DEGREE PROJECT IN COMPUTER SCIENCE AND ENGINEERING, SECOND CYCLE, 30 CREDITS

STOCKHOLM, SWEDEN 2017

Remote heart rate estimation by evaluating measurements from multiple signals

KRISTOFFER UGGLA LINGVALL

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTER SCIENCE AND COMMUNICATION

Remote heart rate estimation by evaluating measurements from multiple signals

KRISTOFFER UGGLA LINGVALL

Master in Computer Science
Date: June 26, 2017
Supervisor: Iolanda Leite
Examiner: Danica Kragic
Swedish title: Pulsmätning på avstånd genom viktning av mätvärden från olika signaler
School of Computer Science and Communication

Abstract

Heart rate can say a lot about a person's health. While most conventional methods for heart rate measurement require contact with the subject, these are not always applicable. In this thesis, a non-invasive method for pulse detection is implemented and analyzed. Different signals from the color of the forehead—including the green channel, the hue channel and different ICA and PCA components—are inspected, and their resulted heart rates are weighted together according to the significance of their FFT peaks. The system is tested on videos with different difficulties regarding the amount of movement and setting of the scene. The results show that the approach of weighting measurements from different signals together has great potential. The system in this thesis, however, does not perform very well on videos with a lot of movement because of motion noise. Though, with better, less noisy signals, good results can be expected.

Sammanfattning

En människas puls säger en hel del om dennes hälsa. För att mäta pulsen används vanligtvis metoder som vidrör människan, vilket ibland är en nackdel. I det här examensarbetet tas en metod för pulsmätning på avstånd fram, som endast använder klipp från en vanlig videokamera. Färgen i pannan mäts och utifrån den genereras flera signaler som analyseras, vilket resulterar i olika mätvärden för pulsen. Genom att värdera dessa mätvärden med avseende på hur tydliga signalerna är, beräknas ett viktat medelvärde som ett slutgiltigt estimat på medelpulsen. Metoden testas på videoklipp med varierande svårighetsgrad, beroende på hur mycket rörelser som förekommer och på vilket avstånd från kameran försökspersonen står. Resultaten visar att metoden har mycket god potential och att man kan man förvänta sig fina resultat med bättre, mindre brusiga signaler.

Contents

Acronyms

bpm beats per minute. 5, 7, 10, 13, 19, 20, 23, 25, 27, 28, 31, 35

BSS blind source separation. 13

CPU central processing unit. 37

CWT continuous wavelet transform. 7

DFT discrete Fourier transform. 13

FFT fast Fourier transform. 3, 5, 7, 13, 14, 19, 20, 27, 28

FPS frames per second. 21

GPU graphics processing unit. 37

HR heart rate. 16, 23, 30

HSI hue, saturation, and intensity. 6

HSL hue, saturation, and lightness. 6

HSV hue, saturation, and value. 6, 8, 11, 12, 16, 18, 20, 28

ICA independent component analysis. 3–5, 8, 13, 18, 20, 28

JADE Joint Approximation Diagonalization of Eigen-matrices. 4, 18

KLT Kanade–Lucas–Tomasi. 11, 16, 34

PCA principal component analysis. 3, 5, 8, 13, 18, 20, 28

POS Plane-Orthogonal-to-Skin. 2

PPG photoplethysmography. 5, 7, 9, 10, 12, 13, 16, 18, 24, 34

RGB red, green, and blue. 4, 6, 8, 11, 12, 16–18, 20, 28

RMSE root-mean-square error. 5–7, 23, 25, 31

ROI region of interest. 4, 6, 15, 16, 34, 36, 37

SIFT scale-invariant feature transform. 37

SNR signal-to-noise ratio. 6

Chapter 1

Introduction

A person's heart rate can say a lot about the individual. It gives an indication of the physical health of the person, and it tells how strained he or she is, for example. Moreover, it can show how stressed or nervous the person is, which can be used in sports contexts to show how stressed an athlete is in a given situation.

Imagine a football player the moment after the referee has given him the green light to shoot the deciding penalty in the world cup final, or a golf player the second before a putt to win a big tournament. How would a high pulse impact the performance in these situations? Or what would a high heart rate say about the chance that a poker player is bluffing when moving all-in during a crucial hand?

Most heart rate measurement techniques use some device that demands physical contact with the subject's body, which in these situations is problematic. This thesis addresses this problem of measuring sports players' heart rate by using non-intrusive sensing technology.

In the world, small changes constantly happen that the naked eye of a human cannot see. The reason for this is that many of the signals fall below our visual system's limited spatiotemporal sensitivity [37]. One example of what a human cannot see is how the skin color varies due to blood circulation. Machines can detect these small color variations and measure them [26], and this information can be analyzed to calculate the heart rate.

Many attempts have been made earlier to measure human heart rate without the need of intrusive tools, with different level of success. While most works have focused on still subjects with good results, the works where moving subjects were also targeted did not have as

high performance [13, 22, 25, 26, 33]. What mainly differ among these methods is what color channel that is analyzed and how this signal is processed. The proposed method in this thesis attempts to measure heart rate using multiple signals that have produced good results before and weighting these measurements together based on their significance. By using a method like this, the measurement error can potentially be reduced to a level where the system can be used in real-world environments (e.g. outdoor conditions, moving subjects, etc.).

1.1 Problem statement

The central question addressed in this thesis is: "Can the performance of remote heart rate measuring on moving subjects be improved by relying on multiple signals gathered from image data?"

1.2 Contribution

The two main contributions of this work include:

- A novel method for evaluating and weighting together multiple signals' heart rate measurements, that aims to improve the accuracy of existing non-intrusive heart rate monitoring systems.
- A performance evaluation of the proposed method on image data collected in outdoor environments (previous research has mostly been evaluated on datasets collected in indoor environments) of three subjects in different conditions (variation in camera distance, lighting and head movements).

Even though the method is only tested on a set of comparatively basic signals (see which in Section 4.2.2), the expectation is that the method can be used on other more advanced signals as well, and still improve the performance. For example, Wang et al. [36] developed an algorithm called POS (Plane-Orthogonal-to-Skin) for heart rate estimation which produces a well-performing signal, that could be used.

1.3 Delimitation

This thesis focuses on detecting the heart rate by analyzing the color change of the face. Other data, such as thermal data and acceleration data, that can also be used for pulse detection are not utilized because in many real-world applications it is unrealistic to collect this.

Another delimitation is that the subjects in this experiment are all white men, and further research is needed to verify whether our methods apply to other ethnic groups. However, earlier studies [28, 33] have shown that the skin tone has a notable impact on the accuracy of heart rate estimation.

1.4 Thesis outline

The rest of the thesis is outlined as follows. The next chapter (Chapter 2) gives an overview of work done before in the area and in Chapter 3, the reader is provided the background theory needed to understand the subject of remote heart rate monitoring. In Chapter 4, the proposed method to estimate the average heart rate and the experimental setup are explained thoroughly. The results of the experiment are then presented in Chapter 5. In the final chapter (Chapter 6), the results are discussed, and conclusions are drawn before possible work to be done in the future on the system is suggested.

Chapter 2

Related work

In this section, previous work in the field of remote heart rate estimation is summarized. The first studies on the subject are presented, followed by other important research in the field. The chapter is concluded with a section where it is established what work this thesis will be based on.

2.1 Early studies

In 2010, Poh et al. [26] demonstrate that it is possible to measure the heart rate of a human by only observing the small color variations of the skin. An automatic face tracker is used to find the face in the image, and the region of interest is selected as 60 percent of the width and the full height of the face. The color of the area is then separated into the three RGB channels, where the pixels are spatially averaged to generate a red, a green and a blue measurement for each frame, t, called $x_1(t)$, $x_2(t)$ and $x_3(t)$. The measurements are averaged as in Equation 2.1.

$$x_i'(t) = \frac{x_i(t) - \mu_i}{\sigma_i} \qquad \text{for } i = 1, 2, 3 \tag{2.1}$$

where μ_i and σ_i are the mean and standard deviation of x_i, respectively.

An ICA algorithm called Joint Approximation Diagonalization of Eigen-matrices (JADE) [3] is used to decompose the raw signals into three independent source signals. The second ICA component is then selected because it typically contains a strong plethysmographic sig-

nal, and the power spectrum of it is retrieved using FFT. The heart rate is estimated as the highest power of the spectrum within the range $[0.75, 4.00]$ Hz, corresponding to 45–240 bpm. To remove artifacts, results diverging more than 12 bpm from the last measurement (one second before) are rejected and the second highest power that meets the requirements is selected instead.

The results show that the approach of applying ICA to retrieve the PPG signal works well, and a root-mean-square error (RMSE) of 2.29 is achieved on subjects sitting still. This can be compared to an RMSE of 6.00 if the heart rate is obtained just by analyzing the raw data. The RMSEs for when the subjects make small movements in the frame are 4.63 and 19.36, with and without using ICA, respectively.

A year later, Poh et al. [25] improve their method by adding temporal filters before and after applying ICA, which results in an RMSE of 1.24 on subjects sitting still. An alternative for ICA is then presented during the same year by Lewandowska et al. [20], who instead uses PCA. Their results show that PCA performs in line with ICA but is less computationally complex.

2.2 Eulerian video magnification

Another technique to make use of the "invisible" changes in videos is to amplify them so that a human can see them. One method to do this, demonstrated by Wu et al. [37], is called Eulerian video magnification.

In Eulerian video magnification, the video sequence is first decomposed into different spatial frequency bands. The frames of the video are spatially low-pass filtered and downsampled for computational efficiency before temporal processing is performed on each of these bands. The time series corresponding to the value of a pixel in a frequency band is considered, and a band-pass filter is used to extract the frequency bands of interest. A possible frequency range for pulse detection could be $[0.75, 4.00]$ Hz. The temporal processing is uniform for all spatial levels, and for all pixels within each level. The extracted band-passed signals are then multiplied by α, the magnification factor, and the magnified signals are added to the original signals to obtain the final result.

