming window filter [0.75-3 Hz]	N/A	Component with highest peak in PSD
[31]	LUV	N/A	CWT and Inverse- CWT [0.65,3 Hz]	N/A	U Channel (represents Red and Green color indicator)
[32]	RGB	MRC	Hamming window filter	N/A	Green Channel
[33]	RGB	N/A	A pass band Filter [0.7-3.6 HZ]	N/A	Green Channel
[36]	RGB	ICA and EVM	- Five Point Moving Average - Hamming Window Filter [0.75-3Hz] - Butterworth Filter	N/A	Component with highest peak in PSD
[37]	RGB	ICA	- Five Point Moving Average - Hamming Window Filter [0.7-4 Hz]	N/A	Component with highest peak in PSD
[44]	RGB	ICA, PCA	Hamming Window Filter [0.5-3.7 Hz]	N/A	Combination of Red and Green
[39]	RGBMT	ICA	Band pass Filter[0.8-2.2 Hz]	N/A	Combination of Red, Green and Thermal
proposed Method	RGB	ICA	A Low Pass Filter [5 Hz] Hamming Window Filter [0.6-3.3] Butterworth Filter (around estimated HR)	Equalized Using the HSV Color space	ML* technique

2.2. Human Detection Systems

To develop a remote HR measurement system using video, we must first detect the presence of the monitored subject. Furthermore, to employ multiple cameras in the vital sign monitoring system, we need to detect the head pose of the subject with respect to the camera. This allows us to identify the camera that has the clearest view of the subject's face. Hence, in this section we will explore the topics of face and head pose detection.

2.2.1. Face Detection

Face detection is a mature research area that continues to gain attention from researchers due to the applicability of the technology to a wide variety of systems, such as biometric identification, surveillance applications, video conferencing, indexing of image, video databases and intelligent human–computer interfaces. For instance [49]used face detection for patient identification in medical record retrieval.

Face detection is one important step in face recognition systems. Face recognition is one of the popular approaches to biometric identification; hence many biometric systems infuse face recognition with other biometric features such as voice or fingerprints. Frischholz et al. [50] presents a model-based face detection method which combines face, voice, and lip movemenst recognition in a biometric systems. There are many digital image and video databases across the internet as the study of face detection algorithms has become an important part of many content based image retrieval (CBIR) systems. For example, Wactlar et al. [51]provides search and retrieval of TV news and documentary broadcasts based on the CBIR systems focusing on associating names with faces. In our work, we use a face detection scheme to locate the monitored subject in the video frame. Moreover, we propose a vital sign monitoring system consisting of

multiple cameras, where decisions involving which camera to use are based on an estimate of the head's angle with respect to each camera. Human faces are detected using a face detection algorithm [46]then the pose of the human head is computed based on the algorithm features.

There are several approaches to detect a face from a digital image or video frame. The traditional approaches find the facial landmarks (Feature-Based). Facial landmarks may vary from one method to another. An algorithm may choose the position or size of the eyes, nose, cheeks and mouth. For example, Christodoulou et al. [52] consider three features: skin, hair, and others to locate human faces. These landmarks are used to find matching features in the unknown images.

Other approaches use template-matching technique to recognize the face in images (Appearance-Based). Some of these techniques use the entire gray-scale face images as template images. For instance, Brunelli et al. [53] create a facial images dataset and numbers the dataset based on the image entry. Each unclassified image is compared with the dataset. The result of the comparison is a vector including matching score. The image with the highest matching score is the final face category. Furthermore, Sakai et al. [54] use several sub-template images for eyes, nose, mouth and cheekbones to create a new face template . Then, the correlation between unknown images and the face template is calculated to find the best match. Scale-Invariant Feature Transform (SIFT) is another template matching technique where texture features are extracted and stored as a template dataset. These features are known as SIFT key points. Best matches from any new images can be selected based on the location, scale and orientation of the objects with respect to the SIFT feature dataset [55]

2.2.2. Head Pose Detection

Multimodal Human Computer Interaction (HCI) employs multiple modalities including speech, faces [56] and head pose[57]. The human head pose can be described by pitch, yaw and roll (Figure 1). Head pose detection has many applications in HCI research. For instance, the influence of banner advertisements on attention has been studied using head pose and gaze detection algorithms in [58]. Moreover, Stiefelhagen [59] use head pose detection to track the attention of the audience in general meetings; they postulate that if an individual wants to pay attention to an object, he/she typically turns his/her head towards it.

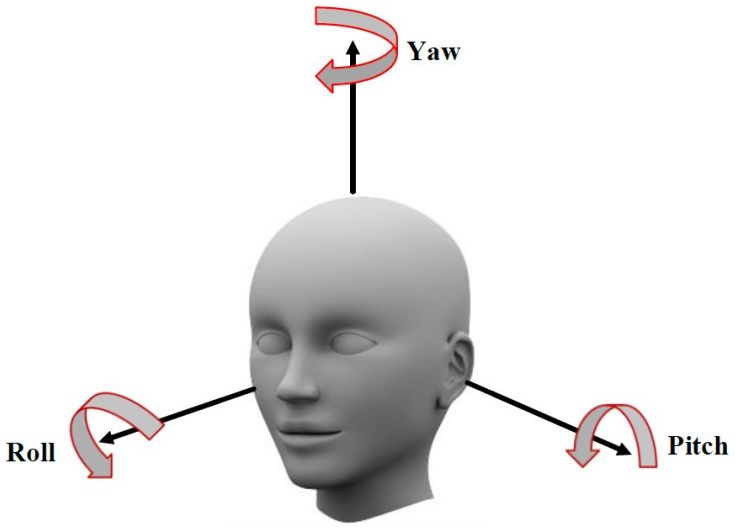

Figure 1: Head pose degrees of freedom

2.3. Independent Component Analysis (ICA)

Independent component analysis (ICA) [60] is a blind source separation technique to separate multivariate signals from underlying sources. Poh et al. [4] use ICA to extract the channel containing BVP information from RGB signal which is obtained from the video recording of the

human face by a webcam. ICA decomposes the RGB signals to three independent source signals. Any of them can carry PPG information. Let us denote the amplitude of the Red, Green and Blue signals extracted from the camera at the time point t by $y_1(t)$, $y_2(t)$ and $y_3(t)$, respectively. Moreover, let $x_1(t)$, $x_2(t)$ and $x_3(t)$ represent the underlying independent source signals after applying ICA. Then, the input of the ICA is a linear mixture of the underlying sources:

$$\boldsymbol{Y(t) = AX(t)} \qquad (1)$$

Where $\boldsymbol{t = [y_1(t), y_2(t), y_3(t)]^T}$, $\boldsymbol{X(t) = [x_1(t), x_2(t), x_3(t)]^T}$ and matrix A contains mixture coefficient a_{ij}. ICA decomposes the underlying sources by finding a demixed Matrix, W. W is the inverse of the matrix A:

$$\boldsymbol{\hat{X}(t) = WY(t)} \qquad (2)$$

To uncover the independent sources, W should be maximize the non-Gaussianity of each source.

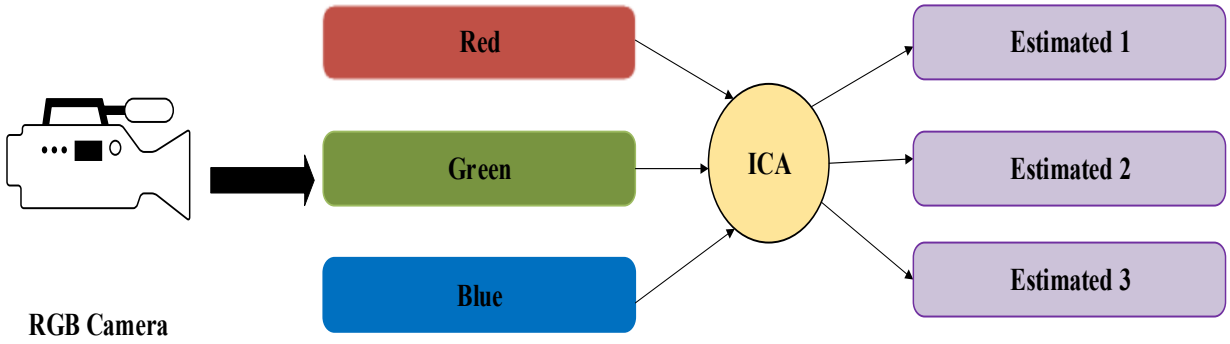

Figure 2: Independent Component Analysis

2.4. Signal Processing

It is difficult to measure HR immediately post-ICA, as the produced signals require noise reduction procedures. In this section, we will briefly describe the type of filters that are commonly used after ICA and how they are employed in remote HR measurement:

- **Moving Average Filter:** is a low pass Finite Impulse Response (FIR) filters which considers *M* sample of the input signal and takes their average to produce a single output point.[5] [36][37]
- **Low Pass Filter:** The range of the heart rate is in the low frequency range. The normal heart rate range of an adult at rest is between 1 to 1.67 Hz. Hence the designed filter removes frequencies above 1.67 Hz.
- **Hamming Window Filter:** is a band pass filter which reduces the ripples in the signal and retrieves a more accurate representation of the original signal's frequency spectrum [29][32][36][37][44].
- **Butterworth Filter:** An adaptive HR filter proposed by [36] to further clean up the signal. The adaptive filter uses an initial HR estimate to build a narrow band IIR Butterworth filter. The filter continuously updates its HR-Estimate (where HR-Estimate refers to the HR estimate in the previous video window) to ensure adaptability to changes in the HR.

2.5. Machine Learning

Machine Learning (ML) is one of the applications of Artificial Intelligence (AI) that enables the system to learn using statistical techniques[61]. In this procedure, the machine learns from a given dataset to make prediction or decisions.

There are two categories of ML techniques based on the learning process:

- Supervised Learning: Data is provided with labels. The machine learns from the existing relationship between the input, labels, and output, to predict the label of future data.

- Unsupervised Learning: No labels are provided. The machine defines a function that describes the structure of the unlabeled data.

An ML system can also be categorized based on its output: classification and regression. We will describe the latter two categories in the following sub-sections.

2.5.1. Classification

Classification modeling is the task of approximating a modeling function from input variables to a discrete output variable. The output of the classification is the label. The scheme function predicts the labels of each given observation.

2.5.2. Regression

Regressing is the task of approximating a modeling function from input variables to a continuous output variables. The output of the regression is a real value, hence the performance of this model must be reported as an error of the predictions.

