B.Sc. Thesis
Bachelor of Science in Engineering

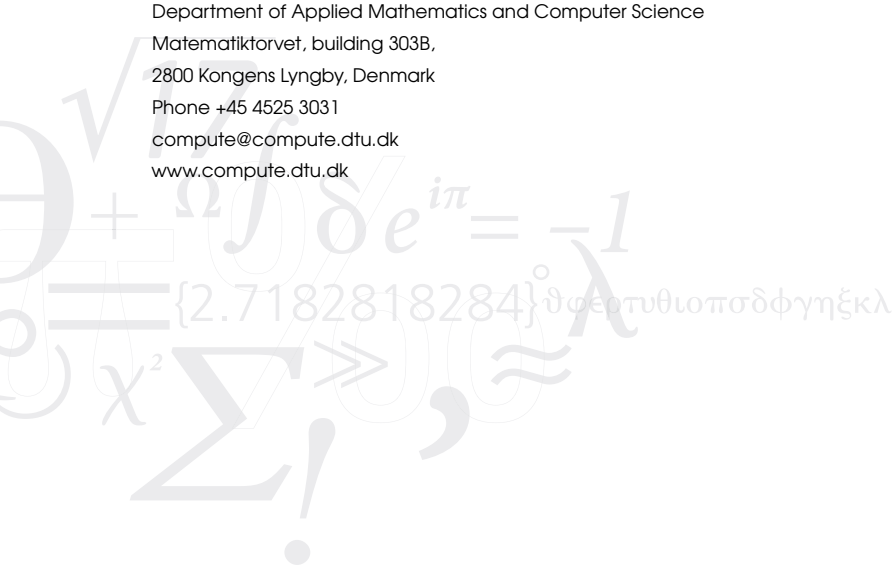
DTU Compute
Department of Applied Mathematics and Computer Science

Camera-based Heart Rate Monitoring

Janus Nørtoft Jensen (s113225)
Morten Hannemose (s113258)

Kgs. Lyngby 2014

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Matematiktorvet, building 303B,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk



Summary

In recent years, measuring of the heart rate using only a cheap camera has become popular. In this thesis we have implemented and tested different non-contact methods for measuring the heart rate. Face detection was used in all of these and ICA for separation of the observed signals into a heart rate signal was used in some. Videos of test subjects with a duration of 60 seconds were recorded. The methods were applied to these and were compared with a reference heart rate monitor.

The results were comparable or better than similar studies, and we achieved high correlation and small errors compared with the reference heart rate. We also found that the choice of colour space did not significantly change the results.

We have extended this method to calculate maps of how visible the heart rate is in different parts of the face, and used these to obtain better heart rate signals. Using these improved signals signals we have calculated the time delay of the heart rate between the hand and the face. We have compared this time difference with the systolic blood pressure, but we were unable to conclude whether there is a significant correlation, due to a too small sample size.

Resume

I de seneste år er pulsmåling med udelukkende med et billigt kamera blevet populært. I denne opgave har vi implementeret og testet forskellige kontaktfrie metoder til at måle puls. Ansigtsgenkendelse blev brugt i dem alle og i nogle blev ICA brugt til at separere de observerede signaler til et pulssignal. Videoer af forsøgspersoner med en længde på 60 sekunder blev optaget. Metoderne blev anvendt på disse og blev sammenlignet med en referencepulsmåler.

Resultaterne var sammenlignelige eller bedre i forhold til lignende studier og vi opnåede høj korrelation og små fejl i sammenligning med referencepulsen. Vi fandt også ud af at valget af farverum ikke påvirkede resultaterne signifikant.

Vi har udvidet metoden til at udregne billeder over hvor synlig pulsen er i forskellige dele af ansigtet og brugt disse til opnå bedre pulssignaler. Ved brug af disse forbedrede signaler har vi udregnet tidsforskelle af pulssignalet mellem hånden og hovedet. Vi har sammenlignet disse tidsforskelle med det systoliske blodtryk, men vi var ikke i stand til at konkludere om der er en signifikant korrelation, på grund af et for lille datasæt.

Preface

This bachelor's thesis was prepared at the department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) under the supervision of Professor Rasmus Larsen, Professor Knut Conradsen and Ph.D.-student Jakob Wilm in fulfillment of the requirements for acquiring a BSc Eng degree in Mathematics and Technology.

The scope of the project was 15 ECTS points and it was carried out from February 7th to June 16th 2014.

Kgs. Lyngby, June 16, 2014

Janus Nørtoft Jensen (s113225)

Morten Hannemose (s113258)

Acknowledgements

We would like to thank our advisors and the rest of our supervision group for valuable advice, feedback and cake throughout the project period.

We are particularly grateful for the assistance given by Anders Roy Christiansen and Henrik Hannemose with their interesting ideas and elegant algorithmic suggestions for our project. Your feedback and interesting discussions have proven very valuable to us.

We are very grateful for the assistance with proofreading provided by Emilie Lundbye Dalsgaard.

Also we would like thank Rasmus Jessen Aaskov for technical assistance on how to access the HPC clusters, which we used for calculation of heart rate intensity maps (HRIMs).

We would especially like to thank all of the people who volunteered as test subjects and allowing us to record videos of them. This thesis would not have been possible without them.

We wish to thank the gardener in 324 for keeping the trees in good shape, by rigorous weekly shakings. Without him there would be leaves everywhere which would have been a really problematic working environment.

We also want to thank the bar *Hegnet* for being a place where great ideas could form, and providing a steady flow of liquids for optimal programming conditions.

Our special thanks are extended to you, the reader, for actually taking the time to read this far. Great job!

Contents

List of Abbreviations

BPM	Beats per minute
BSS	Blind source separation
DTU	Technical University of Denmark
ECG	Electrocardiogram
FFT	Fast Fourier transform
FPS	Frames per second
HRIM	Heart rate intensity map
HSV	Hue, saturation and value
IBI	Interbeat interval
ICA	Independent component analysis
JADE	Joint approximate diagonalization of eigenmatrices
KLT	Kanade-Lucas-Tomasi
Lab	Lightness, a and b
LBP	Local binary pattern
PPG	Photoplethysmography
PTT	Pulse transit time
PTTD	Difference in pulse transit time
RGB	Red, green and blue
RMSE	Root mean square error
ROI	Region of interest
SNR	Signal-to-noise ratio

CHAPTER 1

Introduction

People are becoming more and more interested in their personal health. This is evident from the rapid rise in personal health related applications for smartphones, which today are capable of complex calculations. Vital signs are good indicators of personal health, and measurement of these are of interest. The most commonly measured vital signs are the heart rate and the blood pressure.

Today's most widespread techniques for both of these signs, such as finger monitors or blood pressure cuffs, require contact with the person's skin in order to get a measurement. Over longer periods of time this can cause discomfort and has an inherit risk of loosing contact, e.g. when a person is sleeping. Furthermore, the usual methods for measuring blood pressure cannot produce continuous readings.

Recent advancements in this field have led to automatic non-contact inexpensive methods for continuous monitoring of the heart rate (Poh et al. 2010) using nothing but ambient light and a consumer level camera. Due to the inexpensiveness and the fact that the method is relatively easy to implement in for example a smartphone app gives this method the potential to become widespread.

To our knowledge there is no such simple method for non-contact measuring of blood pressure. However, methods for continuous monitoring of the blood pressure have been shown, where the relationship between the velocity of the blood and the force it exerts on the blood vessels is utilised (Douniama and Couronné 2007).

1.1 Problem statement

In this thesis we will recreate the method developed by Poh et al. (2010) for measuring heart rate with a low-cost camera and only ambient light, in the hope of achieving similar results. We will examine if the choice of colour space has any influence on the results. Furthermore, we will look at alternatives to ICA for extracting the heart rate.

We will look at the error introduced by shortening the length of the video sequence used for the measurement. This will be used to determine an acceptable length for use in possible applications.

Additionally, we aim to expand on this method in order to create a HRIM of the face to show where the heart rate is most visible. This can possibly be used to detect

abnormalities in the facial blood flow. The new knowledge about the HRIM can be used to improve the detection of heart rate in ordinary videos.