2.3 HR measurement in realistic situations

Even though the results in the works presented previously were promising, the methods mainly performed under well-controlled conditions. In 2014, Li et al. [22] present a method for heart rate estimation that works better when the video illumination and the subjects' motions cannot be controlled. By analyzing the background, the authors try to remove color changes in the face appearing because of variations in lighting, to get better results. To improve the method further, the region of interest is not a square around the face like in most approaches before, but a narrower semicircle built up using feature points in the area under the eyes, down to the chin.

The new method is evaluated and compared to other methods on a public database called MAHNOB-HCI [30], which according to the authors could be considered as rather difficult cases for pulse detection. Their method yields an RMSE of 7.62 on the dataset. In comparison, the improved method by Poh et al. [25] yields an RMSE of 13.60 on the same dataset.

Another work addressing the problem of remote heart rate estimation under more realistic circumstances, is DistancePPG by Kumar et al. [18]. Their approach is to track different regions of the face and to combine the skin-color change signals from these into a single signal with improved signal-to-noise ratio (SNR). The signals are combined using a weighted average, where the weights depend on the blood amount and light intensity in the face region.

2.4 Alternative color spaces

While the papers presented before used the RGB color space in their methods, other color spaces can be used for heart rate estimation as well. In 2015, Tsouri et al. [33] investigate the benefits of using alternative color spaces for this purpose. Besides all three color channels of RGB, all parameters of HSV, HSL, HSI, XYZ, CIE XYZ and CIE YUV are tested and the performance is analyzed. It is shown that the hue parameter has the best results, followed by Y of CIE XYZ, U of CIE YUV and the green channel of RGB.

In another paper, Bousefsaf et al. [1] show a different way to obtain the heart rate of a moving subject in a webcam video. A face tracker

is used to get the interesting region of the image, and a skin detector is used to get the pixels containing photoplethysmography (PPG) information that could be used to measure the heart rate. The colors are converted to a color space called CIELUV (parameters L*, u* and v*), where the PPG fluctuations have the most impact on the u* channel. In contrast to many other methods, the Fourier transform is not used in favor of the continuous wavelet transform (CWT). According to the authors, CWT can detect rapid frequency changes in time due to its variable window width, which the Fourier transform cannot. The method works well and scores an average RMSE of 2.33 on moving subjects.

2.5 Other cameras

There are also other non-invasive methods for pulse detection, which do not examine the color of the skin. For example, the heart rate can be extracted by studying the temperature of the skin. Pavlidis et al. [10] present a method on how to analyze thermal imagery for this purpose, where the results show that the detected heart rate agrees to the reference heart rate with 98 percent. The measurements are made with the knowledge that the correct heart rate is in the 60–100 bpm interval. Other limitations of this method are that it requires video sequences of at least two minutes and that the subjects are sitting still in front of the camera. Consequently, this method cannot be applied directly to the problem of this thesis.

Infrared cameras can also be used for heart rate estimation. Qiu et al. [27] use an Intel RealSense 3D camera to do this with good results. A combination of the Viola–Jones face detector and a supervised method is used on the infrared and the depth data to track facial landmarks in the image. These landmarks are then used to extract the cheek regions of the face. The averaged temporal data is denoised using a global self-similarity filter [6] before the signal is temporal filtered, detrended and smoothed. Then, the signal is band-passed filtered to $[0.7, 4.0]$ Hz, using a Hamming window-based method, corresponding to 42–240 bpm. The FFT of the signal is then inspected for the maximum power, which is assumed to be the pulse frequency f_p. The heart rate is, finally, calculated as $60 f_p$. The method is evaluated on their own dataset containing 10 videos of still subjects in varying illumination and has an RMSE of 3.66.

2.6 Extension of related work

In this thesis, a combination of earlier presented methods is used to improve accuracy. The approach suggested by Poh et al. [26] is used as a starting point, which includes the way a face is tracked and how the face colors are averaged. Their approach with ICA on RGB is also used, along with the green channel of RGB and the hue channel of HSV proved to be good by Tsouri et al. [33], and the PCA on RGB shown by Lewandowska et al. [20]. As an addition, both ICA and PCA are used on HSV to generate six additional signals.

A Fourier transform is applied to the resulting 14 signals, and the heart rates are extracted from the frequency domain in a similar manner as Poh et al. [26] did. A novel method is then used to calculate the significance of each measurement before they are weighted together into a final heart rate estimate.

Chapter 3

Background

In a remote heart rate monitoring system, different components are needed. The technique used to estimate a heart rate remotely from camera images is called photoplethysmography and is described in the first section of this chapter. To be able to use this technique reliably, different components are needed. To begin with, the face in the video has to be detected and tracked. The color of the face region then has to be inspected so it has to be represented in a proper way. This color information then has to be processed and analyzed in order to get a heart rate estimate. This chapter describes the background methodologies relevant for the proposed remote heart rate measuring system.

3.1 Photoplethysmography

Photoplethysmography (PPG) is a simple technique used to detect volumetric changes in different parts of the body by examining the color of the skin [34]. Most of these changes happen because of fluctuations in blood volume in the organ, and photoplethysmography can thus be seen as a measurement of the pulse wave traveling through the body.

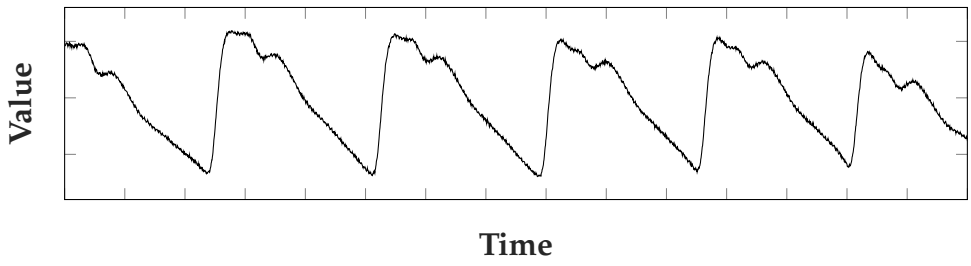


Figure 3.1: An example of a PPG signal [4].

An example of a PPG signal can be seen in Figure 3.1. By studying the time between two consecutive peaks, Δt, the heart rate (in bpm) can be calculated as $\frac{60}{\Delta t}$.

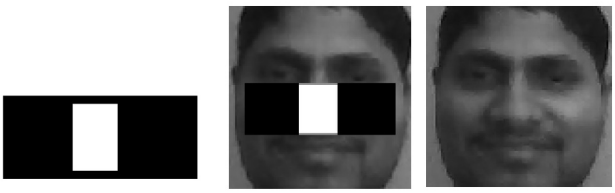
3.2 Face detection and tracking

A face tracking algorithm typically contains two steps. First, the face has to be located in the image, and then some feature points in it have to be followed. Here, two of the most commonly used methods to achieve this are described.

3.2.1 Detection with Viola–Jones

One of the most known and commonly used face detection algorithms is Viola–Jones [35]. The idea behind the algorithm is simple: a window is slid over the image, in which a classifier decides whether it contains an object (in this case a face) or not. The classifier, which consists of several simpler classifiers that are applied successively until a candidate is rejected or all the stages are passed, is trained on a set of face and non-face images. Haar-like features, of which examples can be seen in Figure 3.2, are used to make the decisions. To be able to find faces of not only one size, the classifier is designed so it can be resized and slid over the image repeatedly in different sizes.

(a) A Haar feature that looks like the eye region which is darker than the upper cheeks is applied onto a face.



(b) A Haar feature that looks like the bridge of the nose is applied onto a face.

Figure 3.2: Haar features explained [35].

Once a face has been detected in the image, the known position can be used in the next frame to speed up the algorithm and to track the face. This is simply done by only searching for a face in regions close to the previous position. If a face is not found, the algorithm should preferably check the entire frame for a face.

3.2.2 Tracking with KLT

Kanade–Lucas–Tomasi (KLT) feature tracking is a fast algorithm to track feature points in a video and is often used for face tracking [23, 31]. KLT starts by detecting good features to track, which often are corner-looking points (intersection of two edges), by inspecting the minimum eigenvalue of every 2×2 gradient matrix of the image. A Newton-Raphson method is then used to track the points by minimizing the difference between two frames.

3.3 Color spaces

There are different models for representing colors. Two of the most used models are RGB and HSV, which are used in this thesis.

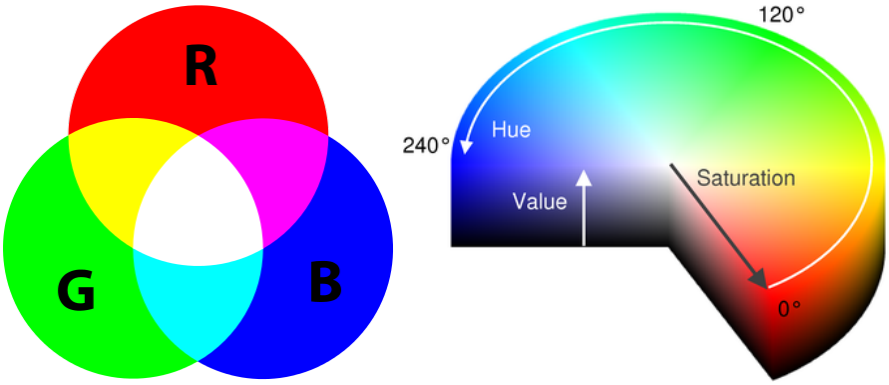
3.3.1 RGB

RGB is a very commonly used color representation model, having three parameters—R (red), G (green), and B (blue)—each of them with a value in range $[0.0, 1.0]$. It is an additive color model, which means that the resulting color is the one where the parameters are mixed [12]. For example, black is represented by the triplet $(0, 0, 0)$, blue by $(0, 0, 1)$ and white by $(1, 1, 1)$. The RGB color model is visualized in Figure 3.3a and is used throughout this project.