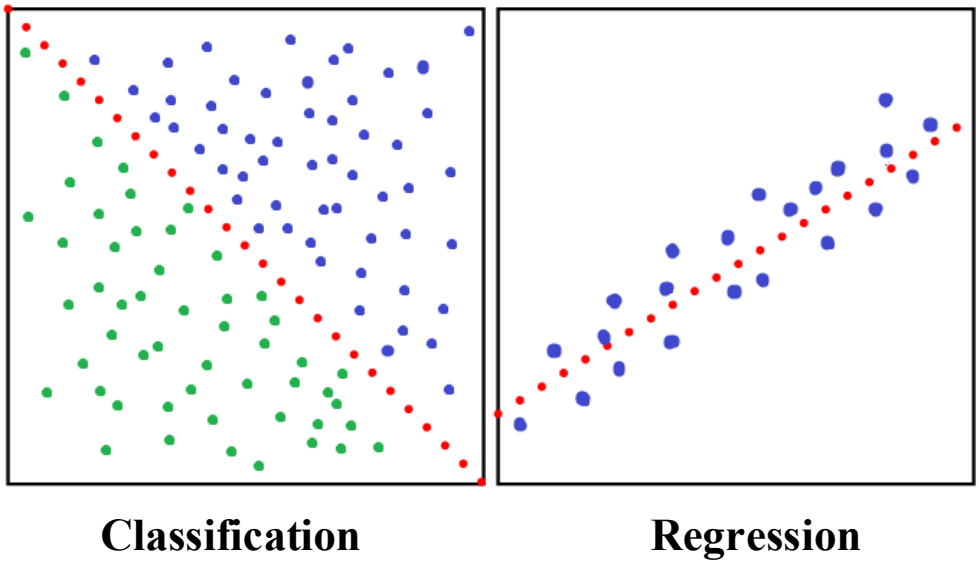

Figure 3: Machine Learning categories based on their output

Chapter 3. Proposed Method

We base our approach on the system proposed by Poh et al. [4]. Fig. 1a depicts the steps necessary to produce a HR estimation from video using the Poh et al. [4] technique. First, they find the face of the still subject using a face detection mechanism. Then, they select a rectangular area of the face, which includes 60% of the width and the full height of the detected face as a Region of Interest (ROI). They decompose the ROI into the RGB channels by averaging the Red, Green and Blue amplitude values over all the pixels for each frame. They present the RGB signals to the ICA module to separate them into three independent components. One of the independent components produced by the ICA module is most reflective of BVP. However, it is not always clear which one. Poh et al. [4] heuristically choose the second component. Conversely, in another study by the same authors, they choose the signal with the maximum integrated power spectrum [37].

In the proposed approach, we alleviate the assumption of subject stillness. Hence, we suppose that the subject is given freedom to roam the environment. The algorithm shall measure the HR given that it receives a video that captures at least 50% of the subject's face. Figure 4 depicts the steps involved in processing a facial video to estimate HR using the proposed approach. To describe our method, in the following subsections, we discuss only the differences between the proposed approach and Poh et al. [4]. We highlight our contributions in red in Figure 4.b.

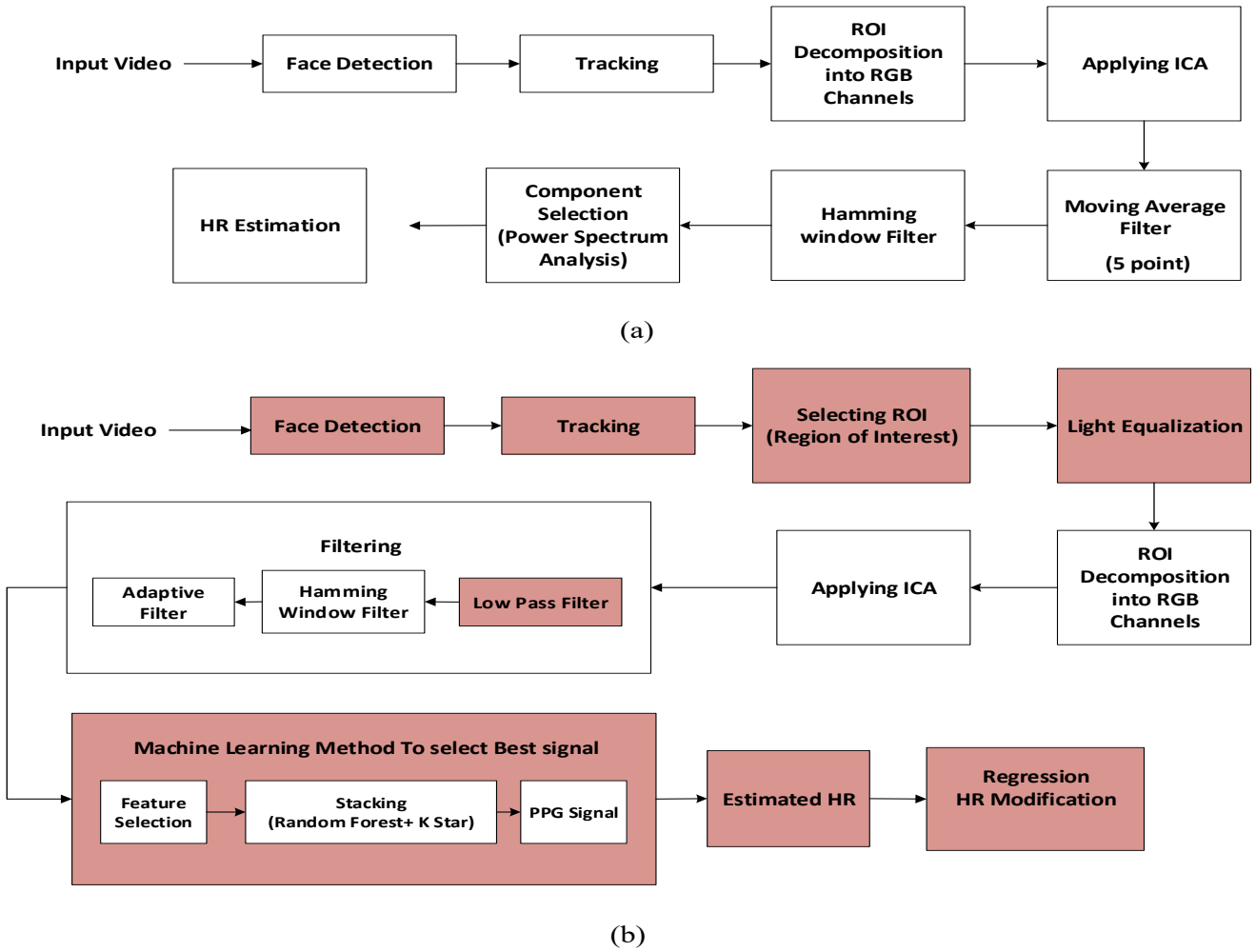

Figure 4: Method flowcharts of (a) Poh et al. [4] (b) proposed method

3.1. Single Camera Approach

In movement, the angle of the subject's face and the camera change frequently. In this thesis, we present two approaches for remotely measuring HR from video: the single camera and the multi-camera approaches. For the single camera approach, we assume that the face of the subject is visible in all video frames. Hence, one camera is used to capture the face of the subject. In the multi-camera approach described in Section 3.2. we utilize multiple cameras to capture the subject's face under a variety of head poses. In this section, we discus the single camera approach.

3.1.1. Face Detection

We use a face detection method and a robust face tracker to find and track the face in a video recorded by a standard RGB camera. In our experiment, we used a 24-bit RGB camera recording at a 30-frame per second and 640×480 pixels per frame. The face detection method (OpenFace)[19] finds the face in the first frame. However, after initial experiments, we found that this algorithm is unable to detect the face if the subject is more than two meters away from the camera. To address this problem, we add Kanade–Lucas–Tomasi (KLT) [47][48][62] feature tracker to the face detector. Using the tracker, we can track the detected face as the subject moves. The maximum allowed distance of the subject can now be seven meters from the camera. The combination of these two methods returns the face detected in each frame. We elect to consider the entire face of the subject as the ROI to collect enough facial information when the size of the face decreases due to the increase in the distance from the camera.

3.1.2. Light Equalization

Objects look different under varying light conditions. Human beings recognize objects in images with little effort despite lighting changes. However, it is challenging for computer vision systems to achieve the same feat.

The noise generated by fluctuations in the light levels can disrupt continuous measurement of HR. This is especially important given that the proposed approach, as opposed to previous ones [4][5], supports subject motion and roaming. Hence, changes in lighting levels on the face are bound to occur in an unevenly lit environment. To address this issue, we decompose the ROI pixels into the H (Hue), S (Saturation) and L (Lightness) color components. We calculate L′ by normalizing the values of L in the ROI across the frames of the video window. We then replace

L_{xyi} by L'_{xyi} for every pixel at (x, y)in frame i. We leave the values of the H and S components unchanged. For a given video window composed of n frames, we calculate L'_{xyi} using equation (1) where we denote the mean of $L_{xyi}$for pixel (x, y) over n frames by M_{xy} and the standard deviation by S_{xy}.

$$\boldsymbol{L'_{xyi} = \frac{L_{xyi} - M_{xy}}{S_{xy}}} \qquad (3)$$

In a continuous monitoring scenario, the video window refers to the length of video we use to calculate one HR value. Although different window sizes can be employed by the system's operator, for our experiments, we adopted a standard window length of 60 seconds (i.e. approximately 1,800 frames for a 30 frames-per-second sampling rate). We reconstruct the ROI using the H, S and L′ color components. Therefore, we decrease light fluctuation in the frames of a single video window. Figure 5 shows the changes in lighting across three frames before (Figure 5a) and after (Figure 5b) light equalization.

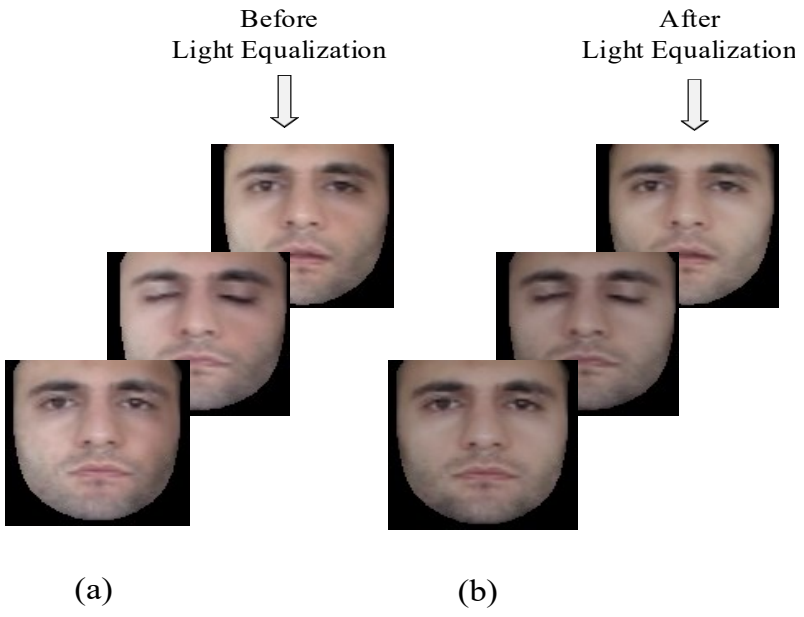

Figure 5: Face images (a) before and (b) after light equalization. The intensity of the light is different over the video frames. The right side of the face is darker than the left side in (a). The lightness of the face is equalized in (b) after the proposed algorithm

3.1.3. Signal Processing

To form the raw signals, the ROI is separated into the RGB channels then spatially averaged over all pixels. The raw signals from the Red, Green, and Blue channels are de-trended using a technique based on the smoothness priors approach [63]. The signals are de-trended by a smoothness parameter of λ=2000. Moreover, the de-trended signals are normalized by subtracting the mean and dividing by the standard deviation. After normalizing and de-trending, the RGB signals are presented to the ICA module. The order of the independent components produced by the ICA module is random. It is difficult to measure HR immediately post-ICA, as the produced signals require noise reduction procedures.