Using the heart rate measurement method we want to examine the relationship between the time difference of when blood from a heartbeat reaches the hand and face and the blood pressure.

CHAPTER 2

Background

2.1 Vital Signs

Vital signs are physiological measurements which assess the most basic body functions. They are good indicators of the general health of an individual. Two vital signs are the heart rate and the blood pressure which we will describe briefly in this section.

Heart Rate

The heart rate is defined as the number of times the heart beats per minute. A healthy human adult usually has a resting heart rate between 60 and 80 beats per minute (BPM). The heart is a muscle and, like all other muscles, becomes stronger and more efficient through exercise. Therefore an individual in excellent physical condition, such as a marathon runner, can have a resting heart rate as low as 35 BPM (Scanlon and Sanders 2007). According to Tanaka et al. (2001) the maximum heart rate is strongly correlated with age. The relationship is approximately given by $\mathrm{HR}_{\mathrm{max}} = 208 - 0.7 \cdot \mathrm{Age}$. For a 20-year old this gives $\mathrm{HR}_{\mathrm{max}} = 194$. For measurements in this thesis we have chosen to limit the possible values for the heart rate from 40 to 200 because of the above.

Blood Pressure

Blood pressure is the force the blood exerts against the walls of the blood vessels. Blood pressure is usually measured in mmHg (millimeters of mercury). A blood pressure reading yields two numbers; a higher number, the systolic blood pressure, describing the maximum pressure, and a lower number, the diastolic blood pressure, which describes the minimum pressure. The normal range for the systolic blood pressure lies between 90 to 120 mmHg and the diastolic blood pressure ranges from 60 to 80 mmHg (Scanlon and Sanders 2007).

2.2 Photoplethysmography

The word photoplethysmography (PPG) is composed of the words ‘photo’ and ‘plethysmography’. The latter comes from the greek word plethysmos which means increase, and is a term describing the measurement of changes in volume in different parts of the body (Verkruysse et al. 2008). These changes are often due to fluctuations in the

amount of blood in a specific body part. Hence by plethysmography is meant the detection of the pulse wave travelling through the body. The word photo refers to the use of light for plethysmography. The term photoplethysmography was first used by Hertzman and Spealman (1937). From a PPG signal it is relatively easy to derive the heart rate from the time between two consecutive peaks. An example of a signal is shown in Figure 2.1.

Two variations of PPG are used. The first is called transmission mode PPG and with this method light is transmitted through tissue (e.g. a finger) and the changes in the transmitted light caused by the pulse wave is measured by a probe on the other side. The other method is called reflectance mode PPG and, as the name suggests, the probe is placed on the same side as the light source and it measures pulsatile changes in the back-scattered light. The latter method is the one investigated in this thesis with ambient light as the light source and a camera as the probe.

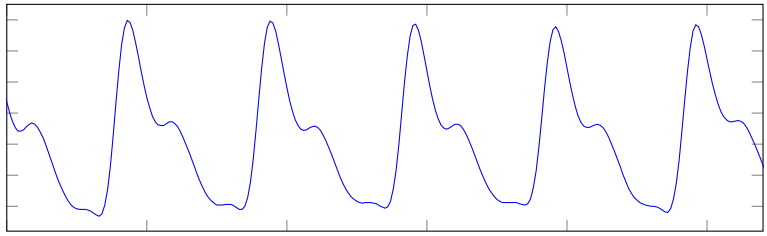


Figure 2.1: Example of a PPG signal.

Hemoglobin absorption spectra

A human has 4 to 6 litres of blood of which 52% to 62% is plasma. The remaining 38% to 48% are different blood cells with the largest part of them being red blood cells (Scanlon and Sanders 2007). The plasma is mostly transparent so the majority of the absorbtion of the blood stems from the red blood cells. The function of the red blood cells is to carry oxygen. To do this each cell contains a vast amount of the protein hemoglobin (Hb) which can bond oxygen molecules. When blood is flowing from the heart to the body, it contains oxygenated hemoglobin (HbO_2) and when it flows back to the heart, it contains deoxygenated hemoglobin (Hb). As PPG measures fluctuations in the amount of blood in the body part by its colour, it is relevant to examine the absorption spectra of Hb and HbO_2. These spectra are shown in Figure 2.2 on the next page in the range 350-700 nm. The figure has been made using data found in Zijlstra et al. 2000. It shows that both Hb and HbO_2 have absorption peaks between 530 nm and 590 nm which corresponds to green and yellow colours.

We note that although Figure 2.2 on the facing page has been made using data found by in vitro experiments, that is by transmission of light trough blood removed from

the body, a similar absorption spectra is found in (Cui et al. 1990) for blood still in the body (in vivo).

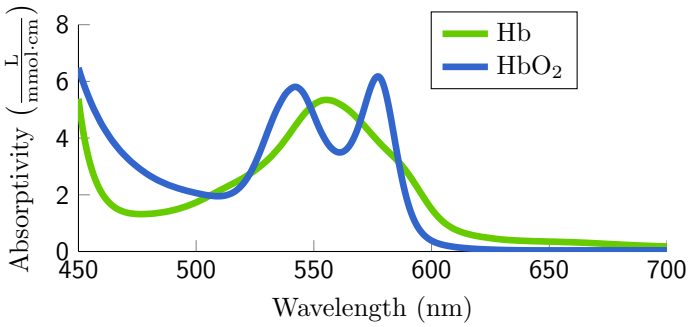


Figure 2.2: Absorption spectra for hemoglobin (green) and oxyhemoglobin (blue) in the range of visible light.

Since PPG measures fluctuations in the amount of blood, it is important that oxygenated and deoxygenated blood have similar absorptions, so that fluctuations are measured similarly for both types of blood. Otherwise the observed colour would also be affected by the concentration of oxygen in the blood. The absorption spectra of Hb and HbO_2 in Figure 2.2 look similar, with high absorptivity in the green colour range and some in the blue and with low absorptivity in the red colours. This implies that oxygenated and deoxygenated blood have similar absorptions.

Since the wavelenghts corresponding to green colours have large absorptivities, changes in blood volume will be more evident at these wavelengths whereas the changes will be much less evident at red wavelengths since the absorptivity is much lower here. Thus in the RGB colour space we expect the strongest PPG signal to be present in the green channel.

2.3 Pulse transit time

PTT is defined as the time it takes a pulse wave to travel between the heart and another place on the body within the same cardiac cycle (a complete heartbeat from its beginning to the beginning of the next). There is a strong dependency between the PTT and the pressure that the pulse wave exerts on the blood vessels, which is the blood pressure (Douniama and Couronné 2007). To compute the PTT two measurements are required. One proximal (the heart) and one distal (the body part). The difference in PTT between two distal points (such as a toe and a finger), which we denote by PTTD can be measured directly using PPG at both points (Nitzan et al. 2002).

2.4 Colour Spaces

Throughout the thesis we will be working with different colour spaces. A colour space is a way of describing colours through numbers. These numbers are often called components and a typical colour space has three or four. We will be working with the RGB (red, green and blue), the HSV (hue, saturation and value) and the Lab (lightness, a and b) colour spaces.

RGB (Red, green and blue)

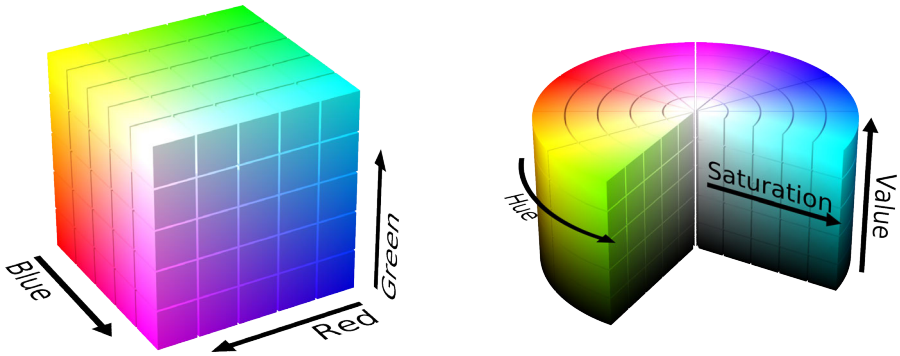
The RGB colour space is the most commonly used colour space. In this space colours are represented as a mixture of red, green and blue, hence the name. Each of these components can have a value between 0 and 1. Red is for example represented by the triplet $\{1, 0, 0\}$ and white is represented by $\{1, 1, 1\}$. An RGB colour can be viewed as a point in $[0, 1]^3 \subset \mathbb{R}^3$ with the components as cartesian coordinates. This is visualised in Figure 2.3(a) on the facing page. The choice of red, green and blue as the primary colours is connected to the physiology of the human eye, whose three most common photoreceptor cells responds most to wavelengths corresponding respectively to red, green and blue colours (Bowmaker and Dartnall 1980).