3.3.2 HSV

HSV (hue, saturation, and value) is a color model that represents an RGB color in a cylindrical coordinate system [12], to make the representation more intuitive. The hue parameter is the angular dimension and represents the color with red at 0°, green at 120° and blue at 240°. Saturation, ranging from zero to one, states how saturated the color is.

Finally, value describes the brightness in a 0–1 range. A visual explanation of the HSV color model can be seen in Figure 3.3b.



(a) The RGB color model. **(b)** The HSV color model.

Figure 3.3: Visualizations of the RGB and the HSV color models.

For remote PPG, HSV has some advantages over RGB [33]. The color information of an object, which is the information of interest, is found solely in the hue parameter. This parameter does not depend on the brightness of the object and should be invariant to light changes. In RGB, this information is distributed across all three primary colors.

To convert a color in RGB to HSV, Equation 3.1 is used [2].

$$
\begin{aligned}
C_{min} &= \min(R, G, B) \\
C_{max} &= \max(R, G, B) \\
\Delta &= C_{max} - C_{min} \\
H &= \begin{cases} 0^\circ & \text{if } \Delta = 0 \\ 60^\circ \cdot \left(\frac{G-B}{\Delta} \bmod 6\right) & \text{if } C_{max} = R \\ 60^\circ \cdot \left(\frac{B-R}{\Delta} + 2\right) & \text{if } C_{max} = G \\ 60^\circ \cdot \left(\frac{R-G}{\Delta} + 4\right) & \text{if } C_{max} = B \end{cases} \\
S &= \begin{cases} 0 & \text{if } C_{max} = 0 \\ \frac{\Delta}{C_{max}} & \text{if } C_{max} \neq 0 \end{cases} \\
V &= C_{max}
\end{aligned}
\tag{3.1}
$$

3.4 Component analysis

When dealing with camera footage used for heart rate estimation, motion and illumination artifacts are substantial obstacles [13, 17, 22, 24]. To overcome these, blind source separation (BSS) can be used to separate the PPG signal from other signals influencing the color change of the skin. Two of the most commonly used methods to do this are independent component analysis (ICA) and principal component analysis (PCA). ICA decomposes mixtures of source signals into components that are, if not completely independent, as independent as possible [5]. PCA, on the other hand, identifies the principal components by using an orthogonal matrix composed by the eigenvectors of the original variables' covariance matrix [5].

3.5 Fourier transform

A Fourier transform is used to transfer a signal from the spatial (time) domain to the frequency domain. A well-used algorithm to compute the discrete Fourier transform (DFT) is FFT (fast Fourier transform). For heart rate detection, the frequency domain of a PPG signal can be inspected for peaks in the region corresponding to a reasonable heart rate. A particular frequency in Hz, f, is easily translated to a heart rate, p, as $p = 60f$.

Something to consider when using FFT for heart rate estimation is the frequency resolution. That is, how wide the bins are that the frequencies of the PPG signal are placed in. The bin size, b, in bpm, for FFT, is defined as $b = \frac{60}{L_s}$, where L_s is the video length in seconds. This means, for example, that a video of 10 seconds only would lead to a resolution of 6 bpm. A visualization of how the FFT resolution depends on the length of the video can be seen in Figure 3.4. If one, for example, would like to have a resolution of 2 bpm, that would require 30 seconds of video.

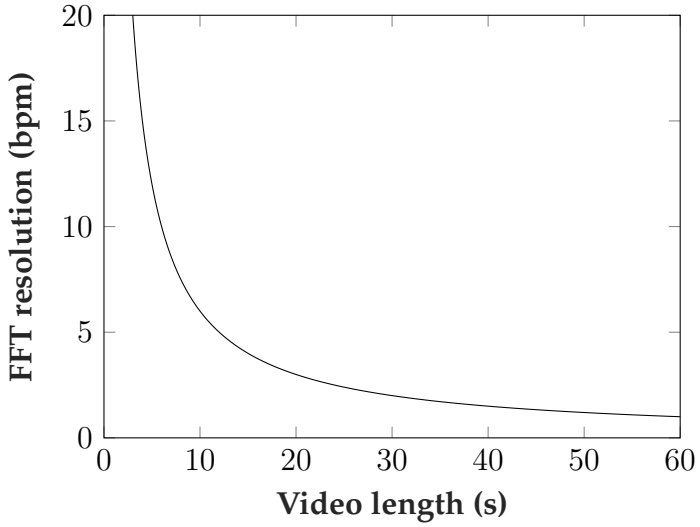


Figure 3.4: The frequency resolution for FFT depending on the length of the video.

To overcome the problem with a low frequency resolution, zero-padding can be used. The signal is then padded with zeros up to the length required to achieve a desirable resolution. A problem with this, though, is that zero-padding "washes out" the signal in the power spectrum. This is the case since the added part of the signal does not contain any data, which means that if a signal gets its length doubled by zero-padding, only half of it contributes to the FFT result.

Chapter 4

Method

In this section, the conducted experiment is explained in detail. First, the proposed method is described thoroughly. The system consists of three main steps—ROI detection, signal processing, and heart rate estimation—where each main step is divided into a handful of smaller steps. An overview of the system architecture can be seen in Figure 4.1. Subsequently, the experimental setup is described including the video setup and how the ground truth was measured, followed by a definition of the resulted dataset. Finally, it is explained how the performance was measured.

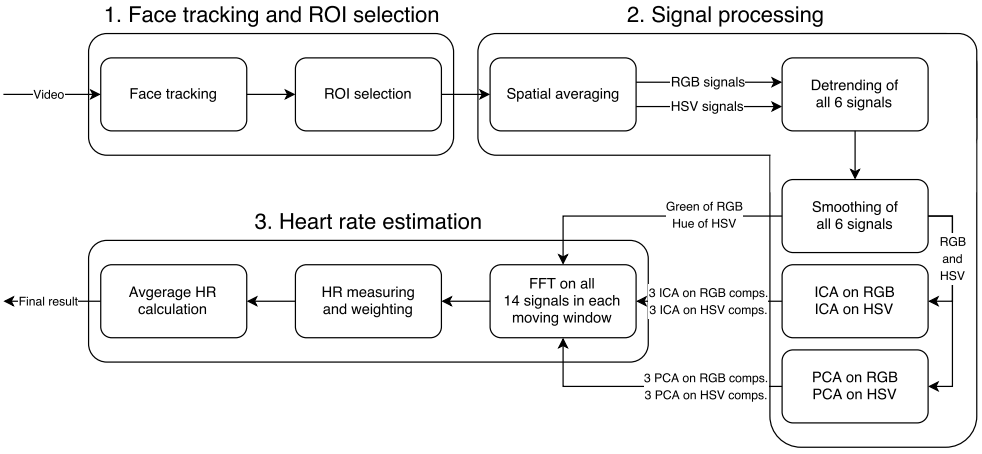


Figure 4.1: A schematic diagram of the system architecture.

4.1 Face tracking and ROI selection

To detect the face in the video, MATLAB's built-in Viola–Jones algorithm [7] was used, and the KLT algorithm [32] was used to track it. If the face detection was not able to locate a face in a video frame, the last know position of it was used instead.

When the face had been located and bounded by a box, the region of interest (ROI) was selected. According to earlier studies [8, 15, 19], the forehead has proven to be the part of the face containing the strongest PPG signal and was therefore selected. A rather simple method to determine the forehead region was used, where the center 60 percent of the width and the second uppermost tenth of the height was selected (see Figure 4.2).

Figure 4.2: A frame from one of the videos (T2) with subject S1. The detected face is the big, outer box and the ROI is the smaller box inside it. The ground truth HR can be seen on the watch face.

4.2 Signal processing

The RGB values of the pixels inside the ROI were spatially averaged, excluding the 5 percent highest and lowest values, which resulted in three raw signals, $x_1(t)$, $x_2(t)$ and $x_3(t)$ for each time frame, t. From the raw RGB signals, the HSV values were calculated using Equation 3.1, producing three more raw signals, $x_4(t)$, $x_5(t)$ and $x_6(t)$.

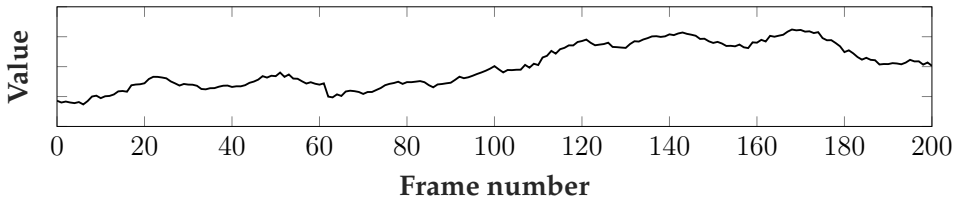
4.2.1 Detrending and smoothing

Considering that only small color changes are of interest, the signals were detrended to eliminate larger fluctuations caused mainly by an alternation of illumination. This was done by subtracting a 30-frame (one second) moving average from the signals as shown in Equation 4.1, resulting in six new signals, $x'_1(t), x'_2(t), \ldots, x'_6(t)$.

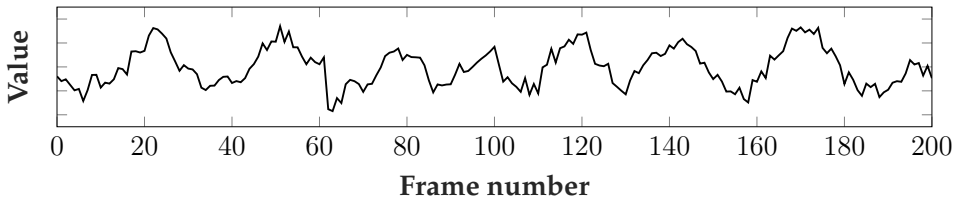
$$x'_i(t) = x_i - \frac{1}{31} \sum_{k=\max(1,t-15)}^{\min(t+15,N)} x_k \qquad \text{for } i = 1, 2, 3 \tag{4.1}$$

where t is the current frame and N is the total number of frames in the video sequence. An example of detrending, demonstrated on the green RGB channel, x_2, can be seen in Figure 4.3b.