We apply a Low Pass Filter with a cut-off frequency of 5 Hz to the ICA output signals. The cut-off frequency is selected based on the frequency range of a typical HR signal. Furthermore, we apply a Hamming window filter to reduce the ripples in the signal and retrieve a more accurate representation of the original signal's frequency spectrum.

Next, we use a narrow band second order Butterworth filter proposed by [36]to further clean up the signal. This filter uses an initial HR estimate to build a narrow band IIR Butterworth filter with the following cut-off frequencies:

$$L\text{-}Cutoff = \alpha \times HR\_Estimate \quad (4)$$

$$H\text{-}Cutoff = \beta \times HR\_Estimate \quad (5)$$

where $\alpha < 1$ and $\beta > 1$.

Per Alghoul et al. [36], we adopt an α value of 0.8 and β value of 1.2. The filter continuously updates its HR-Estimate (where HR-Estimate refers to the HR estimate in the previous video window) to ensure adaptability to changes in the HR.

3.1.4. Channel Selection

ICA produces three independent components (i.e. signals) in a random order [64]. One of these independent components best reflects the PPG information. Therefore, making an assumption about which component contains the most useful information based on the order of the output of the ICA module contradicts the nature of ICA. Monkaresi et al. [5] use ICA and a ML model to predict the best component. They indicate that the output of their algorithm chooses the third component as the PPG carrying signal among all others, hence this is the channel they adopt. Poh et al. [4] heuristically choose the second component as the PPG reflecting signal.

Therefore, we propose a ML-based algorithm for independent component selection post-ICA. In the proposed method, the signal selection runs every time a video is processed to extract the HR. No assumption is made about the order in which the independent components are outputted by the ICA module. The independent component carrying the PPG information is dynamically calculated.

There are commonly two ways to calculate the HR after filtering is concluded [65]. The first one is to count the number of peaks in the time domain PPG signal. The second one, used by Poh et al. [37] and referred to as the MPA (maximum peak among all peaks) approach, is to identify the highest amplitude of the frequency domain representation of the PPG. The component with highest peak in the frequency domain corresponds to the HR. We adopt the latter technique as it is less sensitive to minor disturbances in the time domain signal that might introduce noise artifacts that mimic PPG peaks. We heuristically choose ten features to feed to the ML algorithm. They are categorized into two-feature sets: statistical descriptor features (e.g. mean and standard deviation of the PPG signal in the time domain) and physiological features (e.g. HR and HRV). We run a feature selection algorithm, namely PCA (Principal Component Analysis) [66], to select the most

useful features. This reduces the number of features to five. We use these features to train our predictive model. The first three features are obtained from the frequency domain. They are the amplitude of the highest peak, its corresponding frequency and ratio of the highest peak over the total power. If the $y(t)$ is the PPG signal in the frequency domain, the third feature can be written as

$$\boldsymbol{\omega = \frac{max|y(t)|^2}{E[|y(t)|^2]}} \qquad (6)$$

where $max|y(t)|^2$ is the maximum power and $E[|y(t)|^2]$ is the expected value of $|y(t)|^2$.

The fourth and fifth features are in the time domain and are the standard deviation and mean of the distance between troughs and peaks in the BVP signal.

We performed preliminary testing on several classifiers to select the ones that yield the highest accuracy and precision. For instance, we tested decision tree, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). In Section IV, we present the results we obtained from the best three classifiers: KNN, Random Forest, and K-star. KNN and K-Star are instance-based learning algorithms. KNN stores instances of the training data to classify the new test data. Each test data point is considered in the same class of the nearest instance to the known test data point. The k-neighbors nearest to the test instance are selected and then the average of their predicted values is assigned to the test instance. K-Star uses entropic distance of instances to classify the test dataset based on the probability of transforming instances into another by randomly choosing between all possible transformations. It provides a consistent approach to handle the real valued attributes and missing values. In fact, Cleary et al. [67] compared several instance based algorithm in classification. They utilize several popular datasets commonly used in ML literature. They conclude that K-star performs well across all dataset and has superiority over all instance-based

algorithms. Random Forest is an ensemble learning method which utilizes multiple decision trees to train the dataset and returns the mean prediction or mode of the classes from each decision tree The Random Forest method has several strengths that renders it an attractive classification approach. It can estimate which features are important for classification (inner feature selection algorithm). It reduces the chance of producing a model that over fits the training data and there is no need to perform tree pruning [68][69].

Furthermore, we propose to combine Random Forest and K-star classifiers through a stacking scheme to further improve the accuracy and precision of the algorithm. The basic principle of this algorithm is that a group of "weak learners" can be combined to realize a "strong learner". Stacking mixes several classifiers to make a final prediction from those of the combined classifiers.

The input to the ML algorithm is the extracted features from each component and the output is a "Yes\No" label of whether the component carries PPG information. Hence, the HR value calculated from the independent component labelled as "Yes" by the ML algorithm is the one adopted as the correct HR. Although unlikely, there is a possibility that we obtain two or three "Yes" classifications from the three evaluated independent components. Therefore, the algorithm calculates a final HR estimate by averaging all HR values obtained from all components that correspond to a "Yes" classification. Similarly, if we obtain three "No" classifications, then we average the HR values obtained from all three components. We provide details about the dataset we use to train and test this algorithm in Section 4.1. Also, we present the training procedure in Section 4.1.3.

3.1.5. HR Adjustment through Linear Regression

The HR calculated using the frequency domain based MPA approach is shifted with respect to the time domain based HR estimate [70] . In fact, in our preliminary investigation, we observed that the estimated HR derived through the MPA approach exhibits a linear shift with respect to the HR extracted from the ECG sensor through an R peaks count. We use linear regression to model this phenomenon. A linear regression model finds a linear relationship between one dependent variable and other explanatory or independent variables. We use this model to further correct the HR we extract from the video PPG signal using the MPA approach.

We create a dataset that includes the HR obtained from ECG (as a dependent variable) and the video PPG based HR estimated through the MPA approach (as an explanatory variable). We randomly divide the available dataset into two separate partitions: training and testing. We develop our model on the training set and use the testing set for predictive model assessment and refinement.

Given our dataset $\{y_i, x_i\}_{i=1}^{n}$ of n data points, the model takes the form

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_i + \varepsilon_i} \qquad (7)$$

where i=1,… ,n and β_0 and β_1 are model parameters β_1 is the slope and β_0 is the intercept of the linear model. The relationship is modeled in the presence of the error variable ε. We assume $E(\varepsilon_i)=0$ for all i. We use the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals, to estimate β. Hence, we need to solve:

$$\boldsymbol{\min_{\beta_0,\beta_1} \sum (y_i - \beta_0 - \beta_1 x_i)^2} \qquad (8)$$

The solution is given by:

$$\widehat{\boldsymbol{\beta_1}} = \frac{\boldsymbol{Cov}\,(\boldsymbol{x},\boldsymbol{y})}{\boldsymbol{Var}(\boldsymbol{x})} \quad (9)$$

$$\widehat{\boldsymbol{\beta_0}} = \sum(\boldsymbol{y_i} - \boldsymbol{\beta_1 x_i}) \quad (10)$$

The goodness of the fit is assessed in Section V by the R^2 measure, and when there is only one explanatory variable, is given by:

$$\boldsymbol{R^2} = \frac{\boldsymbol{Cov}\,(\boldsymbol{x},\boldsymbol{y})^2}{\boldsymbol{Var}\,(\boldsymbol{x})\boldsymbol{Var}(\boldsymbol{y})} \quad (11)$$

R^2 shows that how close the data fits the regression model. The range of R^2 is between 0 and 1. The value of 0 indicates that the regression model does not explain the variation in the predicted variable around its mean, and 1 indicates that the model explain perfectly the variation in the predicted variable around its mean. We presume that a shift between the HR obtained from the ECG sensor and HR obtained through the MPA approach from facial PPG does exist and can be linearly modeled. Such model allows us to improve the HR estimation as evidenced by the results presents in Chapter 4.

3.2. Multi-Camera Approach

In Section 3.1, we investigate the development of an HR estimation method for restricted subject movement using a single camera. We assume that the subject's face is visible to the camera at all time. However, in this section, we expand the movement area of the subject to include the entire room. In this approach, we detect the face and its pose during the movement of the subjects and switch between different cameras to capture the most appropriate facial frames for HR estimation.

3.2.1. Pose Detection

Pose detection is a general problem in Computer Vision to detect the position of a subject. There exists numerous algorithms to estimate the position and orientation of objects. These methods usually find key-points' locations that describe the orientation of the object. In our problem, we detect the facial landmarks to find the face in an image and then calculate the pose of the face with respect to the camera. We use multiple cameras to capture the most appropriate frames to extract the HR information. We use The Openface [46] face detection algorithm to identify the location of the facial images and landmarks for the estimation of the subjects' head pose. Openface identifies sixty-eight landmark points including points on the nose, eyebrows, chin, eye corners and mouth for face detection and pose estimation [46].

3.2.2. Camera Selection

We assume there are N cameras in the room. Hence, we divide the room into ω non-intersecting partitions, where ω is calculated using equation (12). We place the cameras along the perimeter of a circle co-centered with room. We ensure that the distance between adjacent cameras is equal for all cameras.

$$\boldsymbol{\omega = \frac{360^{\circ}}{N}} \qquad (12)$$

The system continuously selects every time-period t the most appropriate camera to measure the HR. The head pose angle of the subject is calculated with respect to each camera's axis. We refer to this angle as A_i where i is the index of the camera and ranges from 0 to N. Hence, the head pose angle of a subject looking directly at the camera is 0 degrees. Figure 6 shows the different head pose degrees with respect to each camera.

To calculate HR, for time-period *t,* we use the frames captured by the camera corresponding to the smallest A_i.