HSV (Hue, saturation and value)

Definition

The HSV colour space is a way of representing colours using cylindrical coordinates as opposed to the cartesian coordinates of the RGB colour space, which is seen in Figure 2.3(b) on the next page. Hue, the angular dimension, describes the colour with red being at 0 radians, green at $\frac{2\pi}{3}$ radians, and blue at $\frac{4\pi}{3}$ radians. Saturation describes, as the name suggests, how saturated a colour is, or in another word how colourful it is. Geometrically saturation corresponds to the distance from the center axis and ranges from 0 to 1, with 0 giving grey colours and 1 giving the primary colours and linear mixtures of adjacent pairs of these. Value, which geometrically is the height, and lies between 0 and 1, is the darkness of a colour relative to the saturation with 1 resulting in no darkening and 0 completely black. HSV can be transformed from RGB by the non-linear transformation (Bradski 2000).

$$
\begin{aligned}
V &= \max\{R, G, B\}, \\
S &= \begin{cases} \frac{V - \min\{R,G,B\}}{V} & \text{if } V \neq 0 \\ 0 & \text{otherwise,} \end{cases} \\
H &= \begin{cases} \frac{60(G-B)}{V - \min\{R,G,B\})} & \text{if } V = R \\ 120 \frac{60(B-R)}{V - \min\{R,G,B\}} & \text{if } V = G \\ 240 \frac{60(R-G)}{V - \min\{R,G,B\}} & \text{if } V = B. \end{cases}
\end{aligned}
$$



(a) RGB (b) HSV

Figure 2.3: Visualization of the RGB and HSV colour spaces.

Motivation

A possible advantage of using the HSV over the RGB colour space is that hue and to some degree saturation is only dependent on the colour of an object and not on the intensity of the light reflected by it, whereas all three components in RGB are dependent on both. Hence hue and to some extent saturation should be uncorrelated with changes in light. Thus the components in HSV will be less correlated with each other than the components in RGB. The transformation from RGB to HSV of the signals is non-linear and is therefore not possible through ICA. The separation of light dependance gained from the transformation is thus not linear either. In certain situations where the heart rate cannot be found using RGB, it is therefore perhaps possible to find it using HSV.

Lab (Lightness, a and b)

Definition

The Lab colour space consists of a lighting component L and two colour components a and b. Ligthness, which is the central axis in Figure 2.4 on the following page, ranges from black at 0 to white at 1. The colour components are based on opponent colours and the idea that a colour cannot be both red and green or both blue and yellow. For the component a positive values indicate amounts of red and negative values indicate amounts of green, and similarly for the b component yellow is positive and blue is negative. A visualization of the colour space can be seen in Figure 2.4 on the next page.

Motivation

The two colour components a and b are uncorrelated with light changes, and the motivation for using Lab colours is thus the same as for HSV, since the transformation from RGB to Lab is non-linear as well.

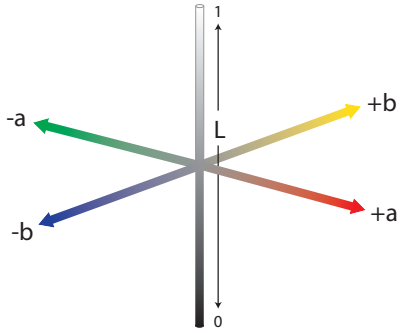


Figure 2.4: Visualization of Lab colour space.

2.5 Independent component analysis

Independent component analysis (ICA) is a special case of the broader concept of blind source separation (BSS). BSS deals with separation of a set of mixed signals into a set of source signals with little to no information about the source signals. A common example is the situation where multiple people speak at the same time in the same room. If one were to record the sound in the room with multiple microphones, the obtained signals would be a mixture of the different people speaking. The aim of BSS in this context is to try and demix the signals into a set of signals each containing the speech signal of a single person. This concept can be seen in Figure 2.5. The ICA

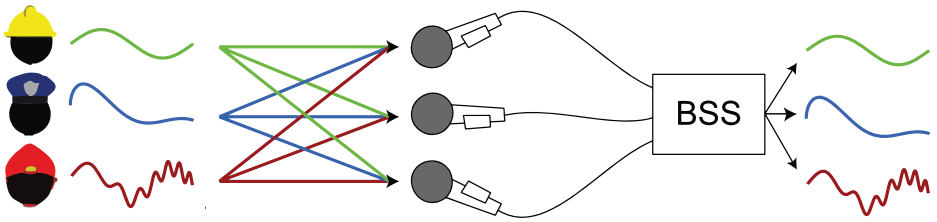


Figure 2.5: Illustration of the concept of BSS.

model assumes that the source signals are non-Gaussian and statistically independent.

It also assumes that the observed signals are linear mixtures of the source signals i.e.

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t),$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^\top$ is a column vector containing the observed signals, $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_m(t)]^\top$ is a column vector containing the source signals and $\boldsymbol{A}$ is a $n \times m$ linear mixing matrix. The number of recovered sources cannot be greater than the number of observed signals hence $m \leq n$. In most applications m is choosen to be equal to n to recover the largest number of source signals, hence $\boldsymbol{A}$ is square $n \times n$ matrix. ICA attempts to approximate the inverse of $\boldsymbol{A}$ as a demixing matrix $\boldsymbol{W}$ for which

$$\widetilde{\boldsymbol{s}}(t) = \boldsymbol{W}\boldsymbol{x}(t),$$

where $\widetilde{\boldsymbol{s}}(t) = [\widetilde{s_1}(t), \widetilde{s_2}(t), \ldots, \widetilde{s_m}(t)]^\top$ is an estimate of of the vector of source signals $\boldsymbol{s}(t)$. The central limit theorem states that a sum of independent random variables is more Gaussian than the variable themselves. Hence to demix the observed signals the most and thus obtaining the source signals $\boldsymbol{W}$ most maximize the non-Gaussianity of each recovered signal. As the sources are assumed to be non-Gaussian, this is in fact possible. Different algorithms uses different means of measuring non-Gaussianity such as kurtosis or negentropy.

JADE (Joint approximate diagonalization of eigenmatrices)

The ICA algorithm used in this thesis is called JADE (Cardoso and Souloumiac 1993). A free *MATLAB*-implementation has been made available by Cardoso (2013), who is also an author of the original algorithm. The input given to JADE is the observed $\boldsymbol{x}$ and the output is the demixing matrix $\boldsymbol{W}$.

2.6 Face tracking

Viola-Jones face detection

A commonly used face detection algorithm is Viola-Jones (Viola and Jones 2004). The original algorithm is based on calculating a number of Haar-like features very fast using the integral image, but we have used the local binary patterns (LBPs) cascade supplied with OpenCV (Bradski 2000) because of faster computation times.

The basic principle is the same as with Haar-like features. Observe Figure 2.6 on the next page. Each point p_i has the value of the sum of all the pixels in the respective rectangle in the 3×3 grid. The LBP feature of p_0, is found by thresholding points in a circle with regard to the center point, p_0. The LBP binary 8-bit descriptor of Figure 2.6 on the following page is thus given by

$$[p_0 > p_1, p_0 > p_2, \ldots, p_0 > p_8]\,.$$

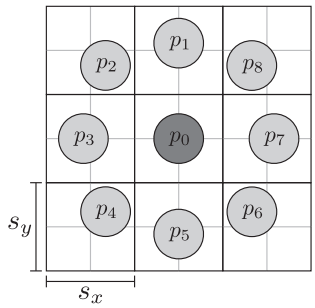


Figure 2.6: Local binary pattern (LBP)

Each rectangle has dimensions (s_x, s_y) where Figure 2.6 has $s_x = s_y = 2$, but in other cases the rectangles may be nonsquare.