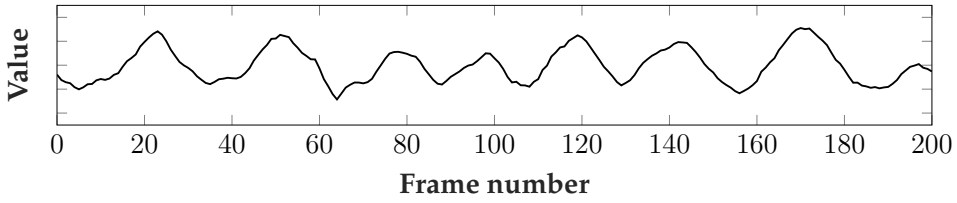
To remove high-frequency noise and to make the signals clearer, the detrended signals were smoothed using a five-frame moving average. The result of this is shown in Figure 4.3c.



(a) Raw signal.



(b) Detrended signal.



(c) Detrended and smoothed signal.

Figure 4.3: Signal processing steps shown on the green channel of RGB (x_2). The video was a close-up of the forehead, recorded indoors.

4.2.2 Generating signals

One of the biggest problems with remote pulse detection on moving subjects is artifacts caused by motion, which make it difficult to recover the PPG signal from the raw signal. Many different approaches have been tested by others to overcome this (see Chapter 2 for examples), and no method is optimal and outperforms the other methods in all scenarios. Hence, the suggested method in this thesis is to use multiple signals and weight the heart rate measurements from these together.

Because of this, 12 further signals were calculated using ICA and PCA on the detrended and smoothed color channels. Altogether, 14 signals were used:

- The green channel of RGB (x'_2).
- The hue channel of HSV (x'_4).
- All three components generated using ICA on RGB.
- All three components generated using ICA on HSV.
- All three components generated using PCA on RGB.
- All three components generated using PCA on HSV.

The green channel of RGB and the hue channel of HSV were selected because it has been shown before that these are good signals for PPG extraction [33]. For ICA, the JADE algorithm was used, which has been proven to be a suitable method for this purpose [5].

4.3 Heart rate estimation

After the signals had been extracted, heart rates for all of them had to be measured and weighted together. How this was done is described in this section.

4.3.1 Moving window FFT

To get the heart rate corresponding to a particular signal, FFT was used. Since the heart rate might vary over time, causing the pulse to get spread out in the power spectrum, a moving window was used. The window was 10 seconds long and shifted forward with one second. Different sizes of the moving window, as well as how much it shifted, were tested beforehand on a subset of the data, which led to these values. The results of these tests are, however, not presented in this thesis.

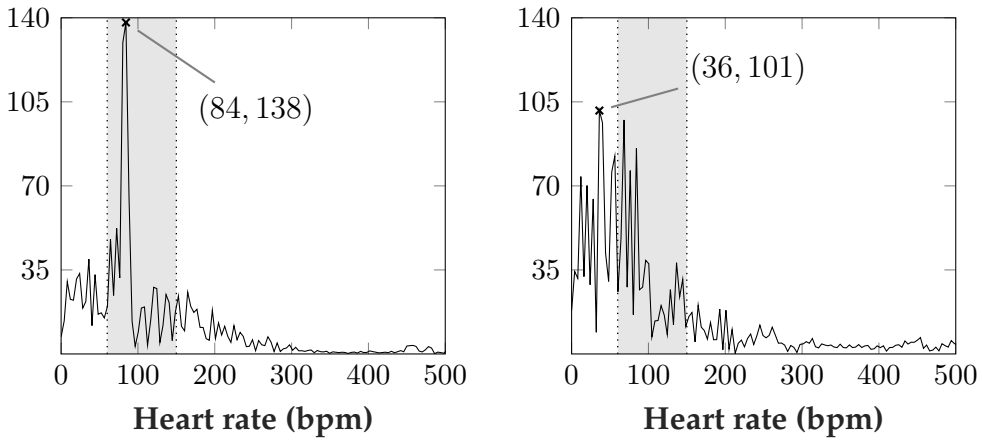
Each signal was zero-padded to a length of 450 frames, to get a frequency resolution of 4 bpm (see Section 3.5 why this is the case). The resolution was chosen because it was high enough for the wanted accuracy, but not so large that it required too much zero-padding. To make the values for the different signals' frequency domains comparable, each signal was normalized using Equation 4.2 before the FFT was calculated.

$$x''(t) = \frac{x'(t) - \mu}{\sigma} \tag{4.2}$$

where $x'(t)$ is the value of a signal in frame t, μ is the mean of the signal and σ is the standard deviation of it.

4.3.2 HR measuring and weighting

For each 10-second window, the FFT was calculated for all 14 signals. Signals with an FFT containing a significant peak in the frequency range corresponding to a possible heart rate were desirable. This range was defined as 1–2.5 Hz, corresponding to 60–150 bpm. If the highest value of the FFT was outside this range, the measurement was rejected (by receiving a score of zero) and did not contribute to the final result. Figure 4.4 shows how the signals can vary in quality, where Figure 4.4a has a clear peak in the right region, and Figure 4.4b is spread out and has its peak outside the heart rate range.



(a) FFT of the third component of PCA on RGB. A clear peak at 84 bpm can be seen.

(b) FFT of the third component of ICA on HSV. No clear peak can be seen.

Figure 4.4: Two different signals' FFTs for the same video sequence window (first window of T1), where (a) is a much more relevant result than (b). The shaded areas mark the defined heart rate range, and the crosses denote the peaks.

When evaluating a result where the peak is in the correct range, two factors are especially important for the significance of it: how big the peak is and its size compared to other peaks. If a peak with a high value is found, one cannot be confident of the result if an almost as high peak is found 20 bpm away. The score, s, of a heart rate measurement was because of this calculated as in Equation 4.3.

$$s = m_1 \cdot \frac{m_1}{m_2} = \frac{m_1^2}{m_2} \tag{4.3}$$

where m_1 is the highest peak and m_2 the second highest peak at least 8 bpm (two bins) away from m_1. This threshold was used because measurements within two bins (0–8 bpm) can be considered as almost identical and should not harm the significance of the result.

4.3.3 Average HR calculation

When all heart rates and their corresponding scores had been calculated for all 10-second windows, the final estimate of the average heart rate, $\bar{p}$, could be calculated. Three different methods were used for this, where one utilized all measurements in each time window, and two of

them only used the measure with the highest score in each window. For the last two, one of them used the weighted average and the other one the non-weighted.

Equation 4.4 describes the first method, where W is the number of windows, and $p(w, f)$ and $s(w, f)$ are the heart rate and its corresponding score for a specific signal, f, in a window, w. In short, the equation took the weighted average of all resulted heart rates. This method was called *AllWeighted*.

$$\bar{p} = \frac{1}{\sum_{w=1}^{W} \sum_{f} s(w, f)} \sum_{w=1}^{W} \sum_{f} p(w, f) \cdot s(w, f) \tag{4.4}$$

Equation 4.5 describes the second method, which calculated the weighted average of the heart rates with the best score in each window. The notation is the same as in the equation above, and this method was called *BestWeighted*.

$$\bar{p} = \frac{1}{\sum_{w=1}^{W} \max_{f} s(w, f)} \sum_{w=1}^{W} p(w, \arg\max_{f} s(w, f)) \cdot \max_{f} s(w, f) \tag{4.5}$$

Equation 4.6 describes the last method, *BestNonweighted*, which calculated the non-weighted average of the heart rates with the highest score in each window. The notation is the same as before.

$$\bar{p} = \frac{1}{W} \sum_{w=1}^{W} p(w, \arg\max_{f} s(w, f)) \tag{4.6}$$

4.4 Experimental setup

For the evaluation of the proposed method, a dataset of videos recorded outdoors was needed. Since there are few outdoor environment corpora available, where none could be found publicly, we collected our own.

The videos were filmed with a Blackmagic Design Studio Camera 4K using a Panasonic H-FS014042 objective. They were shot in 1920x1080 pixels resolution at 30 FPS in a raw, uncompressed format with 8-bit color depth. To lower the file size and the computational

time, the video files were compressed with H.264 to MP4 file format using HandBrake [14]. It has been shown that compressing a video this way does not decrease the resolution required for applying the methods employed in this work, but rather improves it [29]. In each video, the ground truth heart rate was measured using a Garmin Forerunner 210 with its "premium soft strap heart rate monitor" [11], where the face of the watch was recorded in the videos. The video files were inspected manually, and the average heart rates were calculated.

The video dataset consisted of a total of fifteen 20–35 seconds long videos split equally among three different subjects. The subjects are described in Table 4.1 and the videos in Table 4.2. All videos were recorded outdoors because it is the intended environment for the final system. Since the performance was to be measured during different natural moving situations, the data was collected in three conditions depending on the type of head movement. These types were:

- **Still –** Only very small, natural movements.
- **Panning –** The head changes position but is always facing the camera.
- **Rotation –** The head is rotating and is not always facing the camera.

Table 4.1: A description of the subjects.

#	Gender	Age	Skin tone
S1	Male	23	Light
S2	Male	24	Light
S3	Male	29	Light

Table 4.2: A description of the videos in the dataset. T is the length of the video, HR is the average heart rate with the lowest and the highest measurements in parenthesis and D is the distance from the subject to the camera. The light value "half sun" means that the sun came from the side and only half of the face was exposed.