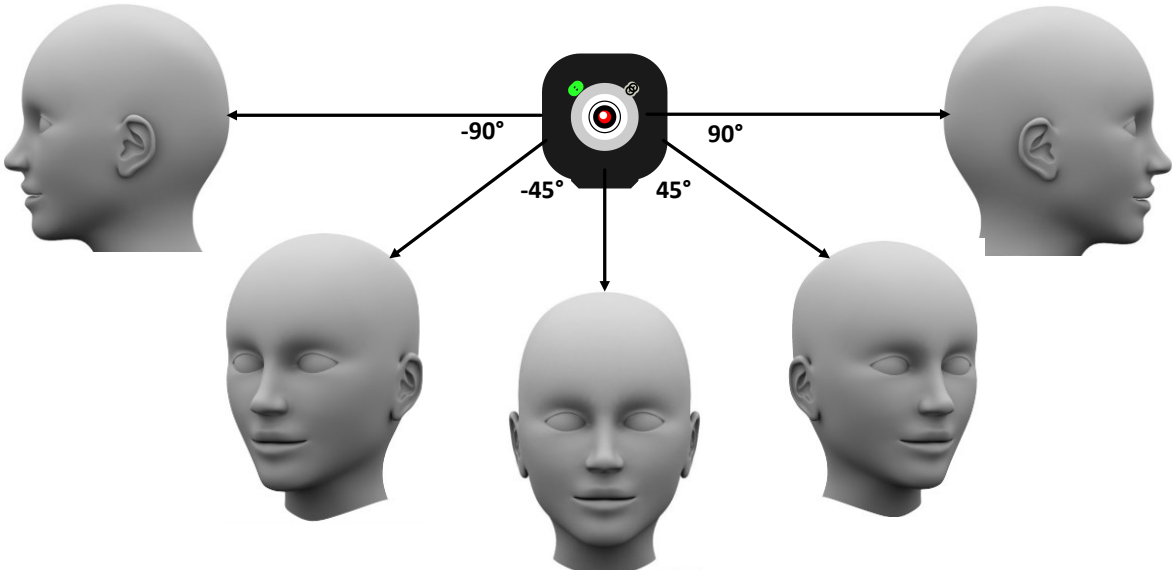

Figure 6: Head Pose degrees with respect to a camera

For the experimental scenario presented in Section 4.1.1.2 and Section 4.1.2.2, we employ four cameras placed on each corner of a square-shaped room. Therefore, video frames depicting a subject's head pose angle between [-45°, 45°] are used to estimate the HR. If we assume an HR monitoring session lasts *M* measurement time-periods, hence:

$$T = M \times t$$

Where *T* is the length of the recording session. Therefore, to measure HR during *T*, we collect *M* sets of frames from the *N* available cameras. The *M* sets of frames are considered the input video to the HR estimation method. Hence, all steps of processing post-frames compilation are identical to those of the single camera approach. To construct the raw RGB signals, the selected ROI (similar to the single camera approach) are spatially averaged over all pixels. The rest of method is the

same as the algorithm explained in section 3.1. Figure 7 shows an example of how the frames are compiled from *N* cameras to produce a single facial video.

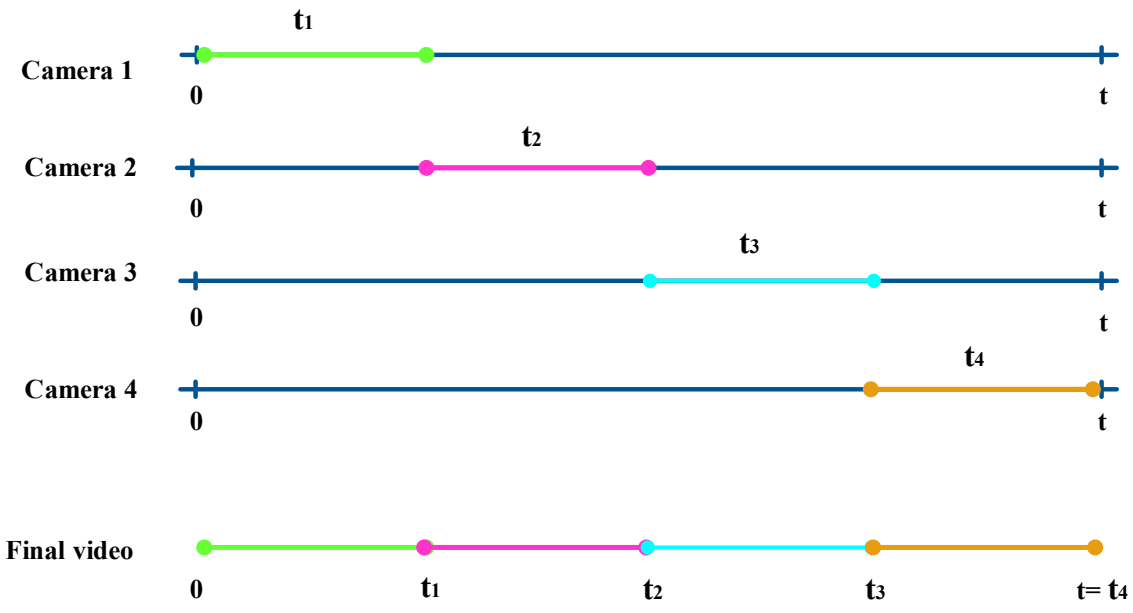

Figure 7: Face images extraction from multiple camera in stationary mode

Extracting HR from the face of the moving subjects with one camera has one important condition, at least 50% of the face of the subject should be visible in the camera. Otherwise the algorithm stops until detection of the face in next frames. Using multiple camera enables the algorithm to catch the face with other cameras if one of the cameras lost the face. However, of course, there are no guarantees that one of the cameras captures the face at an appropriate angle. For instance, the subject might come very close to the camera which crops her/his face out of the frame.

In the experimental scenario of Section 4.1.1.2 and Section 4.1.2.2 we consider a video time-period *t* of 1 second (video sampling rate is 30 fps).

Chapter 4. Experiment, Results, Evaluation and Discussion

The experiments discussed in this section were approved by the Office of Research Ethics and Integrity at the University of Ottawa (Certificate Number: H02-17-13). We collected a data set with eleven volunteer participants (4 female and 7 male) who signed consent forms to participate in this study. The data set was used in the experiments described in Sections IV-A and IV-B.

4.1. Dataset

We collected a dataset to train and/or test the light equalization, ML, and linear regression methods proposed in Sections III-B, III-D, and III-E respectively. The dataset collection procedure was approved by the Office of Research Ethics and Integrity at the University of Ottawa (Certificate Number: H02-17-13). Eleven adult subjects (4 female and 7 male) volunteered to participate and signed a consent form. We collected 220 video segments of the 11 subjects where each video segment is 1 minute long. The dataset consists of two subsets: Stationary and Movement Mode. The Stationary Mode subset is used in the experiment described in Section 4.1.1. The Movement Mode subset is employed in the experiment detailed in Section 4.1.2.

4.1.1. Stationary Mode Subset

4.1.1.1. Single Camera Mode:

We asked participants to sit in front of a camera for 5 minutes with minimum movement. They wore a physiological sensor (Zephyr Bioharness) that collects an ECG signal at a sampling rate of 250 HZ. The data collected by the sensor is used as ground truth to evaluate the accuracy of our results. Fig. 8a shows the experiment setup for stationary mode.

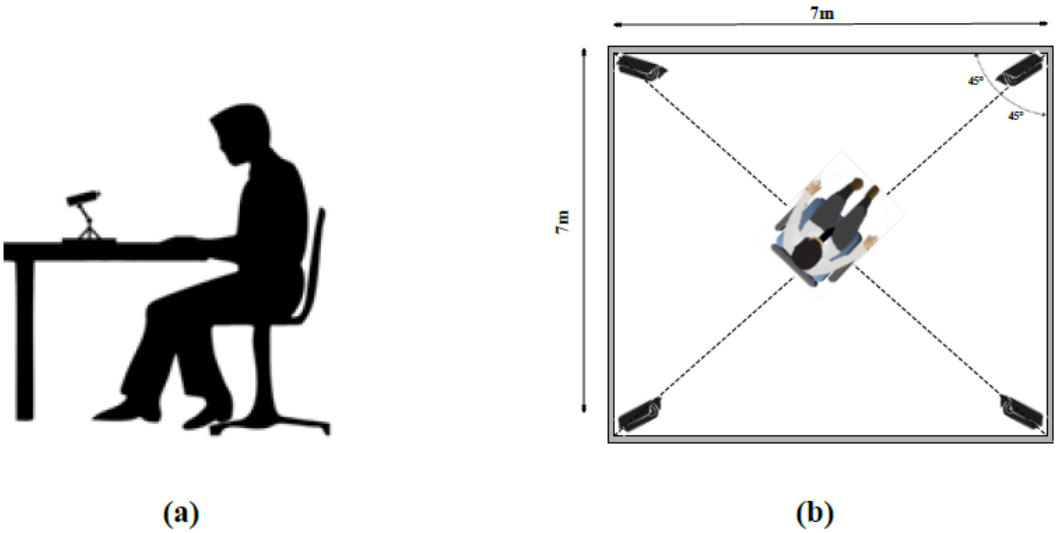

Figure 8: Experimental setup in stationary mode for (a) Single camera, (b) Multiple camera

The videos are collected by a conventional RGB camera (Logitech Webcam C270) at 30-frames per second and 640×480 pixels per frame—while the camera is mounted on a stable tripod at a one-meter distance from the subject. The lighting in the room is a mixture of natural sunlight coming from windows and fluorescent ceiling light.

Table 3 describes the conditions for the Stationary Mode subset collection.

Table 3: Description of the stationary data subsets with single camera

Parameters	**Stationary Mode**
Experiment length per subject (second)	300
# of participants	11 (4 F ,7 M)
Distance from camera (meter)	1
Recording rate (fps)	30
# of 1 min. video segments	55
# of frames	99000

4.1.1.2. Multi-Camera Mode:

We asked participants to sit in front of one camera and rotate towards the next camera with a speed of 30 degree/second. We asked them to stop for 10 second while facing each camera with minimum movement. They wore a physiological sensor (Zephyr Bioharness) that collects an ECG signal at a sampling rate of 250 HZ. The data collected by the sensor is used as ground truth to evaluate the accuracy of our results. Fig. 8b shows the experiment setup for stationary mode using multiple camera. The videos are collected by four conventional RGB camera (Logitech Webcam C920) at 30-frames per second and 640×480 pixels per frame—while the cameras are mounted on the four corners of a 4×4 meter square-shaped room and the subject is seated at the center of the room. The lighting in the room is a mixture of natural sunlight coming from windows and fluorescent ceiling light. Table 4 describes the conditions for the Stationary Mode subset collection with multiple cameras.