For efficient calculation of LBP features for any s_x, s_y the so called integral image $\boldsymbol{II}$ is introduced. For an image $\boldsymbol{X}$, $\boldsymbol{II}$ is defined as

$$ii_{i,j} = \sum_{i=1}^{i} \sum_{j=1}^{j} x_{i,j},$$

where $x_{i,j}$ and $ii_{i,j}$ denote pixels in $\boldsymbol{X}$ and $\boldsymbol{II}$, respectively. This can be calculated very efficient and recusively, under proper edge conditions, by

$$ii_{i,j} = ii_{i-1,j} + ii_{i,j-1} + x_{i,j}.$$

The integral image makes it possible to calculate the sum of any rectangle in the image, by looking at only 4 values in the integral image. Since calculation of an arbitrary 3×3 LBP feature requires 9 rectangles, one would expect $9 \cdot 4 = 36$ values to be needed. However, since some of the values from the integral image are used multiple times, the calculation of an 3×3 LBP feature is possible with only 16 lookups in the integral image.

For the face detection in OpenCV, all detection is done in a window of size 24×24. The 9 rectangles of the LBP are laid out in a 3×3 grid located somewhere in this image. The sum of all the pixels in each rectangle is calculated using the integral image, and thresholded with regard to the rectangle in the middle in order to calculate the 8-bit LBP feature. To each binary 8-bit number a value of success or failure is assigned. In the 24×24 window we can vary the (x, y) offset of the upper left rectangle and s_x, s_y of the 3×3 grid (Liao et al. 2007). This makes it possible to generate approximately 8500 different LBPs within a 24×24 image.

The algorithm uses a moving window over the entire image. First, faces are detected at the smallest allowed scale such as 50×50. If a region does not contain a face, the

window moves a few pixels to the right, and iteratively covers the whole image. After the entire image has been searched, the size of the window is multiplied by a factor (OpenCV default 1.1) and the image is searched with the larger moving window and so on.

Using a large dataset and boosting, a strong classifier can be created as the weighed vote of many weak classifiers. In order to speed up classification, a cascaded classifier is created. For LBPs in OpenCV this consists of 20 stages, where each stage contains from 3 to 10 LBPs. Dependent on the outcome of each LBP, a positive or negative number is assigned. The sum of these numbers are compared with a threshold specific to the current stage, to determine whether the stage has failed. If a stage fails, the window does not contain a face, and the algorithm moves on to the next window. However, if all 20 stages are passed, a face has been detected in that area. Such a detection can be seen in Figure 4.1 on page 17.

KLT feature tracking

Kanade-Lucas-Tomasi (KLT) feature tracking is a very fast algorithm for tracking of feature points in a video. The first step of the algorithm is detection of points for tracking. KLT does this by the the minimum eigenvalue algorithm, which detects points that look like corners and looks for attributes which empirically are known to permit stable tracks. An example of detected points is shown in Figure 2.7.

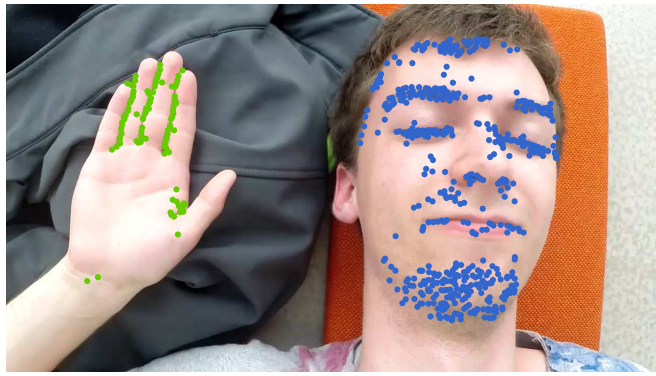


Figure 2.7: Detected feature points in the hand (green) and face (blue) of a subject.

For an adjacent pair of frames the algorithm calculates an approximate translation of each point, and continued application of this leads to a track of all features through the entire sequence of frames. If a small window around a point has changed too much during tracking, the point is assumed to have been wrongly tracked and is discarded.

The tracking from frame to frame attempts find a displacement of the point that minimises the squared differences of the pixels in a small window around the point.

This window can contain a weighting, such as a Gaussian kernel. The displacement vector is calculated iteratively using first order Taylor approximations of the displacement. This usually converges to the correct displacement vector after a few iterations (Tomasi and Kanade 1991).

2.7 Previous work

As mentioned in Section 2.2 on page 3 the term photoplethysmography has been used since the 1930s. For many years all PPG methods required contact with the subject. In the 2000s the idea of non-contact PPG arose. However, at first, expensive cameras and dedicated light sources were used for this.

Verkruysse et al. (2008) were the first to show that PPG signals could be measured remotely from a video of a face using a simple consumer level digital camera and ambient light. Colour movies of test subjects were recorded at either 15 or 30 FPS with a pixel resolution of either 640×480 or 320×240. A region of interest (ROI) was chosen manually and a raw signal was calculated as the average of the pixel values in the ROI. The raw signal was band-pass filtered using a 4th order butterworth filter. The heart rate could then be extracted using fast Fourier transform (FFT). The green channel contained the strongest signal, but the signal was also present in the red and blue signal.

In a paper (Poh et al. 2010) a group from MIT presented an automated way of measuring the heart rate. Videos of 12 subjects (10 males, 2 females) with varying skin colours was recorded using a standard built-in webcamera on a Macbook Pro by Apple Inc. The videos were recorded at 15 FPS and with a pixel resolution of 640×480. For each video Viola-Jones face detection was used to obtain a ROI. Raw signals were found by averaging the pixel values in this ROI. The three RGB signals were then seperated into three independent signals using ICA via the JADE algorithm. It was found that the second component of the ICA typically contained the strongest plethysmographic signal, and it was therefore chosen for simplicity. The heart rate frequency was found in this component as the largest peak in the FFT.

In a second paper (Poh et al. 2011) the group from MIT expanded on their method. Out of the three independent components found with ICA the one having the highest peak in its power spectrum was selected for further analysis. This signal was smoothed with a moving average filter and band-pass filtered using a Hamming window. Then a cubic spline function was used to interpolate the signal at a sampling frequency of 256 Hz. The peaks of the signals were then found in order to form a time series of interbeat intervals (IBIs) from which the heart rate could be calculated directly.

A second group from MIT demonstrated in a paper (Wu et al. 2012) a method of enhancing subtle colour changes and imperceptible motion which they called Eulerian Video Magnification. With their method they could amplify pre-specified frequencies, e.g. frequencies close to a persons heart rate. The output of the algorithm is the input video but with the wanted frequencies amplified. Extraction of the heart rate from

this requires further processing. Also since it requires knowledge about a persons heart rate to enhance the colour changes the blood flow induces, it is not immediately useful for measuring the heart rate.

Several applications of this concept for measuring heart rate have been released. Microsoft uses a combination of this method and an infrared camera to measure heart rate in their Kinect 2 sensor for Xbox One. Several smartphone apps have also been released e.g. "What's my Heart Rate" (ViTrox 2013) for Android and iOS and Philips' "Vital Signs Camera" (Philips 2013) for iOS.

In a conference article (Douniama and Couronné 2007) it was shown that there is a strong linear correlation between what is called the PTT and the systolic blood pressure and some correlation with the diastolic blood pressure. In the study an electrocardiogram (ECG) and a finger PPG sensor were synchronously recorded for 15 healthy subjects. The PTT was calculated as the time between the main peak of the ECG and the front slope of the following pulse wave from the PPG monitor.

CHAPTER 3

Data

3.1 Heart rate measurement

Our data was gathered from 16 subjects (11 male and 5 female), primarily Caucasians and with ages ranging from 21 to 27 years old. For each person a video of one minute length was recorded using the built-in webcam of a laptop (Lenovo X220). The capture was done at a pixel resolution of 640×480 (RGB, 8-bits per channel, manual white balance) at 30 FPS and stored as H.264 encoded video with an average bitrate of 706 kbps. The participants were asked to sit as still as possible during the recording. Figure 3.1 shows a thumbnail from each video. Written consent was obtained from all participants prior to recordings.