#	Subj.	T (s)	HR (bpm)	D (m)	Light	Head movements
T1	S1	31	82 (79–85)	0.5	Sun	Still
T2	S1	34	86 (81–93)	3.0	Sun	Still
T3	S1	31	83 (80–85)	6.0	Sun	Still
T4	S1	35	79 (76–85)	3.0	Sun	Panning, rotation
T5	S1	32	100 (97–104)	2.5	Shade	Panning, rotation
T6	S2	31	78 (74–83)	0.5	Half sun	Still
T7	S2	31	80 (77–83)	2.0	Half sun	Still
T8	S2	35	79 (76–83)	4.0	Sun	Panning
T9	S2	35	79 (76–84)	2.5	Shade	Still
T10	S2	31	90 (77–107)	2.5	Half sun	Panning
T11	S3	32	79 (75–85)	1.0	Sun	Still
T12	S3	31	96 (87–105)	2.5	Sun	Still
T13	S3	31	77 (74–81)	3.0	Sun	Panning, rotation
T14	S3	21	82 (81–85)	4.0	Shade	Panning
T15	S3	20	89 (88–90)	3.0	Shade	Panning, rotation

4.5 Performance evaluation

The three different methods (AllWeighted, BestWeighted and BestNonweighted) presented in Section 4.3.3 were used to estimate the average heart rates for the videos in the dataset, and the **(absolute) errors** were calculated. Because of the way the ground truth heart rate was measured, the instantaneous heart rates in the different windows were not considered, which implied that measurements like root-mean-square error (RMSE) were not applicable.

To get an indication of how clear the PPG signals were, the **weighted standard deviations**, as well as the **regular standard deviations**, of the heart rate measurements were calculated on all video sequences. The former was calculated both for all results and for the best-weighted results in each time window, and the latter was calculated exclusively for the best-weighted results. The weighted standard deviation was calculated using Equation 4.7.

$$\sigma_w = \sqrt{\frac{1}{\frac{M-1}{M}\sum_{i=1}^{N} s(i)} \sum_{i=1}^{N} s(i)\left(p(i) - \bar{p}\right)^2} \tag{4.7}$$

where N is the number of heart rate estimates, M the number of non-zero weights (scores), $p(i)$ is a heart rate estimate and $s(i)$ its corresponding score, and $\bar{p}$ is the weighted average of the heart rates.

Finally, the **number of frames where the face detector failed to find a face** was counted. This was done because it is important to know for an understanding of the system's performance.

Chapter 5

Results

The results are presented in three sections. Firstly, the general performance of the different averaging methods is showed and analyzed. Secondly, the test cases are divided into categories based on what type of head movement they contain to investigate how this affected the performance. Finally, it is inspected in which videos and in how many frames the face tracker did not find a face and how much this had an impact on the performance in these test cases.

5.1 Averaging methods

We start by analyzing the absolute errors and comparing each heart rate prediction against its ground truth. Figure 5.1 shows how the proposed method performed on average, with systems using only one signal as a comparison. It is clear that the proposed method outperformed the single signal systems, but it is hard to see any difference between the three averaging methods. BestNonweighted performed a bit better than the other methods with an average absolute error of 6.7 bpm, and the hue channel was the signal that performed best individually with an error of 16.1 bpm on average. An average absolute error of 6.7 bpm is marginally lower than the RMSE of the system by Li et al. [22] (7.62), which can be seen as the work with the most similar dataset even though it was recorded indoors. However, it should be said that RMSEs and average absolute errors are not the same measurements and thus this is not a perfect comparison.

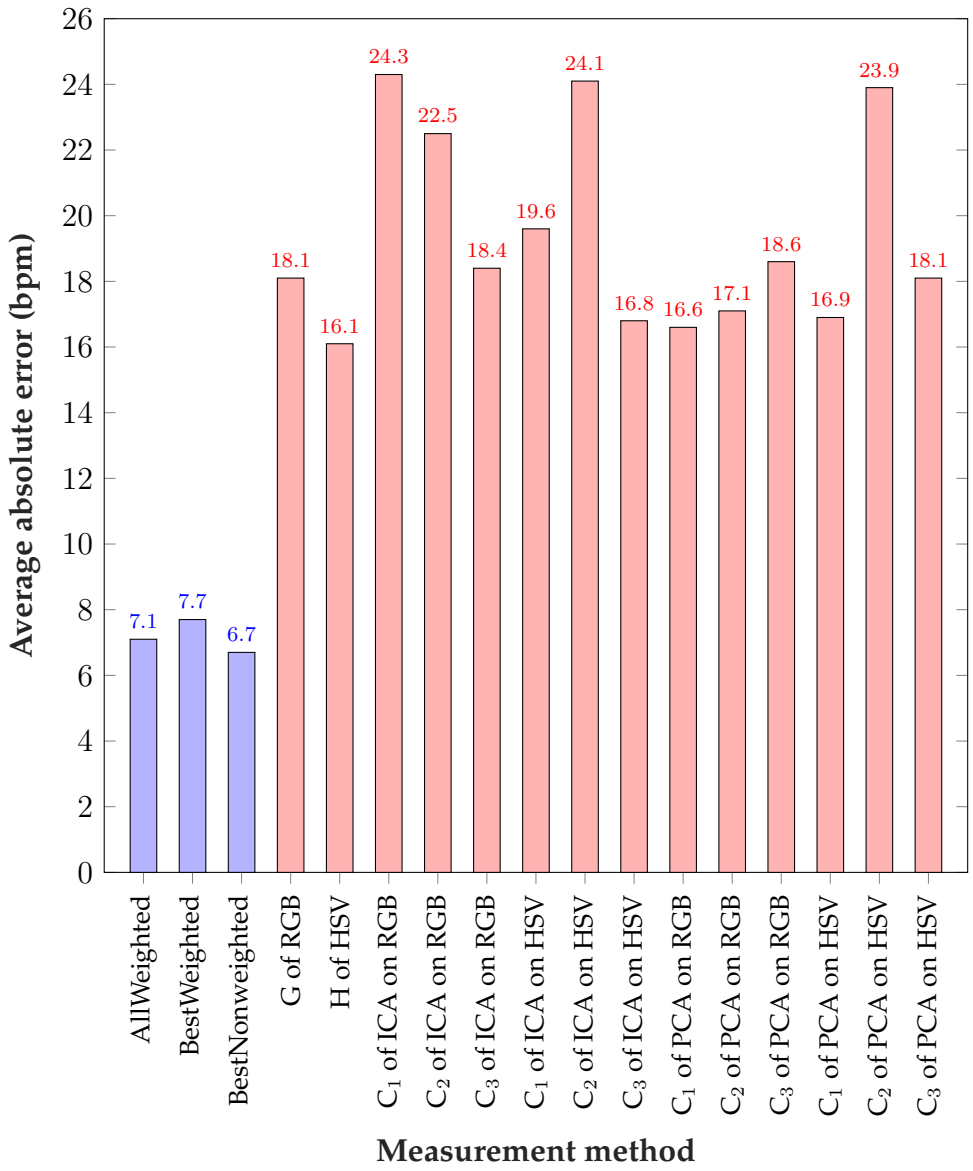


Figure 5.1: The average absolute errors for the proposed methods (blue bars) compared to when each signal is used individually (red bars).

In Figure 5.2, one can see how the three methods guessed compared to the ground truth (the diagonal line). Once again, it is difficult to see a difference between the different averaging techniques. However, what is interesting is that the developed system tends to undershoot. The measured heart rates are in almost every case (35 out of 45 estimates, i.e. 78 percent) lower than the ground truth.

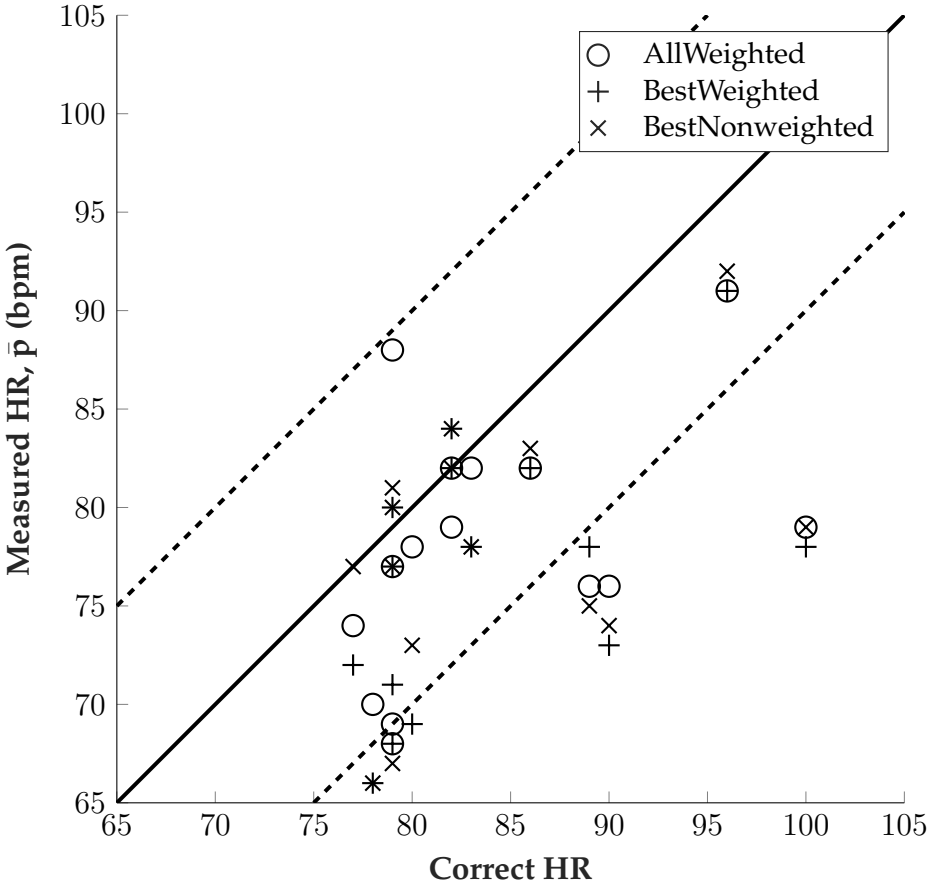
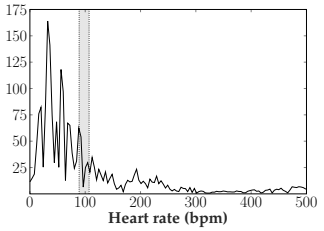
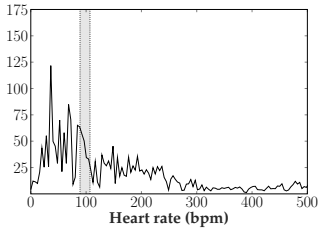


Figure 5.2: Heart rate estimates for the different methods compared to the ground truth. A measurement on the solid line is exactly correct, and a measurement within the two dashed lines is at most 10 bpm off.