Table 4: Description of the stationary data subsets with multiple camera

Parameters	**Stationary Mode**
Experiment length per subject (second)	180
# of participants	10 (3 F ,7 M)
Distance from camera (meter)	0-2
Recording rate (fps)	30
# of 1 min. video segments	30
# of frames	54000

4.1.2. Movement Mode Subset

4.1.2.1. Single Camera Mode:

As opposed to the Stationary Mode, for the Movement Mode subset, we asked the subjects to perform two tasks:

1- Restricted Movement: move in the room for 15 minutes by walking, talking, or lying down given that at least 50% of their face is visible to the camera (see Fig. 3a). Sections of the video where less than 50% of the face was visible were not considered.

2- Unrestricted Movement: move freely in the room without any consideration about the camera (they can turn away from the camera). This task was done to compare the performance of the single camera against multiple camera in the same situation.

The dimensions of the room are 7m×7m. The rest of the experimental conditions, including lighting, camera, and sensor are identical to the ones used for the experiment described in Section IV-A. Figure 9 shows the room setup for subjects in movement.

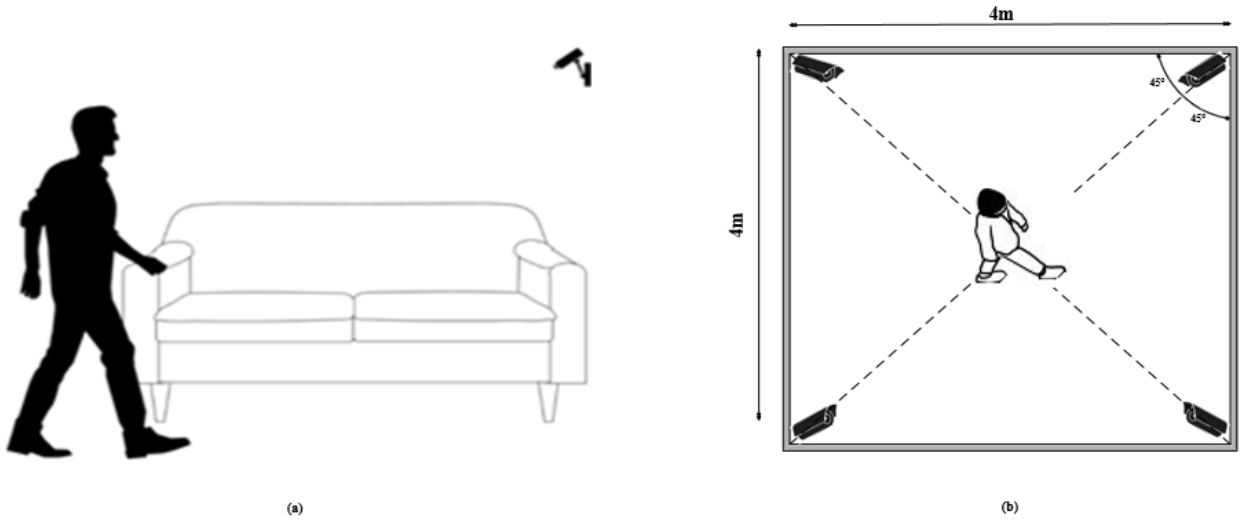

Figure 9: Experimental setup in movement mode for (a) Single camera, (b) Multiple camera

Table 5 describes the conditions for the Movement Mode subset collection with single camera.

Table 5: Description of the movement mode data subsets with single camera

Parameters	**Movement Mode**
Experiment length per subject (second)	**900**
# of participants	**11 (4 F ,7 M)**
Distance from camera (meter)	**0-7**
Recording rate (fps)	**30**
# of 1 min. video segments	**165**
# of frames	**297000**

4.1.2.2. Multi-Camera Mode:

For the Movement Mode subset, we asked the subjects to move in the room for 4 minutes by walking, talking, or lying down (see Fig. 9b). The dimensions of the room are 4m×4m. The subjects can move freely without any constraint. In this experiment at least one of the cameras can capture the face based on the subject's angle with respect to each camera. The rest of the experimental conditions, including lighting, camera, and sensor are identical to the ones used in the experiment described in Section 4.1.1. Table 6 shows the conditions for the movement mode subset collection with multiple cameras.

Table 6: Description of the movement mode data subsets with multiple camera

Parameters	**Movement Mode**
Experiment length per subject (second)	**240**
# of participants	**10 (3 F ,7 M)**
Distance from camera (meter)	**0-7**
Recording rate (fps)	**30**
# of 1 min. video segments	**40**
# of frames	**72000**

4.1.3. Dataset Training

We train the ML algorithm we employ for channel selection post-ICA (proposed in Section 3.1.4) using a randomly chosen 70% of the dataset (i.e. 154 video segments). We reserve 30% of the dataset for later testing. Each video is processed using the proposed method described in Section III and depicted in Fig. 1b up to the Filtering stage. The ICA module produces three independent components. Hence, from the 154 video segments, we extract a total of 462 independent components (3 from each video). We manually attach a label of "Yes" or "No" to each independent component based on the ground truth HR as measured by the ECG sensor. Evidently, the "Yes" label corresponds to the independent component that best reflects the HR compared to the two other ones produced from the same video segment. That is, the "Yes" labelled independent component is the most appropriate to estimate the HR from the facial PPG. To establish the label for an independent component, we calculate the absolute difference between the HR estimated from that component and HR obtained from the ECG sensor. We suppose that the independent component that renders the smallest absolute difference is the one that best reflects the HR, and consequently is the one that is granted the "Yes" label. The other two are given the "No" label.

During training, the dataset is randomly shuffled to maintain generalized learning. By shuffling the data, we ensure that each data point creates an independent change on the model, without being biased by previous ones. We use a 10-fold cross validation to assess the fitness of the model before evaluating it with the testing data set.

4.2. Single Camera Mode

In this section, we provide an evaluation of the proposed contributions with a single camera. The dataset used in the experiments is described in Sections 4.1.1.1 and Section4.1.2.1.

4.2.1. HR Estimation in Stationary Mode

We present the results of the HR estimation from facial video in the Stationary Mode. Hence, we use the Stationary Mode data subset described in Section 4.1.1.1 for all assessments.

4.2.1.1. ICA Independent Component Channel Selection

The ICA module outputs three channels, one of which contains the most accurate representation of the PPG signal. To establish the ground truth with respect to the best component carrying the PPG information, we calculate the absolute difference between the HR estimated from each independent component and HR obtained from the ECG sensor. We suppose that the independent component that renders the smallest absolute difference is the one that reflects best the PPG information. Fig. 4 shows an example of the independent components post ICA and filtering.

To develop an independent component selection scheme, we train several ML algorithms on the given dataset. During training, the dataset is randomly shuffled to maintain generalized learning. By shuffling the data, we ensure that each data point creates an independent change on the model, without being biased by previous ones. Table 7 presents the classification results for the four considered ML schemes: KNN, Random Forest, K-Star and Stacked Random Forest and K-Star. The Stacked Random Forest and K-Start yields the best results in terms of Precision, Recall, and Accuracy.

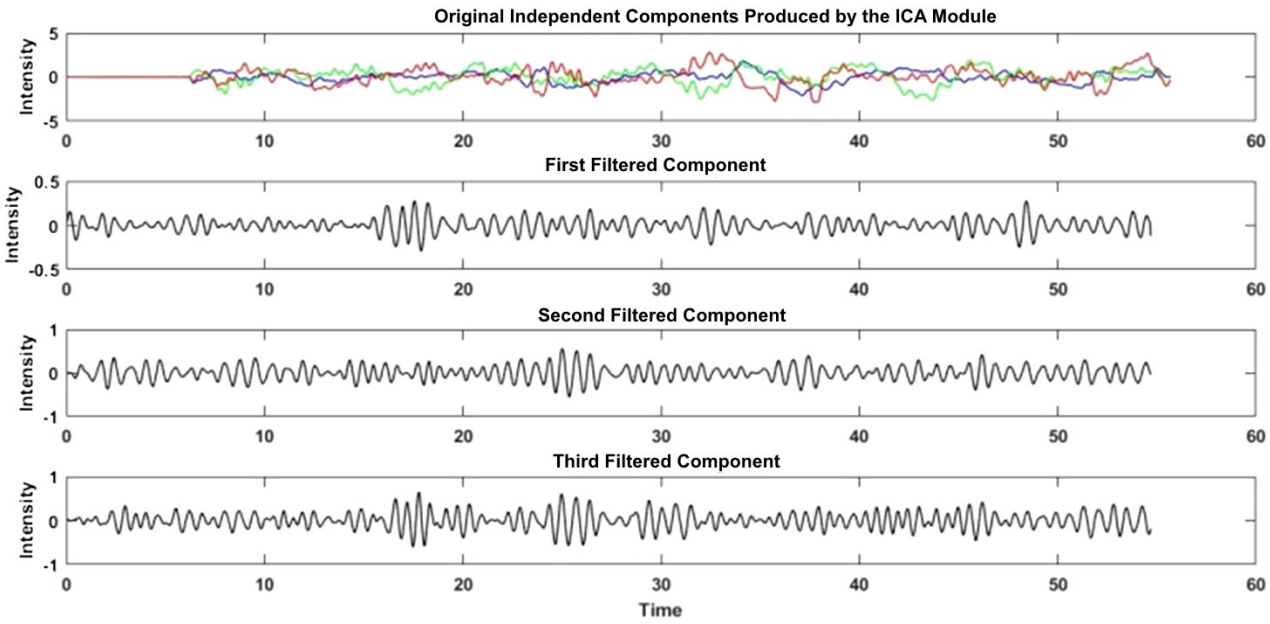

Figure 10: Output of the ICA and noise reduction procedures

This is unsurprising given that stacking two or more classifiers with reasonable performance typically increases the overall accuracy of the model [69]. Hence, we will adopt this scheme as our proposed ML-based method.

Table 7: Performance of the proposed ML Technique

Performance Metrics	KNN	Random Forest	K-Star	Stacking C (Random Forest, K-Star)
Precision	0.848	0.854	0.852	0.870
Recall	0.845	0.855	0.849	0.869
Accuracy (%)	84.539	85.191	85.213	86.871
Area under ROC Curve (AUC)	0.869	0.912	0.871	0.934

To select the independent component carrying the PPG signal, Poh et al. [4] heuristically choose the second component. Monkaresi et al. [5] propose a ML technique to choose the best independent

component produced by the ICA module. They select nine features to train their ML algorithm. The features are the frequency with the highest magnitude in each component before and after noise filtering and the index of the spectral peak in the Power Spectral Density (PSD) of each component, after the application of the noise reduction method. Their algorithm utilizes ICA+KNN to select the best signal out of the three independent components. They find that the third component produces the lowest RMSE compared to the others. However, they did not report the performance of their ML scheme. They conclude that the third component is the most probable signal that carries the HR information.