Figure 3.1: Thumbnails from video recordings

During each video recording, the participant's PPG signal was monitored using a commercially available PPG sensor (CMS50E by Contec Medical Systems Co. Ltd.). The monitor samples the subject's PPG signal in the finger at 60 Hz (8-bit per sample) and saves it as a `csv` file for later processing.

3.2 Blood pressure estimation

For the blood pressure estimation we have gathered videos of 10 persons (6 males and 4 females) holding their hand next to their face. The subjects range from 21 to 25 years old. The videos were recorded using a GoPro Hero3+ Black Edition (1280×720 at 120 FPS "Narrow" field of view, encoded with H.264 at 30618 kbps). Before processing the videos were downsampled by a factor of 2 to 640×360. In Figure 3.2 on the following page a frame from a video can be seen. The subjects

were laid down and asked to keep their fingers together and straight for the recording. We measured the blood pressure of the subjects before and after the video using a standard wrist blood pressure monitor, the SBM09 from the company Sanitas.

Figure 3.2: Still image from video in the blood pressure dataset.

3.3 FPS corrrections

When recording a video, the frames are not always equidistantly spaced, but they may have small or large jumps in times between two frames. This is an issue because the FFT expects uniformly spaced input data. The cause of this can be that the encoder has been unable to keep up with the camera's frame rate, and has dropped one or more frames. To our advantage, we know what these differences in inter-frame times are because the video contains information about when to display each frame, and not just the average frame rate. These times can be accessed in *MATLAB* using the `mmread` function (Richert 2009). Using this information we can resample the signal with spline interpolation to get a uniformly sampled signal.

CHAPTER 4

Method

4.1 Heart rate measurement

We have tried a number of different methods for extracting the heart rate from the video. The methods do though have several things in common.

ROI extraction

In order to find a PPG signal we are only interested in the part of the image that contains many facial skin pixels, what we denote as the ROI. Since the person might move slightly during the video, we find the region of interest (ROI) for each frame in the video. Firstly, we perform face detection for all frames in the video. If there in some frames are no faces detected, the face from the previous frame is used. In the case of multiple detected faces, we select the one closest to the previously detected face.

After parsing the entire video the bounding boxes of the detected faces are smoothed using a 1 second moving average filter. Since the bounding box contains some background, we crop 20% of both the left and right side. This cropped area is what is referred to as the ROI. An example of a bounding box with corresponding ROI can be seen in Figure 4.1.

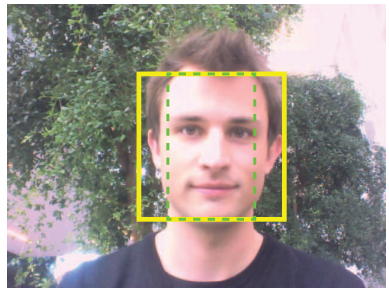


Figure 4.1: Bounding box of detected face in yellow, and selected ROI in green.

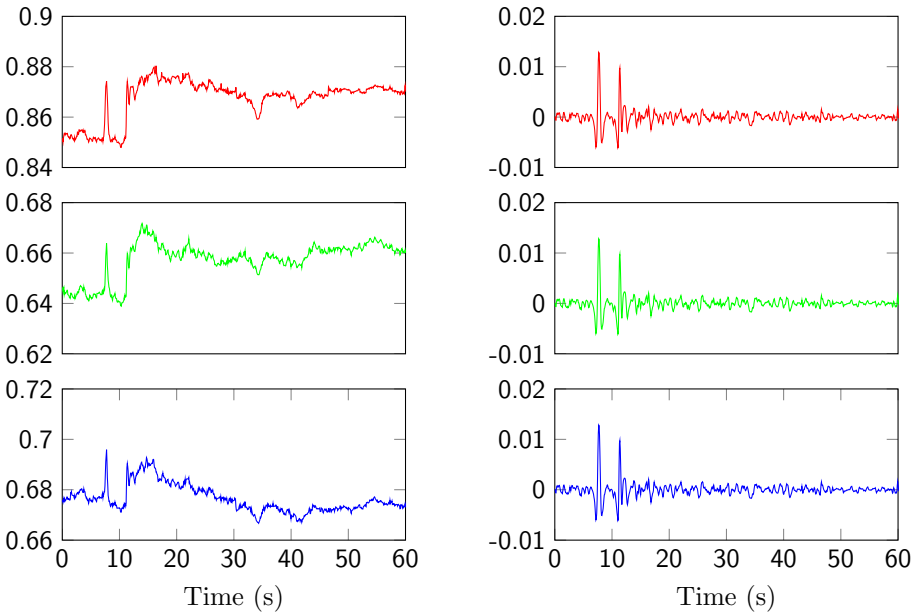
Spatial averaging of ROI and detrending

For each frame, the average value of all pixels in the ROI is calculated. This yields $y_{i,j}$ as the average value of colour $j \in \{\text{red}, \text{green}, \text{blue}\}$ in the ROI of the i'th frame. For an entire movie and a specific colour, j, this is a signal $\boldsymbol{y}_j$.

Over the course of the video there may be long term changes in the averaged ROI values such as lighting changes and exposure adjustments in the camera. To compensate for this, the signal needs detrending. We have performed this by subtracting a moving average smoothed copy of the signal as follows

$$y'_{i,j} = y_{i,j} - \frac{1}{2n+1} \sum_{k=-n}^{n} y_{i+k,j}$$

where $y'_{i,j}$ is the detrended version of $y_{i,j}$. $2n+1$ is the length of the moving average, and it has been determined by cross validation as in Section 4.1 on the facing page. In Figure 4.2 RGB signals are shown before and after they are detrended.



(a) RGB signals

(b) Detrended RGB signals

Figure 4.2: Example of detrending of signals.

Moving average length cross validation

We have used cross validation to determine the optimal length for the moving average filter we use for detrending. We use leave-one-out for the persons, and calculate the optimal length of the filter for each group of 15 persons. Then we use the median length as the length for our experiments.

PPG signal identification

After calculation of the ICA on the detrended signals, it is not known which component is the sought after PPG signal. To identify this, we begin by normalizing the variance of each ICA component. Afterwards the FFTs of all the signals are calculated. Then the one with the largest amplitude in the interval from 40 to 200 BPM is chosen as the PPG signal. The frequency of this peak, f is converted to BPM, which yields the heart rate.

$$\text{Heart Rate} = \frac{60 \cdot \text{FPS} \cdot \text{idx}}{\text{Length of FFT}}$$

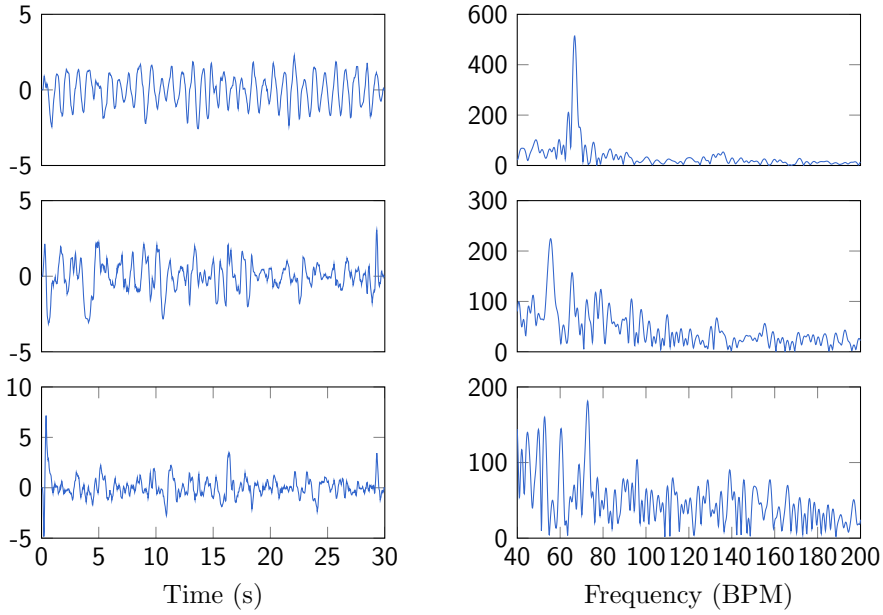
where idx is the zero-based index of the peak in the Fourier transform. In Figure 4.3 on the following page the three ICA components and their Fourier transforms are shown. It is immediate that component 1 has the highest peak, hence this is chosen as the heart rate.