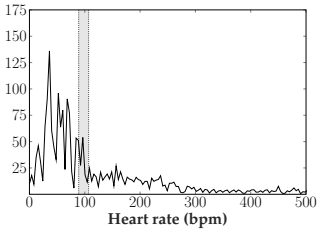
To understand why the system undershoots, test case T10 can be studied. In Figure 5.3, the FFTs for all signals in the first 10-second window is shown. It can be seen that no signal has its peak in the right region of the FFT, but in fact at a much lower frequency. This explains why the heart rate estimate is so low, at least for this test case, and in the next chapter, possible reasons for why it looks like this are discussed.



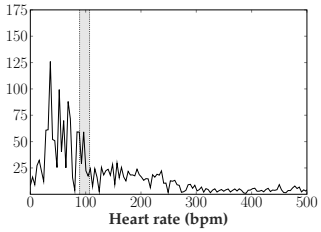
(a) G of RGB.



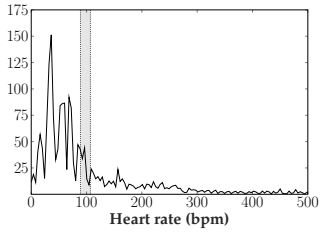
(b) Hue of HSV.



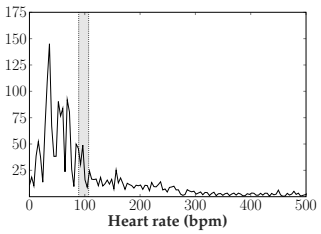
(c) C_1 of ICA on RGB.



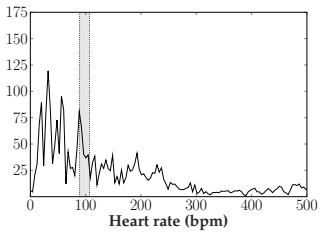
(d) C_2 of ICA on RGB.



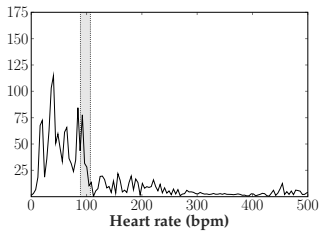
(e) C_3 of ICA on RGB.



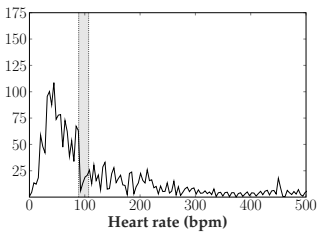
(f) C_1 of ICA on HSV.



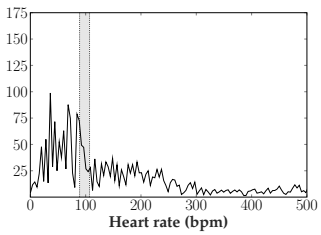
(g) C_2 of ICA on HSV.



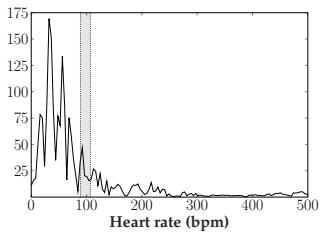
(h) C_3 of ICA on HSV.



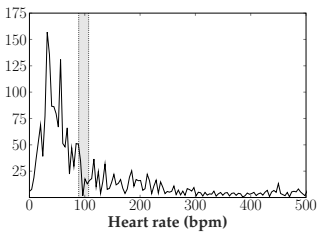
(i) C_1 of PCA on RGB.



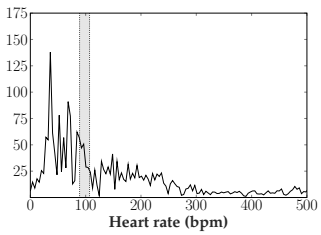
(j) C_2 of PCA on RGB.



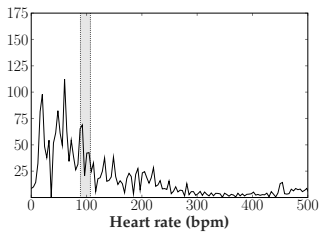
(k) C_3 of PCA on RGB.



(l) C_1 of PCA on HSV.



(m) C_2 of PCA on HSV.



(n) C_3 of PCA on HSV.

Figure 5.3: All signals' FFTs for the first window of T10. The true heart rate was within the shaded area (89–107 bpm).

The absolute errors for each test case can be seen in Figure 5.4. What is notable is that the three test cases with the lowest performance—T5, T10, and T15—were the videos with the most movement. By also studying the standard deviations, one can see that the two test cases with the lowest standard deviation for the "best" measurements (red and brown bars), T1 and T14, also got good heart rate estimates. On the other hand, the estimates for T6, T8, T10 and T15, where the standard deviations were small as well, were poor. Thus, a limited measurement distribution might increase the likelihood of a good result but is no guarantee for it.

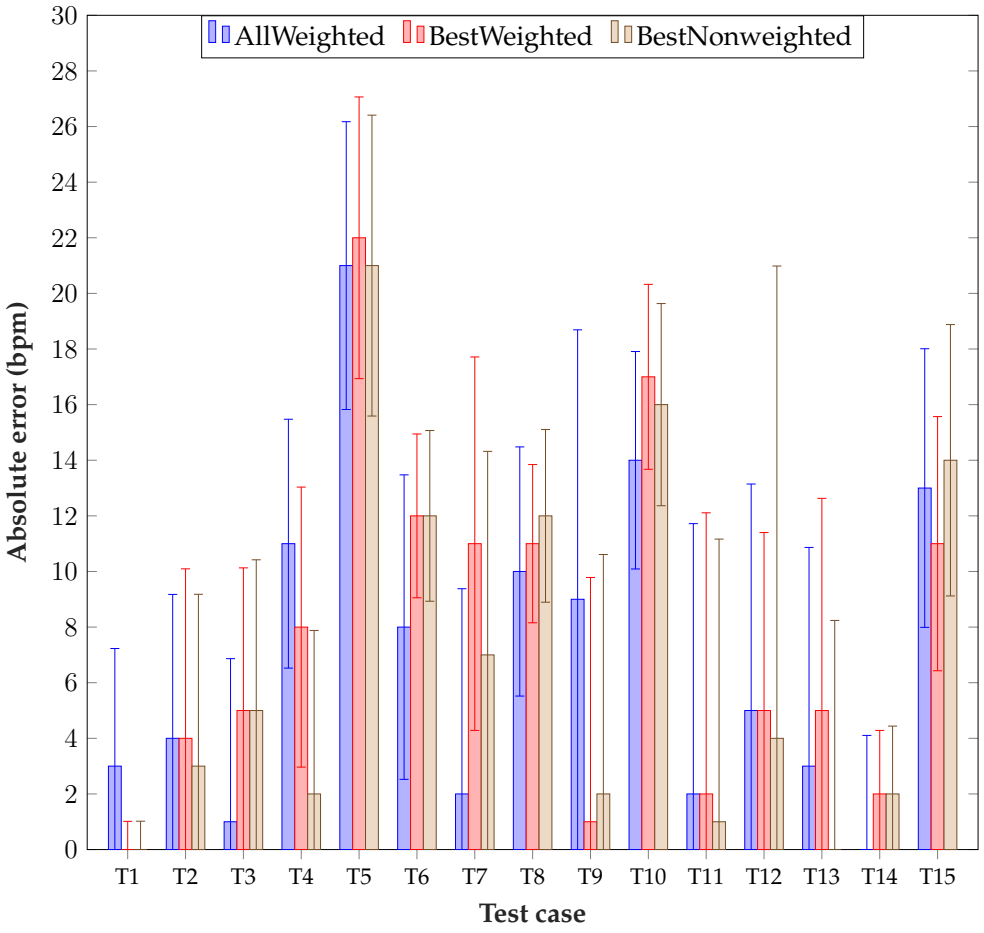


Figure 5.4: The absolute errors for the different averaging methods on each test case. The standard deviation, or the weighted standard deviation, of the measurements, is denoted for each estimate. For AllWeighted and BestWeighted, $\frac{\sigma_w}{2}$ is displayed over and under the average, and for BestNonweighted, $\frac{\sigma}{2}$ is shown the same way instead.

5.2 Head movements in video

In this section, the test cases are split up into the three different head movement categories (still, panning and rotation). This is done to investigate whether the prediction error increases when the head moves compared to when it is still, and whether the type of movement (panning versus rotation) has any impact. If the head in a video sequence both panned and rotated, like for all test cases containing rotation, it was placed in the rotation category.

In Figure 5.5, all heart rate measurements are displayed once again but this time divided by video type. It is clear that the videos with a still head had the best estimates, but no clear difference can be seen between videos with head panning versus videos with head rotation. What is also noteworthy is that the undershooting tendency is unmistakably smaller for still videos compared to videos with movement.