We calculate the absolute difference between the HR obtained from ECG and the HR obtained from each of the three independent components outputted by the ICA module for a video segment. We count the independent component that results in the lowest absolute difference to be the one carrying the useful HR information. Through this process, we found that for our dataset, the HR is best calculated as follows by the three channels: 13.11% by the first component, 30.58% by the second component, and 56.31% by the third component. Although it seems from these results that the PPG signal is mostly carried by the second and third components, we cannot select a single channel that consistently best carries the PPG information. Table 8 compares the proposed ML-based method (Stacked Random Forest and K-Star) with that of Poh et al. [4] and Monkaresi et al. [5]. The proposed scheme produces better results compared to the approaches by Poh et al. [4] and Monkaresi et al. [5] in terms of Precision, Recall, and Accuracy. The ICA produces the independent components in a random order [64] .Hence, it might be counterproductive to assume that one component, based on its order as it appears in the output of the ICA module, will always best reflect the PPG signal. Hence, the application of the proposed ML algorithm to evaluate all independent components and resolve the best one removes any reliance on the order at which the

independent components appear in the output. In the proposed machine learning scheme, the application of the PCA technique to narrow our initial feature set allowed us to preserve the most relevant features to solve the problem. Furthermore, the application of stacking permitted us to leverage the strengths of the Random Forest and K-Star algorithms.

Table 8: Comparison of Poh et al. [4], Monkaresi et al. [5], and the proposed ML scheme for channel selection

Performance Metrics	**Poh et al.** [4]	**Monkaresi et al.** [5]	**Stacking C (Random Forest, K-Star)**
Precision	0.271	0.463	0.870
Recall	0.266	0.455	0.869
Accuracy (%)	26.658	46.284	86.871
Area under ROC Curve (AUC)	51.356	60.472	0.934

4.2.1.2. HR Estimation before Light Equalization and Regression

Table 9 shows the comparison of the Poh et al. [4] and Monkaresi et al. [5] HR estimation techniques with the proposed method (without the application of light equalization and regression). We can observe the positive effect of accurately selecting the independent component through the proposed ML-based technique on the results. For instance, the RMSE decreased by 37% and 21% compared to Poh et al. [4] and Monkaresi et al. [5], respectively.

Table 9: Comparison of Poh et al. [4], Monkaresi et al. [5] and proposed method before light equalization in stationary mode

Parameters	Poh et al [4]	Monkaresi et al. [5]	Proposed method
Selected component	2nd	3rd	-
Mean Bias (bpm)	-0.12	0.91	0.79
SD of Bias (bpm)	3.11	2.03	1.34
Upper limit (bpm)	6.09	3.97	2.62
Lower limit (bpm)	-6.33	-2.15	-1.83
RMSE (bpm)	2.38	1.89	1.50
Corr. coefficient	0.98*	0.98*	0.98*

*: ($p<0.004$)

4.2.1.3. HR Estimation After Light Equalization and Before Regression

Table 5 shows the improvement in HR estimation when light equalization is applied. We achieved lower RMSE in comparison with Poh et al. [4] and Monkaresi et al. [5]. By utilizing light equalization, the fluctuation of light across frames in a window is stabilized. These fluctuations can possibly introduce noise to the signal as they alter the skin color from which the PPG signal is derived. Hence, their reduction results in a cleaner signal as it is evidenced by the results of Table 10.

Table 10: Comparison of Poh et al. [4], Monkaresi et al. [5], and proposed method after light equalization in stationary mode

Parameters	Poh et al. [4]	Monkaresi et al. [5]	Proposed method
Selected components	2nd	3rd	-
Mean Bias (bpm)	-0.12	0.91	0.78
SD of Bias (bpm)	3.11	2.03	1.32
Upper limit (bpm)	6.09	3.97	2.58

Lower limit (bpm)	-6.33	-2.15	-1.03
RMSE (bpm)	2.38	1.89	1.47
Corr. coefficient	0.98*	0.98*	0.98*

*: ($p<0.004$)

4.2.1.4. HR Estimation after Light Equalization and Regression

The proposed regression model in Section III-E finds a relationship between the HR calculated from the video derived PPG signal and HR extracted from the ECG sensor. Using this model, we are able to eliminate any systematic bias between the two HR extraction methods. Hence, when we applied the regression model, we obtained a lower mean and standard deviation of bias and a lower RMSE. Table 11 summarizes the details of the results before and after applying the regression model.

Table 11: Comparison of the estimated HR after light equalization and before and after regression in Stationary Mode

Parameters	**Before Regression**	**After Regression**
Mean Bias (bpm)	0.78	0.70
SD of Bias (bpm)	1.32	1.25
Upper limit (bpm)	2.58	3.15
Lower limit (bpm)	-1.03	-1.75
RMSE	1.47	1.38
Corr. coefficient	0.98*	0.99*

*: ($p<0.004$)

4.2.2. HR Estimation in Movement Mode

In this section we use the dataset from Section 4.1.2.1. Table 12 compares the estimated HR with subjects' movements before and after the light equalization. In comparison to the stationary mode, the positive effect of light equalization is larger in the presence of movement as the light projected on the face fluctuate more in this situation. As the subject displaces in the unevenly lit room, the amount and direction of light reflecting on the face varies, thus requiring the need for light equalization to reduce the negative effect of such phenomenon on the quality of the retrieved PPG signal. In fact, the application of light equalization decreases the RMSE by 2% for the experiment with stationary subjects (Table 10) compared to a decrease of 30% for the experiment with subjects in movement.

Table 12: Comparison of the estimated HR before and after light equalization in Movement Mode

Parameters	**Before Light Equalization**	**After Light Equalization**
Mean Bias (bpm)	2.12	1.47
SD of Bias (bpm)	3.96	3.10
Upper limit (bpm)	7.76	5.78
Lower limit (bpm)	-3.48	-3.46
RMSE	2.43	1.69
Corr. coefficient	0.88*	0.89*

*: ($p<0.004$)

Table 13 shows the results of the proposed algorithm when light equalization and regression are applied. We achieved a 34% reduction in RMSE compared to the experiment when regression was not applied.

Table 13: Comparison of the estimated HR after light equalization and before and after regression in Movement Mode

Parameters	Before Regression	After Regression
Mean Bias (bpm)	1.47	1.20
SD of Bias (bpm)	3.10	1.77
Upper limit (bpm)	5.78	4.70
Lower limit (bpm)	-3.46	-2.30
RMSE	1.69	1.12
Corr. coefficient	0.89*	0.93*

*: ($p<0.004$)

Fig. 11 reports the results of the regression model and bland Altman plot on the Actual HR and our estimated HR. The Bland Altman plot shows the high agreement between the actual HR and the estimated HR.

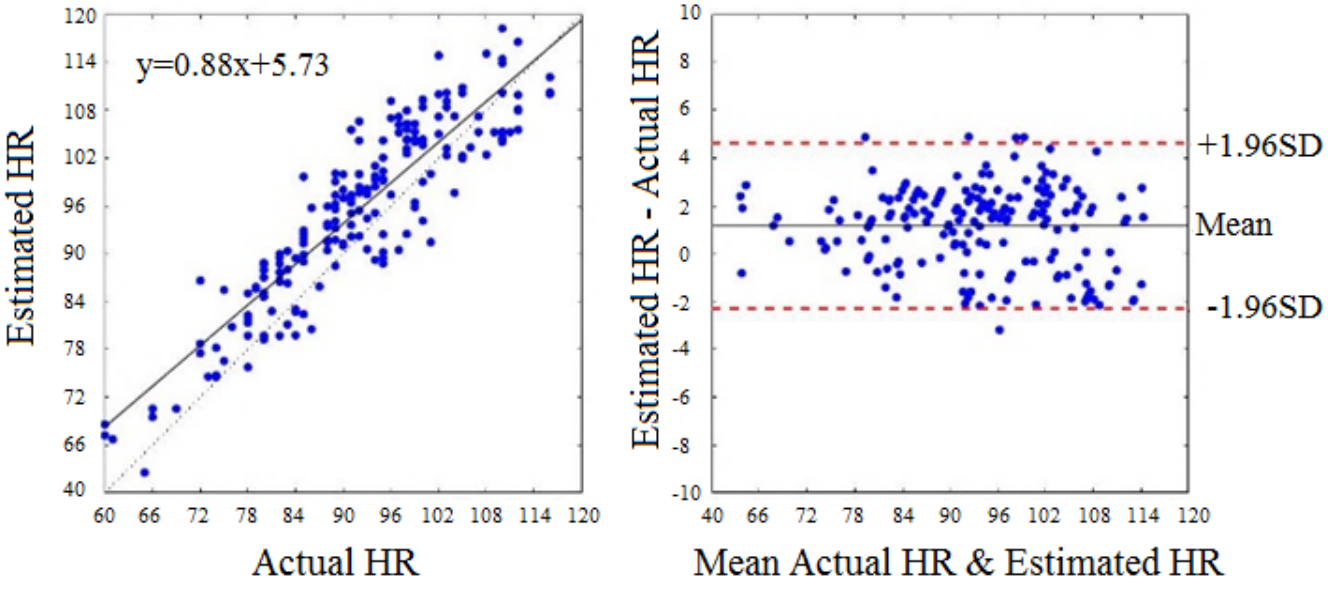

Figure 11: Statistical Results of (a) Regression model (b) Bland and Altman plot

4.3. Multi-Camera Mode

In this section, we provide an evaluation of the proposed contributions using multiple cameras. The dataset used in the experiments is described in Sections 4.1.1.2 and 4.1.2.2.

4.3.1. HR Estimation in Stationary Mode

In this section, we present the results of the HR estimation from facial video in the Stationary Mode. Hence, we use the Stationary Mode data subset described in Section 4.1.1.2 for all assessments.

4.3.1.1. HR Estimation Before and After Light Equalization

As mentioned in section 4.2.2, the light equalization method improves the HR estimation in movement more than stationary mode. However, in this experiment, due to the rotation of the subjects toward each camera, the lighting in the successive video frames needs to be equalized. Table 14 compares the estimation of HR before and after the light equalization.

Table 14: Evaluation of the method before and after Light Equalization

Parameters	**Before Light Equalization**	**After Light Equalization**
Mean Bias (bpm)	0.831	0.801
SD of Bias (bpm)	1.382	1.330
Upper limit (bpm)	2.704	2.606
Lower limit (bpm)	-1.874	1.806
RMSE (bpm)	1.55	1.511
Corr. coefficient	0.98*	0.98*

*: ($p<0.004$)

4.3.1.2. HR Estimation Before and After Regression

Regression has improved the HR estimation in both cases of single and multiple cameras. In Table 15 we provide the result of applying the Regression model on the data to evaluate its effect on the HR estimation.