Alternative PPG signal identification methods

We have also considered several alternatives to identifying the correct PPG signal among the ICA components other than the largest peak in the Fourier transforms. Because of Section 2.2 on page 4 we expect to find the clearest PPG signal in the green channel, since blood has the most absorbtion in the green part of the visible spectrum. Because of this fact, we could identify the PPG signal as the ICA component with the largest relative weight assigned to the green channel. This method is, however, only applicable to the RGB colour space.

Alternatively, we could use the knowledge that the ICA components we need to choose between look radically different. One hopefully looks like a sinusoid, while the others look more or less like Gaussian noise. The kurtosis of a sinusoid is approximately 1.5, while Gaussian noise has a kurtosis of 3. It is thus plausible to look for the component with the minimal kurtosis when seeking the PPG signal.

Under the assumption that the other two components are mostly noise, it could be interesting to select the component with the maximum signal-to-noise ratio (SNR). This would make sense because the PPG signal contains a clear signal in comparison with the other components. We have utlilized the SNR function built in in *Matlab*. This function assumes that the underlying signal is a sinusoid, which is determined from a periodogram of the function. See the *Matlab* reference for further information.



(a) ICA components

(b) Fourier tranforms of the ICA components

Figure 4.3: Example of three ICA components and their respective Fourier transforms.

Fixed weight calculation

After calculation of the heart rate for many videos, we are interested in obtaining an average weight of our colour space, which will yield our PPG signal. We only look at the weights for ICA components identified to contain a PPG signal.

1. All the weights are normalized.
2. The component with the highest mean absolute weight is chosen to have the same sign for all weights.
3. The average weight is calculated.

Heart rate for comparison

For evaluating the quality of our method we need a correct heart rate to compare with. This was found by applying the same detrending and Fourier transform to data from the finger PPG probe. The window used was selected to be as close to the video as possible with an accuracy of ± 0.5 s.

Method overview

We have tried a variety of different methods for extracting a PPG signal. We have outlined these below along with their motivations.

RGB and ICA

In this method the RGB colourspace is left untouched, and ICA is applied after detrending. This is very similar to Poh et al. (2010).

1. Extraction of the ROI and spatially averaging it.
2. Detrending with moving average.
3. ICA analysis of the detrended signals.
4. The PPG signal is identified.
5. The heart rate is computed using the FFT.

RGB with fixed weights

This variation of our method, doesn't include ICA, but uses a fixed weighting of the RGB components to obtain the PPG signal. The fixed weights are chosen as described in Section 4.1 on the facing page.

1. Extraction of the ROI and spatially averaging it.
2. Detrending with moving average.
3. Summing RGB with fixed weights to obtain final PPG signal.
4. The heart rate is computed using the FFT.

Green

From Section 2.2 on page 4 we expect the strongest PPG signal in the green channel, since blood has the most absorbtion in the green part of the visible spectrum. Thus we only to use at the green channel for a very simple method. This is very similar to what Poh et al. (2010) did.

1. Extraction of the ROI and spatially averaging it.
2. Discard the red and blue channels, keeping only the green.
3. Detrending with moving average.
4. The heart rate is computed using the FFT.

HSV and ICA

HSV and ICA is very similar to the RGB and ICA method. The only difference is the colour space used.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to HSV.
3. Detrending with moving average.
4. ICA analysis of the detrended signals.
5. The PPG signal is identified.
6. The heart rate is computed using the FFT.

HSV with fixed weights

Similar to RGB with fixed weights. The fixed weight is found as described in Section 4.1 on page 20.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to HSV.
3. Detrending with moving average.
4. Summing RGB with fixed weights to obtain final PPG signal.
5. The heart rate is computed using the FFT.

Hue

From Section 2.4 on page 6 we expect hue to show the colour, thus making it nice for measurements for differences in colour which should correspond to changes in the amount of blood. Thus we try to only look at hue for a signal.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to HSV.
3. Discard the saturation and value channels, keeping only the hue.
4. Detrending with moving average.
5. The heart rate is computed using the FFT.

Saturation

From Section 2.4 on page 6 we know saturation shows how colourful a given colour is. Changes in the amount of blood, might show in this number, and we thus try to only look at saturation for a signal.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to HSV.
3. Discard the hue and value channels, keeping only the saturation.
4. Detrending with moving average.
5. The heart rate is computed using the FFT.

Lab and ICA

Lab and ICA is again very similar to the RGB and ICA method. The only difference is the colour space used.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to HSV.
3. Detrending with moving average.
4. ICA analysis of the detrended signals.
5. The PPG signal is identified.
6. The heart rate is computed using the FFT.

Lab with fixed weights

Similar to RGB with fixed weights. This fixed weight can be found as described in Section 4.1 on page 20.

1. Extraction of the ROI and spatially averaging it.
2. Convert the signals from RGB to Lab.
3. Detrending with moving average.
4. Summing RGB with fixed weights to obtain final PPG signal.
5. The heart rate is computed using the FFT.

Best linear combination

A linear model like the other methods. We examine many different linear combinations of RGB as seen in Figure 4.4 on the following page. We only need to look at one half of a full sphere, since the other half would yield the same results, only with opposite sign. All linear combinations are normalized with respect to variance, and the PPG signal is selected as the one with the largest peak. Due to linearity of the Fourier transform we only calculate three Fourier transforms, one for each RGB channel. Then we can calculate the Fourier transform of an arbitrary linear combination of RGB as a linear combination of the RGB Fourier transforms.

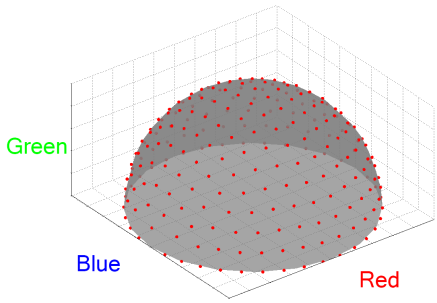


Figure 4.4: Each red dot represents a unique linear combinations of RGB as examined in "Best linear combination" method.

1. Extraction of the ROI and spatially averaging it.
2. Detrending with moving average.
3. Calculating most linear combinations of RGB.
4. The PPG signal is identified.
5. The heart rate is computed using the FFT.

Comparison of methods

To compare the different methods for measuring heart rate we test the equality of the variances of their differences with the reference heart rate. Usually an F-test is used for testing equality of variances. However since this is very sensitive to non-normality we have instead used the Brown-Forsythe test (Brown and Forsythe 1974) which is less sensitive to non-normality. We do this as we do not expect the errors of the different methods to be normally distributed.

4.2 Short time heart rate estimation

In our heart rate measurements we have used windows of either 30 or 60 seconds. We want to examine the relationship between the error and the length of the window.

We have taken 6 test subjects (3 men and 3 women) from the data set who generally gave good measurement results. Their heart rates are calculated with the RGB and ICA method with windows length ranging between 1 and 60 seconds with an increment size of 0.5 seconds.

4.3 Heart rate intensity map

In this section we present a method for measuring how the amplitude of the heart rate is distributed in a body part, e.g. the face.

1. The portion of the video covering the body part is stabilized using KLT feature tracking.
2. Each frame is smoothed using a Gaussian kernel.
3. The heart rate is determined in each pixel using the RGB and ICA method.
4. The median heart rate is chosen as the correct heart rate.
5. In each pixel the amplitude of this heart rate is found.

4.4 Blood pressure estimation

To investigate the relationship between blood pressure and PTTD we use the method described in this section to compute the PTTD between the face and a hand. In order to work under optimal conditions we have decided to record videos of people lying down, which places the hand approximately at heart rate level. This should result in less variance in our PTTDs (Douniama and Couronné 2007). To minimise artefacts from pixel noise and sampling rate, we opted for a high quality camera with a high frame rate.