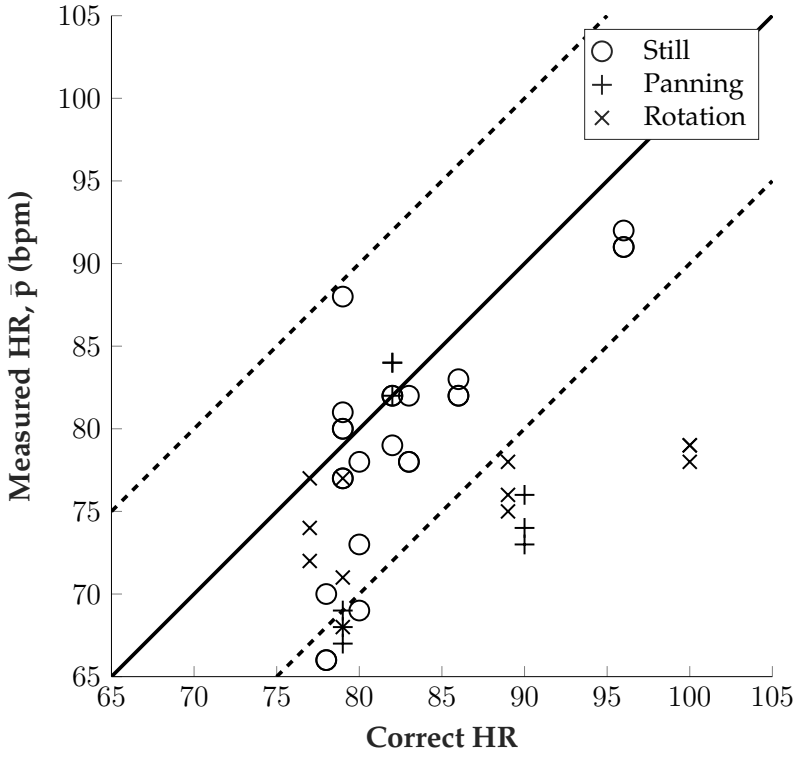


Figure 5.5: HR estimates grouped by what type of head movements the video sequences had.

The average absolute errors for the different types of videos are visualized in Figure 5.6. One can see that the videos with movement are remarkably more difficult for heart rate estimation and that the system only measured the pulse erroneously with 4.5 bpm on still videos. This is in line with the work by Poh et al. [26], where an RMSE of 4.63 was achieved on subjects making very small movements indoors. The videos where the head was also rotating were the most difficult, which was expected since the face was not always facing the camera.

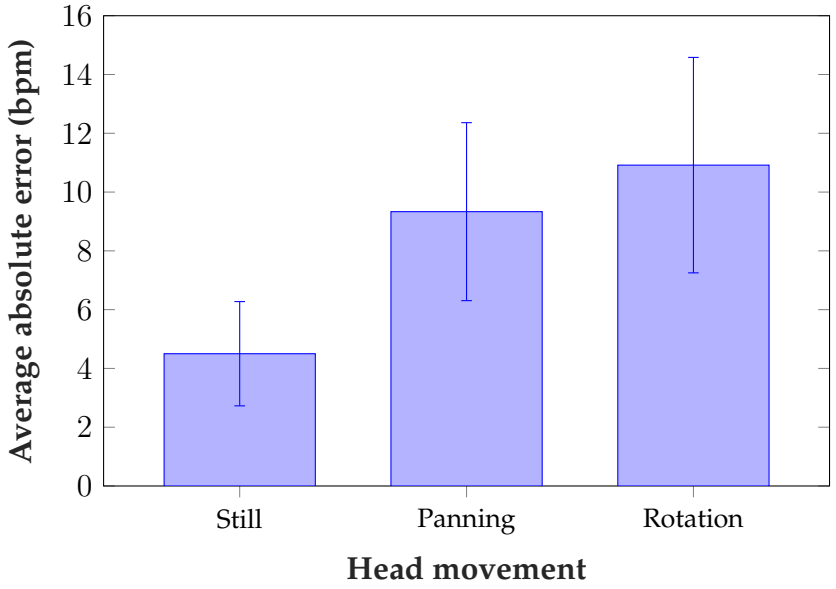


Figure 5.6: The average absolute error for test cases with the same type of head movement. The error bars denote the standard deviation.

5.3 Face tracking errors

To further investigate the main source of the average absolute error, we look at how the face tracker performed and how it affected the system performance if the face was lost. In Table 5.1, it can be seen that the face detector only failed to detect a face in three of the videos, and in one of them (T8) this only occurred in a single frame. What is worth noting is that T5 and T15 both are videos containing rotation of the head, while T8 only contained panning. By looking back at Figure 5.4, one can see that all of the test cases with failed face detection also had poor heart rate estimates. An important remark to do, though, is that

it is possible that the face detector in some frames found something that was not a face (a false positive). Such a case is here classified as a successful detection but might, in fact, be even more problematic than if no face was found.

Table 5.1: The number of frames in each test case for which the face detector failed to find a face.

Test case	**T1**	**T2**	**T3**	**T4**	**T5**
# of frames w/ failed face detection	0	0	0	0	**13**
Test case	**T6**	**T7**	**T8**	**T9**	**T10**
# of frames w/ failed face detection	0	0	**1**	0	0
Test case	**T11**	**T12**	**T13**	**T14**	**T15**
# of frames w/ failed face detection	0	0	0	0	**97**

The significance of a good face detector is made even more indisputable in Figure 5.7, where it can be seen that the error is almost three times as big for videos with failures compared to videos without them. The difference would be even bigger if T8, where only a single frame was lost, was not included in the "Yes" bar.

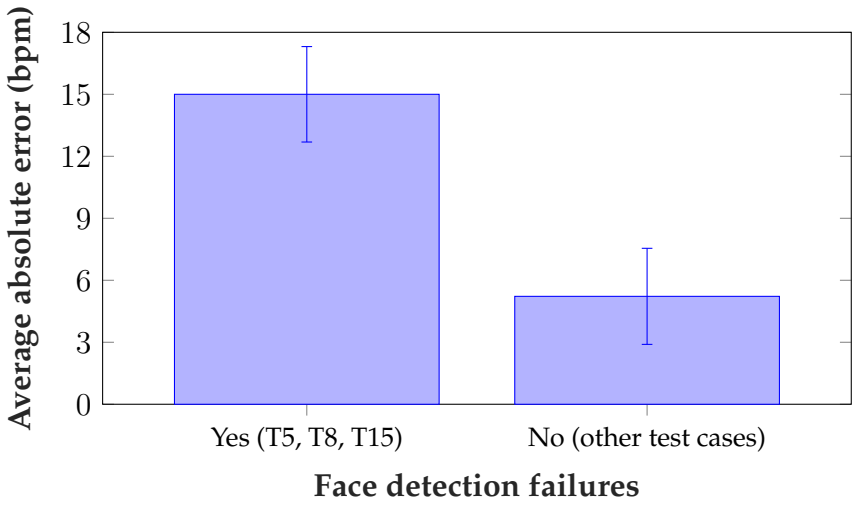


Figure 5.7: The average absolute error and the standard deviation for test cases with face detection failures versus videos without face detection failures.

Chapter 6

Discussion and conclusions

In this chapter, the work done in this thesis is discussed, and conclusions are drawn. At the end of the chapter, possible work to be done in the future is suggested.

6.1 Discussion

The discussion is divided into three parts. First, the main findings from the experiment are discussed, and after that, the experiment itself is criticized. Lastly, the ethical aspects of a heart rate estimation system like this are examined.

6.1.1 Main findings

To start with, the proposed approach enabled the prediction of remote heart rate with low error in outdoor environments. As expected, the way of combining multiple signals by weighting them together based on their significance led to improved results (Figure 5.1). Primarily, the system performed well on still subjects, though. For moving subjects, a clear tendency of undershooting was present (Figure 5.5) and the errors were bigger (Figure 5.6). The overall performance was in line with the system implemented by Li et al. [22] which was tested under similar conditions but indoors. Because of the variations in lighting, outdoor environments can be considered more challenging which makes the proposed approach quite promising compared to the state-of-the-art.

A possible explanation for the undershooting is that artifacts caused by motion have a low frequency and have a significant impact on the result. Since the undershooting is not as apparent on still subjects, this theory is credible. One could believe that the spatial averaging of the pixels in the ROI should make these artifacts negligible, but since the color changes caused by the blood flow are so small, it does not. An alternative explanation for the undershooting could be that the ground truth heart rates were, in fact, lower because of a systematic error in the pulse watch.

Dealing with motion artifacts is not trivial. Detrending reduces them but is obviously not enough to remove them entirely. These motion artifacts could be a result of a non-perfect face tracking or an improper ROI selector. If the tracker followed the face perfectly and smoothly, the ROI pixels would be almost the same in each frame, even if the subject moved. It can be seen in some of the videos that the face tracking in this system could be much better. For example, in T5, T8, and T15, the face is in parts of the video completely lost by tracker (Table 5.1), so that very few forehead pixels end up in the ROI. This affects all windows containing these frames and makes those windows worse for heart rate estimation. For these cases where there are no good data to analyze, it does not matter what averaging method that is used.

To overcome this problem, a better face detection algorithm and face tracker could be used. Both Viola–Jones and KLT feature tracking are quite old algorithms and today there are better and more robust methods for face detection. There are, for example, deep learning methods that have solved this problem with exceptional results [9, 21, 38]. With a better face tracker, the ROI also could be selected more precisely using the feature data, so that the system knows that it has the same part of the face in every frame. These advances can make the real-world applicability of our system possible in the near future. An alternative, or an addition, to a better face tracker, would be to use some form of outlier detection to distinguish between skin and non-skin pixels.

However, if there are only a few frames with weak PPG data, it should not be a significant problem with a well-functioning scoring method. The measurements from the affected windows will then get a low score and not contribute to the final estimate a lot. If the tracker is out of line in a substantial part of the video, so that the majority of the windows are affected, the problem is extremely critical, though.