Table 15: Evaluation of the method after Light Equalization and before and after Regression

Parameters	**Before Regression**	**After Regression**
Mean Bias (bpm)	0.801	0.743
SD of Bias (bpm)	1.330	1.29
Upper limit (bpm)	2.606	2.528
Lower limit (bpm)	1.806	-1.788
RMSE	1.511	1.433
Corr. coefficient	0.98*	0.99*

*: ($p<0.004$)

4.3.2. HR Estimation in Movement

We use the dataset described in Section 4.1.2.2. for this experiment. In this section we estimate the HR signal extracted from a single camera to the signal employed from four camera in unrestricted movement. First, we provide the results of Light Equalization algorithm due to its significant impact on our result. Then, we provide the results of the regression algorithm.

4.3.2.1. HR Estimation Before and After Light Equalization

Light fluctuation in movement is more sensible than in stationary. When the subject moves in the room the distance and angle of the subject change with respect to the light source. In addition different cameras have different angel with respect to the light source. Switching between cameras

can produce new source of noise to the signal. Equalizing Light across the frames removes the unwanted frequencies from the signal. Table 16 provides the result of the estimated HR before and after Light Equalization.

Table 16: Evaluation of the method before and after Light Equalization

Parameters	Before Light Equalization		After Light Equalization	
	Single Camera	Four Cameras	Single Camera	Four Cameras
Mean Bias (bpm)	8.651	1.992	7.922	1.39
SD of Bias (bpm)	7.542	3.871	6.891	2.81
Upper limit (bpm)	14.782	7.585	13.506	5.507
Lower limit (bpm)	-6.131	-5.595	-5.584	-4.117
RMSE (bpm)	6.412	1.912	5.712	1.43
Corr. coefficient	0.80	0.93	0.84	0.95

*: ($p<0.004$)

The result of the Table 15 shows that using more than one cameras can improve the accuracy and decrease the error of the estimation. One camera alone cannot cover all of the angle of a moving person's face. Each frame is spatially averaged and creates a data point of the estimated HR signal. Then, the estimated HR signal loses some of the essential points when some faces of the subjects are missing due to the bad video coverage of a single camera. Hence the error of the estimation is higher with one camera in comparison with multiple cameras.

4.3.2.1. HR Estimation Before and After Regression

Using regression the proposed algorithm robustness increased and the RMSE decreased from 1.43 to 0.96 bpm. Table 17 presents the improvement achieved by the regression algorithm.

Table 17: Evaluation of the method before and after Regression

Parameters	Before Regression		After Regression	
	Single Camera	Four Cameras	Single Camera	Four Cameras
Mean Bias (bpm)	7.922	1.39	6.651	1.081
SD of Bias (bpm)	6.891	2.81	6.341	1.364
Upper limit (bpm)	13.506	5.507	12.428	2.665
Lower limit (bpm)	-5.584	-4.117	-5.774	-1.585
RMSE (bpm)	5.712	1.43	5.081	0.96
Corr. coefficient	0.84	0.95	0.85	0.98

Figure12 presents the performance of the regression model and bland Altman plot. Bland-Altman plot in Figure 12 analyze the agreement between the estimated HR from the proposed algorithm using multiple cameras and the HR extracted from the ground truth.

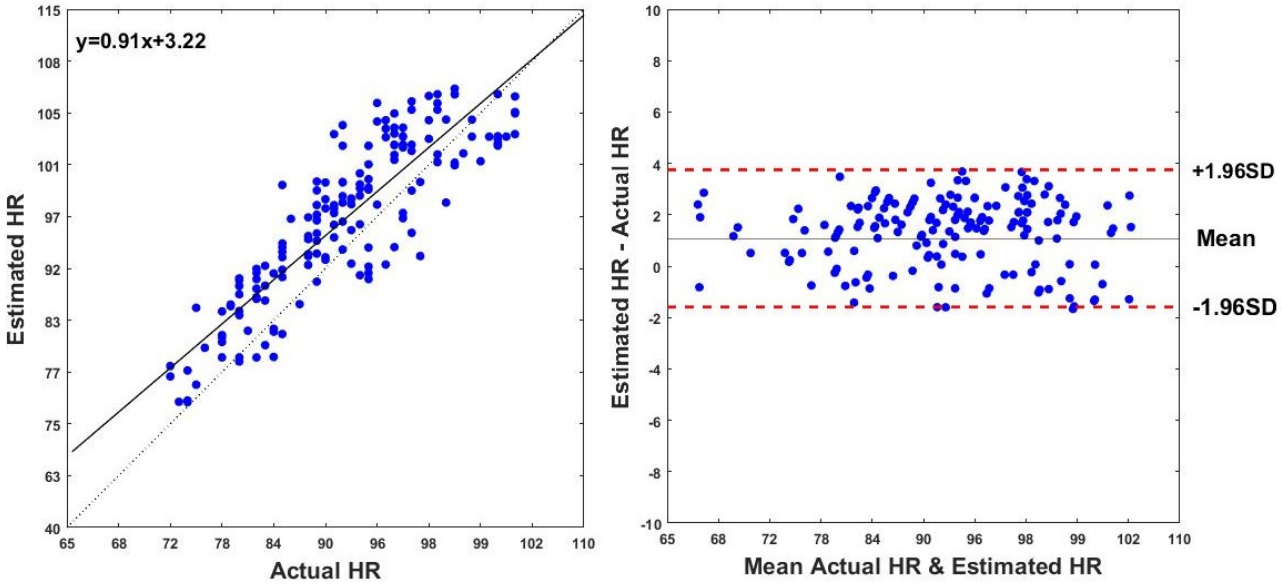

Figure 12: Statistical Results of (a) Regression model (b) Bland and Altman plot In Multiple camera experiment

Chapter 5. Conclusion & Future Works

In this chapter, we conclude the results we reached in this work. Then, we describe ideas for future research in this field.

5.1. Conclusions

We proposed and evaluated a method for the remote measurement of HR using a PPG signal extracted from a video recording of a subject's face. The proposed method is based on the technique presented by Poh et al.'s [1]. However, we extended the latter technique by implementing a light equalization scheme to reduce the effect of spatial and temporal light variation on the HR estimation, a ML method to select the most accurate channel outputted by the ICA module, and a regression model that estimates the relationship between the HR calculated through the MPA scheme and the measurement extracted from an ECG signal.

We distinguish our work from existing ones as follows:

- We propose a light equalization scheme to minimize the negative effect of noise generated from the fluctuation of light on HR estimation;
- We develop a ML algorithm to select the BVP information carrying component after performing ICA on the RGB channels; and
- We propose a linear regression model to improve the accuracy of the HR estimate obtained through the ICA.
- We propose a multiple camera HR monitoring system to allow the monitored subject to roam an environment without restriction.

In our experimental results, we found that the proposed light equalization technique reduces the RMSE by 0.74 bpm for moving subjects and 0.03 bpm for still subjects, the ML technique selects the correct component outputted by ICA with an accuracy of 86.9%, and the application of the regression model to adjust the estimated HR reduced the RMSE by 0.57 bpm for moving subjects and 0.09 bpm for still subjects.

Moreover, using multiple cameras improves the limitation of movement for detection of face while the face is not visible for one camera. In our method we extract and combine partial signals from multiple cameras. If subjects move freely in the room one camera cannot capture face all the time and we would face significant frame loss. We achieved 0.96 bpm for RMSE in estimating the HR using multiple cameras. Our method presents possibility of continuous measurement of physiological vital signs in stationary and movement of the subjects.

5.2. Future Works

We employ several algorithm to extract HR from the face images. There are several improvements that can be made to the proposed method:

1. Deep Learning algorithm can be used to extract the useful features from the raw signals after the ICA, as oppose to manually selecting the features.
2. The HR estimates can be used in an affective computing system to detect the emotions of the monitored individual. In fact, since we are tracking the face across multiple frames, we can combine facial expression estimation with the HR information to realize a multimodal affect recognition system.

3. In this thesis, we do not investigate the implementation of this system on hardware platforms. Therefore, deploying the proposed method from the camera to output of the regression model in a real time scenario may interest researchers in the field.

Bibliography

[1] A. Pantelopoulos and N. G. Bourbakis, "A Survey on Wearable Sensor-Based Systems for Health Monitoring and Prognosis," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 1, pp. 1–12, Jan. 2010.

[2] S. Gontarek, R. P. Remote, D. Mcduff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability The MIT Faculty has made this article openly available . Please share Citation Accessed Citable Link Detailed Terms Remote Measurement of Cognitive Stress via Heart Rate Variability," 2017.

[3] M. Villarroel *et al.*, "Continuous non-contact vital sign monitoring in neonatal intensive care unit.," *Healthc. Technol. Lett.*, vol. 1, no. 3, pp. 87–91, Sep. 2014.

[4] M.-Z. 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, p. 10762, May 2010.

[5] H. Monkaresi, R. A. Calvo, and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam," *IEEE J. Biomed. Heal. Informatics*, vol. 18, no. 4, pp. 1153–1160, 2014.

[6] L. Fanucci *et al.*, "Sensing Devices and Sensor Signal Processing for Remote Monitoring of Vital Signs in CHF Patients," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 3, pp. 553–569, Mar. 2013.

[7] Litong Feng, Lai-Man Po, Xuyuan Xu, Yuming Li, and Ruiyi Ma, "Motion-Resistant Remote Imaging Photoplethysmography Based on the Optical Properties of Skin," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 879–891, May 2015.

[8] Q. Zhang, X. Zeng, W. Hu, and D. Zhou, "A Machine Learning-Empowered System for Long-Term Motion-Tolerant Wearable Monitoring of Blood Pressure and Heart Rate With Ear-ECG/PPG," *IEEE Access*, vol. 5, pp. 10547–10561, 2017.

[9] L. Fanucci *et al.*, "Sensing Devices and Sensor Signal Processing for Remote Monitoring of Vital Signs in CHF Patients," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 3, pp. 553–569, Mar. 2013.

[10] S. Reule and P. E. Drawz, "Heart rate and blood pressure: any possible implications for management of hypertension?," *Curr. Hypertens. Rep.*, vol. 14, no. 6, pp. 478–84, Dec. 2012.

[11] R. M. T. Laukkanen, P. K. Virtanen, R. M. T. Lauk K Anen, and P. K. Virtan En, "Heart rate monitors: State of the art," *J. Sports Sci.*, pp. 3–7, 1998.

[12] R. Macleod and B. Birchler, “ECG Measurement and Analysis,” 2014.

[13] J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, Mar. 2007.

[14] K.-M. Chen, D. Misra, H. Wang, H.-R. Chuang, and E. Postow, “An X-Band Microwave Life-Detection System,” *IEEE Trans. Biomed. Eng.*, vol. BME-33, no. 7, pp. 697–701, Jul. 1986.

[15] Changzhi Li, Jun Ling, Jian Li, and Jenshan Lin, “Accurate Doppler Radar Noncontact Vital Sign Detection Using the RELAX Algorithm,” *IEEE Trans. Instrum. Meas.*, vol. 59, no. 3, pp. 687–695, Mar. 2010.