Difference in pulse transit time

In this section we describe the two different methods we have used to estimate the PTTD.

Fourier transform angle based method

One way of calculating the phase difference between two signals is using the phase information kept in the Fourier transform. For the sequence of N (possibly complex) numbers $x_0, x_1, \ldots, x_N - 1$ the discrete fourier transform is defined as

$$X_k = \sum_{n=0}^{N-1} x_n e^{\frac{-i2\pi kn}{N}}, \ k \in \mathbb{Z},$$

where each X_k is a complex number whose absolute value and angle respectively describes the amplitude and phase of a sinusoidal component of the sequence x_n. The phase of a periodic signal can therefore be found as the phase of that component which has the maximum amplitude. To compare the phases of two signals the sinusoidal components chosen must be same. Instead of finding the phase for the maximum amplitude of each signal, the phase is instead at the maximum of the first signal (this is only plausible when the two maxima are close). This yields two phases θ_1, θ_2 of the periodic signals from which the phase difference can be found by subtracting the two phases

$$\theta = \theta_1 - \theta_2,$$

which subsequently can be converted into a time delay by

$$T = \frac{\theta}{2\pi f}$$

where the f is the frequency of the peak.

Cross correlation based method

Another way of calculating the phase difference between two signals is by using cross-correlation. The cross-correlation for real discrete functions is defined as

$$(f \star g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[m+n].$$

The maximum of the cross-correlation indicates where the two signals are best aligned, hence the time delay, or phase difference, is found as

$$T = \arg\max_t (f \star g)(t).$$

For periodic signals the cross-correlation will be periodic as well, hence the possible values t can take must be limited to one period.

Method overview

The method for estimating the PTTD is outlined is outlined below. When calculating PPG signals for the hand and face they are weighted by their HRIMs to weigh those regions that contain more signal higher, thereby obtaining cleaner PPG signals. This makes the signals easier to compare when calculating the phase difference between them.

1. A HRIM is calculated for the hand and face respectively.
2. A PPG signal is found in each pixel using the RGB and ICA method. Each of these PPG signals are weighted by their respective HRIMs pixel.
3. The PTTD is calculated in one of two ways; based on the Fourier transform or based on the maximum cross-correlation.

After the PTTD has been computed the relationship between the PTTD and the blood pressure can be examined using scatterplots.

CHAPTER 5

Results

5.1 Statistics

In the result section we have used several statistics. They are

Mean error - the mean value of the differences between the measured and the reference heart rate.

Mean absolute error - the mean value of the absolute differences between the measured and the reference heart rate.

Standard deviation of error - the sample standard deviation of the differences between the measured and the reference heart rate.

Root mean square error (RMSE) - the root mean square error of the differences between the the measured and the reference heart rate.

Correlation coefficent - Pearson's correlation coefficient which is a measure of the linear dependence between two variables.

- Significance of correlation coefficient - a p-value is computed using the Student's t-test with the the null hypothesis that the correlation is zero against the alternative hypothesis that the correlation is not zero. A small p-value indicates that the correlation is not zero, and a large value the opposite.

Success rate - this is defined as the percentage of correctly found heart rates. We accept a heart rate as correct if it is within 2 BPM of the reference heart rate.

Success rate ICA - this is defined as the percentage that the correct heart rate is found as the maximum amplitude in one of the three ICA components. Again a heart rate is accepted as correct if it is within 2 BPM of the reference heart rate.

5.2 Face Detection

The quality of the face detection was very good with very few incorrect detections. The confusion matrix in Table 5.1 on the next page shows this nicely. It should be noted that the 1.69% false negatives are all from one video.

Table 5.1: Confusion matrix for face detection

Detected face \Face present	True	False
True	98.3%	0%
False	1.69%	—

5.3 Moving average length cross validation

Detrending of the signals was performed using a moving average. The length was found via cross validiation described in Section 4.1 on page 19 to be 37 frames.

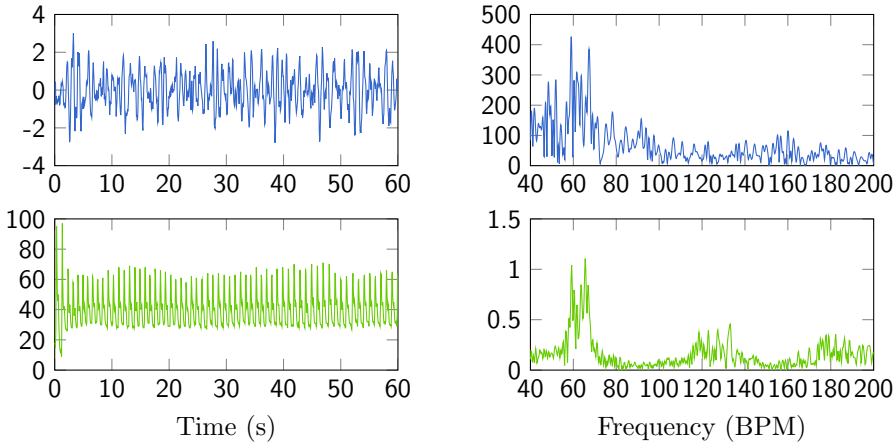
5.4 Heart rate measurement

The different methods for measuring heart rate are denoted by:

M1:	RGB and ICA	**M2**:	RGB and fixed weights
M3:	Green	**M4**:	HSV and ICA
M5:	HSV and fixed weights	**M6**:	Hue
M7:	Saturation	**M8**:	Lab and ICA
M9:	Lab and fixed weight	**M10**:	Best linear combination

Measuring on 60 second videos

Figure 5.3 on the next page shows a complication that might arise when measuring the heart rate over longer time periods. In the Fourier transform of the reference signal (the lower in Figure 5.3(b) on the facing page) there are two peaks, both very large (this is due to the subjects heart rate changing over the 60 seconds). The most significant ICA component might have either as its maximum peak. The upper fourier transform in Figure 5.3(b) on the next page shows an example where the ICA has the "wrong peak" (it is actually the correct heart rate in the latter part of the video). This will result in a wrong measurement although in reality it is not completely wrong. The methods not using ICA have the same issue.



(a) ICA component and reference signal.

(b) Corresponding Fourier transforms.

Figure 5.3: The maximum peak of two signals being very different for similar Fourier transforms.

Averaging 30 second videos

Because of the complications described in the previous section we have chosen to work with 30 second videos instead. To get results for a full 60-second video the different measuring methods are used on 30-second moving windows with 1 second increments. The 31 measurements for each person are then averaged to get a measurement for the full 60 seconds videos. In Table 5.4 on the following page the statistics described in Section 5.1 on page 27 are shown for these average-measured heart rates. All correlation coefficients are greater than 0.9 and very significant. However M6 (Hue) and M7 (Saturation) have greater standard deviations, RMSEs and lower success rates than the rest, suggesting that they are not as good as the rest of the methods. It is seen that M10 (Best linear combination) has a 100% success rate, the highest correlation coefficient, the lowest standard deviation and RMSE.

Table 5.4: Statistics for the 10 heart rate measurement methods used on 60-second videos as averages of 30-second moving windows.

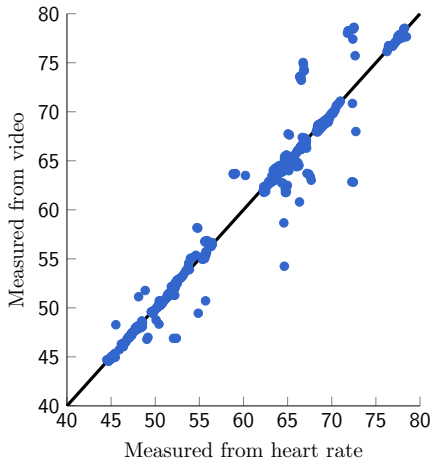
Statistic	M1	M2	M3	M4	M5
Mean bias (BPM)	0.04	-0.07	-0.10	0.03	0.09
Mean absolute error (BPM)	0.47	0.38	0.60	0.45	0.43
Standard devation of error (BPM)	0.84	0.57	0.85	0.76	0.64
RMSE (BPM)	0.82	0.56	0.83	0.73	0.63
Correlation coefficient	$0.996^{\dagger}$	$0.998^{\dagger}$	$0.996^{\dagger}$	$0.997^{\dagger}$	$0.998^{\dagger}$
Success rate	93.8%	100.0%	100.0%	93.8%	100.0%
	M6	M7	M8	M9	M10
Mean bias (BPM)	-1.04	-0.37	0.04	-0.10	0.10
Mean absolute error (BPM)	2.06	1.11	0.47	0.44	0.31
Standard devation of error (BPM)	3.11	1.92	0.84	0.67	0.53
RMSE (BPM)	3.18	1.90	0.82	0.66	0.52
Correlation coefficient	$0.941^{\dagger}$	$0.978^{\dagger}$	$0.996^{\dagger}$	$0.997^{\dagger}$	$0.998^{\dagger}$
Success rate	68.8%	81.3%	93.8%	100.0%	100.0%

† indicates significance at $p < 10^{-7}$.