6.1.2 Criticism of experimental setup

Many modifications could have been made to the conducted experiment that would have improved it. The dataset could have been bigger, to make the results more credible, and it could have been more diversified. All three subjects were white men in age 23–29, and 10 out of 15 test cases (67 percent) had an average heart rate in the 77–83 bpm interval. The heart rates were in this region mainly because these were the natural heart rates for the subjects during the photo shoot, but also because 80 bpm could be seen as a normal heart rate for a player in a slow-paced sport.

Another improvement that could have been made to the experiment is how the ground truth was measured. The accuracy of pulse measurements from a consumer heart rate monitor is arguably not as high as from medical equipment, for example. If a better device had been used, it also means that the heart rate could have been compared in each time window beside the average for the entire sequence.

To better be able to compare the proposed method to other methods, a public dataset of videos could preferably have been used. An alternative for this is the MAHNOB-HCI database [30], used by Li et al. [22]. Whereas this dataset contains rather challenging videos for heart rate estimation with high-quality ground truths, all videos were recorded indoors which make them unsuitable for our purposes.

6.1.3 Ethical aspects

When dealing with camera footage of people, integrity questions are raised. This is an even more critical discussion when physical parameters such as heart rate are extracted from the videos. For a limited study like this, where all subjects have given their consent, this is no problem, but if the system would be used in sports broadcasts in the future, it could be. Whether this is illegal or not to do without explicit permissions from the players is not discussed here. However, this possible legal problem and the ethical complications of a system like this could easily be solved if the players had to agree to this exposure as a part of the tournament terms.

6.2 Conclusions

The main question to be answered in this thesis was whether the performance of remote heart rate measuring on moving subjects could be improved by relying on multiple signals gathered from image data. The short answer to that is yes; according to the results on the small dataset collected in outdoor environments to evaluate this work, it seems like that is the case. The method performed significantly better than when all signals were used individually, regardless of what averaging technique that was used. One can argue that the overall performance of the proposed system would be improved with better signals.

Another conclusion that can be drawn is that the amount of head movement in the video sequence is crucial for the accuracy of a heart rate estimate. It was shown that the general performance on still subjects was acceptable but decreased when the head was panning or rotating in the camera frame. This was partly because it was more difficult to track a moving face, but more importantly because of the introduction of motion artifacts.

6.3 Future work

If this project was to be extended in the future, several things could be improved.

To begin with, the face tracking was far from perfect, especially for faces not looking straight towards the camera. This is a vital part of the pipeline that must be improved to get good data to analyze. Related to this, potential improvements could also be made to the ROI selection. The current method of selecting the forehead does not work well for faces viewed from the side, and it sometimes includes segments of the image that do not contain any skin. One possible solution for this would be to use a skin detector.

The scoring method used to evaluate a heart rate measurement on a signal appeared to work well, but was not compared to other methods for doing this. This could be a topic worth paying attention to in the future as well. More importantly, a better method to filter out motion noise must be found though.

If this system were to be used in any real application in the future, its performance would also need to be improved. For a video of 31 seconds (test case T1), the complete process of estimating the average heart rate takes 160.7 seconds. Out of this, 159.6 seconds are spent on tracking the face and extracting the ROI, which is a factor 5.1 of the video length. This would supposedly be much faster if it were implemented in another programming language than MATLAB, and even faster if it used the GPU. It has been shown that SIFT, which is an algorithm in computer vision to detect and describe local features in images, can be up to ten times faster while using the GPU compared to the CPU [16]. Consequently, there are good premises for making this a real-time system.

Bibliography

[1] Frédéric Bousefsaf, Choubeila Maaoui, and Alain Pruski. "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate". In: *Biomedical Signal Processing and Control* 8.6 (2013), pp. 568–574.

[2] Gary Bradski and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[3] Jean-François Cardoso. "High-order contrasts for independent component analysis". In: *Neural computation* 11.1 (1999), pp. 157–192.

[4] Peter H Charlton et al. "An assessment of algorithms to estimate respiratory rate from the electrocardiogram and photoplethysmogram". In: *Physiological measurement* 37.4 (2016), p. 610.

[5] Eirini Christinaki et al. "Comparison of blind source separation algorithms for optical heart rate monitoring". In: *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*. IEEE. 2014, pp. 339–342.

[6] Thomas Deselaers and Vittorio Ferrari. "Global and efficient self-similarity for object classification and detection". In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, pp. 1633–1640.

[7] *Detect objects using the Viola–Jones algorithm*. mathworks.com. Accessed: 2017-04-07. URL: `https://se.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-class.html`.

[8] Sibylle Fallet et al. "Imaging Photoplethysmography: What are the Best Locations on the Face to Estimate Heart Rate?" In: *Computing in Cardiology 2016*. EPFL-CONF-222366. 2016.

[9] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. "Multi-view face detection using deep convolutional neural networks". In: *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM. 2015, pp. 643–650.

[10] Marc Garbey et al. "Contact-free measurement of cardiac pulse based on the analysis of thermal imagery". In: *IEEE Transactions on Biomedical Engineering* 54.8 (2007), pp. 1418–1426.

[11] *Garmin Heart Rate Monitor*. garmin.com. Accessed: 2017-04-26. URL: https://buy.garmin.com/en-US/US/p/pn/010-10997-00.

[12] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN: 013168728X.

[13] Gerard de Haan and Vincent Jeanne. "Robust pulse rate from chrominance-based rPPG". In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), pp. 2878–2886.

[14] *HandBrake: Open Source Video Transcoder*. handbrake.fr. Accessed: 2017-04-05. URL: https://handbrake.fr/.

[15] Mohamed Abul Hassan et al. "Optimal source selection for image photoplethysmography". In: *Instrumentation and Measurement Technology Conference Proceedings (I2MTC), 2016 IEEE International*. IEEE. 2016, pp. 1–5.

[16] S. Heymann et al. "SIFT implementation and optimization for general-purpose gpu". In: *WSCG '07*. 2007.

[17] Byung S Kim and Sun K Yoo. "Motion artifact reduction in photoplethysmography using independent component analysis". In: *IEEE transactions on biomedical engineering* 53.3 (2006), pp. 566–568.

[18] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. "DistancePPG: Robust non-contact vital signs monitoring using a camera". In: *Biomedical optics express* 6.5 (2015), pp. 1565–1588.

[19] Georg Lempe et al. "ROI selection for remote photoplethysmography". In: *Bildverarbeitung für die Medizin 2013*. Springer, 2013, pp. 99–103.

[20] Magdalena Lewandowska et al. "Measuring pulse rate with a webcam – a non-contact method for evaluating cardiac activity". In: *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*. IEEE. 2011, pp. 405–410.

[21] Haoxiang Li et al. "A convolutional neural network cascade for face detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 5325–5334.

[22] Xiaobai Li et al. "Remote heart rate measurement from face videos under realistic situations". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 4264–4271.

[23] Bruce D. Lucas and Takeo Kanade. "An Iterative Image Registration Technique with an Application to Stereo Vision". In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'81. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679. URL: http://dl.acm.org/citation.cfm?id=1623264.1623280.

[24] Andreia V Moço, Sander Stuijk, and Gerard de Haan. "Motion robust PPG-imaging through color channel mapping". In: *Biomedical optics express* 7.5 (2016), pp. 1737–1754.

[25] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. "Advancements in noncontact, multiparameter physiological measurements using a webcam". In: *IEEE transactions on biomedical engineering* 58.1 (2011), pp. 7–11.

[26] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. "Noncontact, automated cardiac pulse measurements using video imaging and blind source separation." In: *Optics express* 18.10 (2010), pp. 10762–10774.

[27] Qiang Qiu et al. "Low-cost Gaze and Pulse Analysis Using RealSense". In: *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*. MOBIHEALTH'15. London, Great Britain: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015, pp. 276–279. ISBN: 978-1-63190-088-4. DOI: 10.4108/eai.14-10-2015.2261657. URL: http://dx.doi.org/10.4108/eai.14-10-2015.2261657.

[28] Nikolai Grov Roald. "Estimation of vital signs from ambient-light non-contact photoplethysmography". MA thesis. Norway: Norwegian University of Science and Technology, 2013.

[29] Philipp V Rouast et al. "Remote heart rate measurement using low-cost RGB face video: A technical literature review". In: *Front. Comput. Sci* (2016).

[30] Mohammad Soleymani et al. "A multimodal database for affect recognition and implicit tagging". In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 42–55.

[31] Carlo Tomasi and Takeo Kanade. *Detection and Tracking of Point Features*. Tech. rep. International Journal of Computer Vision, 1991.

[32] *Track points in video using Kanade-Lucas-Tomasi (KLT) algorithm*. mathworks.com. Accessed: 2017-04-09. URL: https://se.mathworks.com/help/vision/ref/vision.pointtracker-class.html.

[33] Gill R Tsouri and Zheng Li. "On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras". In: *Journal of biomedical optics* 20.4 (2015), pp. 048002–048002.

[34] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. "Remote plethysmographic imaging using ambient light." In: *Optics express* 16.26 (2008), pp. 21434–21445.

[35] Paul Viola and Michael J Jones. "Robust real-time face detection". In: *International journal of computer vision* 57.2 (2004), pp. 137–154.

[36] Wenjin Wang et al. "Algorithmic principles of remote-PPG". In: *IEEE Transactions on Biomedical Engineering* (2016).

[37] Hao-Yu Wu et al. "Eulerian Video Magnification for Revealing Subtle Changes in the World". In: *ACM Trans. Graph.* 31.4 (July 2012), 65:1–65:8. ISSN: 0730-0301. DOI: 10.1145/2185520.2185561. URL: http://doi.acm.org/10.1145/2185520.2185561.

[38] Zhanpeng Zhang et al. "Facial landmark detection by deep multi-task learning". In: *European Conference on Computer Vision*. Springer. 2014, pp. 94–108.