[16] J. Tu and J. Lin, “Fast Acquisition of Heart Rate in Noncontact Vital Sign Radar Measurement Using Time-Window-Variation Technique,” *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 112–122, Jan. 2016.

[17] R. K. Singh, A. Sarkar, and C. S. Anoop, “A health monitoring system using multiple non-contact ECG sensors for automotive drivers,” in *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, 2016, pp. 1–6.

[18] H. Zhao, H. Hong, L. Sun, Y. Li, C. Li, and X. Zhu, “Noncontact Physiological Dynamics Detection Using Low-power Digital-IF Doppler Radar,” *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1780–1788, Jul. 2017.

[19] Changzhan Gu, Changzhi Li, Jenshan Lin, Jiang Long, Jiangtao Huangfu, and Lixin Ran, “Instrument-Based Noncontact Doppler Radar Vital Sign Detection System Using Heterodyne Digital Quadrature Demodulation Architecture,” *IEEE Trans. Instrum. Meas.*, vol. 59, no. 6, pp. 1580–1588, Jun. 2010.

[20] E. F. Greneker, “Radar sensing of heartbeat and respiration at a distance with applications of the technology,” *Radar Syst. (RADAR 97)*, vol. 1997, no. 449, pp. 150–154, 1997.

[21] S. Bakhtiari *et al.*, “Compact Millimeter-Wave Sensor for Remote Monitoring of Vital Signs,” *IEEE Trans. Instrum. Meas.*, vol. 61, no. 3, pp. 830–841, Mar. 2012.

[22] Ji-Jer Huang, Sheng-I Yu, Hao-Yi Syu, and A. R. See, “The non-contact heart rate measurement system for monitoring HRV,” in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013, pp. 3238–3241.

[23] Nanfei Sun, M. Garbey, A. Merla, and I. Pavlidis, “Imaging the Cardiovascular Pulse,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, vol. 2, pp. 416–421.

[24] M. Garbey, N. Sun, A. Merla, and I. Pavlidis, “Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery,” *IEEE Trans. Biomed. Eng.*, vol. 54, no. 8, 2007.

[25] C. B. Pereira, X. Yu, M. Czaplik, R. Rossaint, V. Blazek, and S. Leonhardt, “Remote monitoring of breathing dynamics using infrared thermography.,” *Biomed. Opt. Express*, vol. 6, no. 11, pp. 4378–94, Nov. 2015.

[26] C. Takano and Y. Ohta, “Heart rate measurement based on a time-lapse image,” *Med. Eng. Phys.*, vol. 29, no. 8, pp. 853–857, Oct. 2007.

[27] G. Balakrishnan, F. Durand, and J. Guttag, “Detecting Pulse from Head Motions in Video,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[28] Y. Sun, S. Hu, V. Azorin-Peris, R. Kalawsky, and S. Greenwald, “Noncontact imaging photoplethysmography to effectively access pulse rate variability,” *J. Biomed. Opt.*, vol. 18, no. 6, p. 61205, Oct. 2012.

[29] 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.

[30] B. Martinez, M. F. Valstar, X. Binefa, and M. Pantic, “Local Evidence Aggregation for Regression Based Facial Point Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 55, pp. 1149–1163, 2013.

[31] F. Bousefsaf, C. Maaoui, and A. Pruski, “Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate,” *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 568–574, Nov. 2013.

[32] M. Kumar, A. Veeraraghavan, and A. Sabharwal, “DistancePPG: Robust non-contact vital signs monitoring using a camera,” *Biomed. Opt. Express*, vol. 6, no. 5, p. 1565, May 2015.

[33] C. H. Antink, H. Gao, C. Brüser, and S. Leonhardt, “Beat-to-beat heart rate estimation fusing multimodal video and sensor data,” *Biomed. Opt. Express*, vol. 6, no. 8, p. 2895, Aug. 2015.

[34] S. Park, M. J. Won, D. W. Lee, and M. Whang, “Non-contact measurement of heart response reflected in human eye,” *Int. J. Psychophysiol.*, vol. 123, pp. 179–198, Jan. 2018.

[35] M. Rubinstein *et al.*, “Eulerian video magnification for revealing subtle changes in the world,” *ACM Trans. Graph.*, vol. 31, no. 4, 2016.

[36] K. Alghoul, S. Alharthi, H. Al Osman, and A. El Saddik, “Heart Rate Variability Extraction From Videos Signals: ICA vs. EVM Comparison,” *IEEE Access*, vol. 5, pp. 4711–4719, 2017.

[37] M. Z. Poh, D. J. McDuff, and R. W. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, 2011.

[38] L. A. M. Aarts *et al.*, “Non-contact heart rate monitoring utilizing camera

photoplethysmography in the neonatal intensive care unit — A pilot study," *Early Hum. Dev.*, vol. 89, no. 12, pp. 943–948, Dec. 2013.

[39] O. Gupta, D. McDuff, and R. Raskar, "Real-Time Physiological Measurement and Visualization Using a Synchronized Multi-camera System," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, pp. 312–319.

[40] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, p. I-511-I-518.

[41] T. M. Mahmoud, "A New Fast Skin Color Detection Technique," *Int. J. Comput. Inf. Eng.*, vol. Vol:2, No:, 2008.

[42] J. M. Saragih, S. Lucey, and J. F. Cohn, "Deformable Model Fitting by Regularized Landmark Mean-Shift," *Int. J. Comput. Vis.*, vol. 91, no. 2, pp. 200–215, Jan. 2011.

[43] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann Publishers Inc., pp. 674–679, 1981.

[44] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, and Jedrzej Nowak, "Measuring pulse rate with a webcam — A non-contact method for evaluating cardiac activity - Semantic Scholar," in *Federated Conference on Computer Science and Information Systems*, 2011.

[45] M. Dantone, J. Gall, G. Fanelli, L. Van Gool, and E. Zurich, "Real-time Facial Feature Detection using Conditional Regression Forests," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[46] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace: an open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[47] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams: a Factorization Method|Part 3 Detection and Tracking of Point Features," 1991.

[48] C. Feichtenhofer and A. Pinz, "Spatio-temporal good features to track," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013.

[49] T. Percy Driver, S. Sundaram, G. Khandelwal, and M. Sahasrabudhe, "Systems And Methods For Patient Identification Using Mobile Face Recognition," 28-May-2009.

[50] R. W. Frischholz and U. Dieckmann, "BiolD: a multimodal biometric identification system," *Computer (Long. Beach. Calif).*, vol. 33, no. 2, pp. 64–68, 2000.

[51] H. D. Wactlar, T. Kanade, M. A. Smith, and S. M. Stevens, "Intelligent access to digital video: Informedia project," *Computer (Long. Beach. Calif).*, vol. 29, no. 5, pp. 46–52,

May 1996.

[52] C. I. Christodoulou *et al.*, “Multi-feature texture analysis for the classification of carotid plaques,” *IJCNN’99. Int. Jt. Conf. Neural Networks. Proc.*, vol. 5, no. February, pp. 3591–3596, 1999.

[53] R. Brunelli and T. Poggio, “Face recognition: features versus templates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.

[54] T. Sakai, M. Nagao, and S. Fujibayashi, “Line extraction and pattern detection in a photograph,” *Pattern Recognit.*, vol. 1, no. 3, pp. 233–248, Mar. 1969.

[55] D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999.

[56] Y.-J. Tu, C.-C. Kao, and H.-Y. Lin, “Human Computer Interaction Using Face and Gesture Recognition.”

[57] U. Weidenbacher, G. Layher, P. Bayerl, and H. Neumann, “LNAI 4021 - Detection of Head Pose and Gaze Direction for Human-Computer Interaction.”

[58] P. Sajjacholapunt and L. J. Ball, “The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness.,” *Front. Psychol.*, vol. 5, p. 166, 2014.

[59] R. Stiefelhagen, “Tracking focus of attention in meetings,” in *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 273–280.

[60] P. Comon, “SIGNAL PROCESSING Independent component analysis, A new concept?*,” *Signal Process. Comon / Signal Process.*, vol. 36, no. 36, pp. 28–314, 1994.

[61] A. Talwar and Y. Kumar, “Machine Learning: An artificial intelligence methodology,” *Int. J. Eng. Comput. Sci.*, vol. 2, no. 12, pp. 3400–3404, 2013.

[62] B. D. Lucas and T. Kanade, “An Iterative Image Registration Technique with an Application to Stereo Vision.”

[63] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, “An advanced detrending method with application to HRV analysis,” *IEEE Trans. Biomed. Eng.*, vol. 49, no. 2, pp. 172–175, Feb. 2002.

[64] J. F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proc. F Radar Signal Process.*, vol. 140, no. 6, p. 362, 1993.

[65] R. Zaman, C. H. Cho, K. Hartmann-Vaccarezza, T. N. Phan, G. Yoon, and J. W. Chong, “Novel fingertip image-based heart rate detection methods for a smartphone,” *Sensors (Switzerland)*, vol. 17, no. 2, pp. 1–12, 2017.

[66] F. Song, Z. Guo, and D. Mei, “Feature Selection Using Principal Component Analysis,” in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, 2010, pp. 27–30.

[67] J. G. Cleary and L. E. Trigg, “K*: An Instance-based Learner Using an Entropic Distance Measure,” New Zealand.

[68] N. Horning, “Random Forests : An algorithm for image classification and generation of continuous fields data sets,” in *International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences*, 2010.

[69] K. Fawagreh, M. Medhat Gaber, E. Elyan, and M. M. Gaber, “Random forests: from early developments to recent advancements,” *Syst. Sci. Control Eng. An Open Access J.*, vol. 2, no. 1, pp. 602–609, 2014.

[70] N. H. Mohd Sani, W. Mansor, K. Y. Lee, N. Ahmad Zainudin, and S. A. Mahrim, “Determination of heart rate from photoplethysmogram using Fast Fourier Transform,” in *2015 International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, 2015, pp. 168–170.

Appendix A- Ethical Approval

File Number: H02-17-13 **Date (mm/dd/yyyy):** 04/06/2017

Université d'Ottawa
Bureau d'éthique et d'intégrité de la recherche

University of Ottawa
Office of Research Ethics and Integrity

Ethics Approval Notice

Health Sciences and Science REB

Principal Investigator / Supervisor / Co-investigator(s) / Student(s)

First Name	Last Name	Affiliation	Role
Hussein	Al Osman	Engineering / Computer Engineering	Supervisor
Hamideh	Ghanadian	Engineering / Electrical Engineering	Student Researcher

File Number: H02-17-13

Type of Project: Master's Thesis

Title: Non-contact extraction of heart rate from video

Approval Date (mm/dd/yyyy)	Expiry Date (mm/dd/yyyy)	Approval Type
04/06/2017	04/05/2018	Approval

Special Conditions / Comments:
N/A