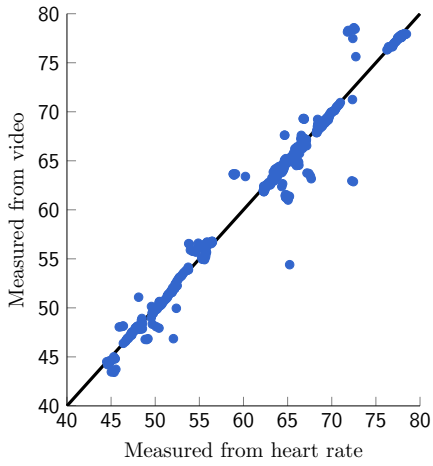
Measuring on 30 second windows

In this section we present result for the 10 measuring methods used on the 30-second moving windows used for the averaging in the the last section. These results are useful in the context of continuous monitoring of the heart rate. In total the moving windows gives $16 \cdot 31 = 496$ data points. Correlation plots for the 10 methods against the reference heart rate are shown in Figure 5.5 on page 33. The correlation plots for M3 (Green), M6 (Hue) and M7(Saturation) are more scattered than the rest which look similar to each other.

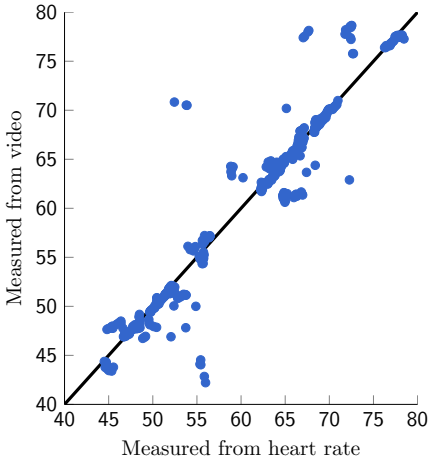
(a) RGB and ICA



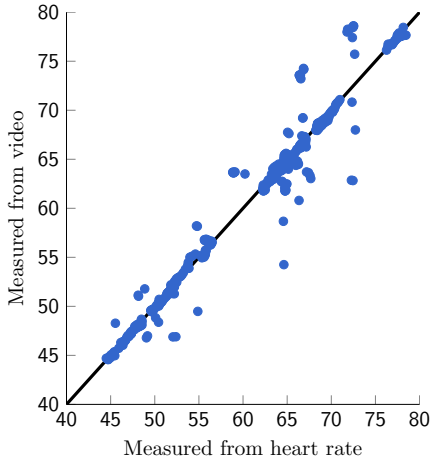
(b) RGB and fixed weights



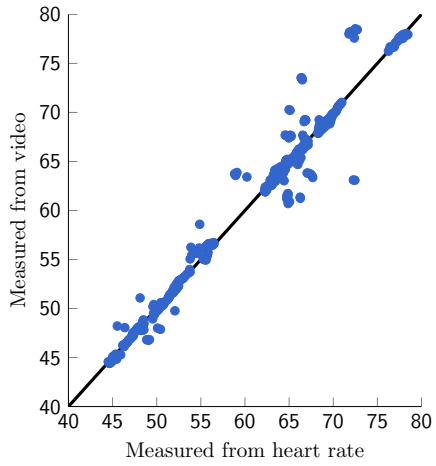
(c) Green



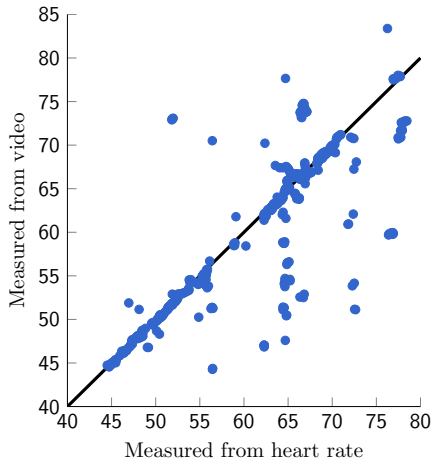
(d) HSV and ICA



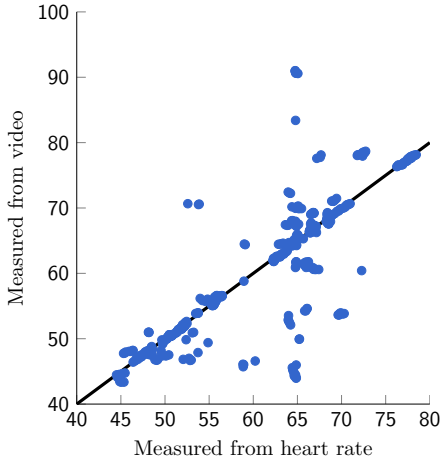
(e) HSV and fixed weights



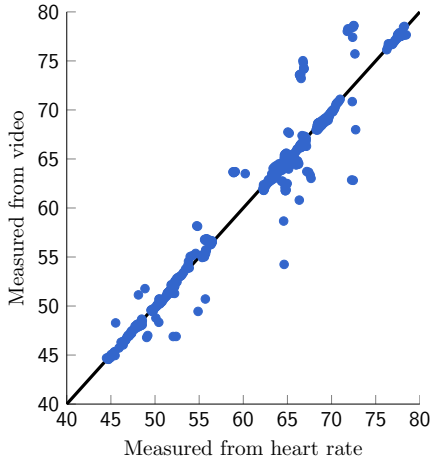
(f) Hue



(g) Saturation



(h) Lab and ICA



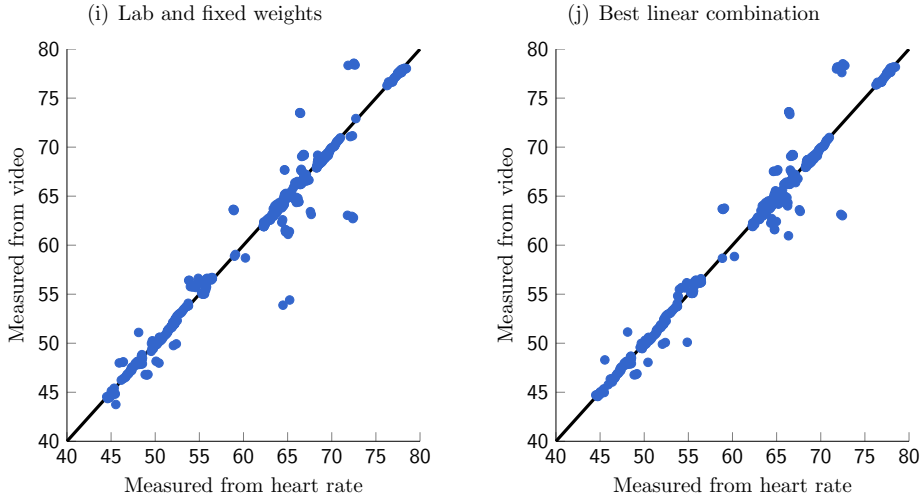


Figure 5.5: Corelation plots of all 10 methods. The black lines indicate the line of equality.

The statistics in Table 5.6 on the next page supports the larger scattering of M3 (Green), M6 (Hue) and M7 (Saturation) as all of these have have greater standard deviations and RMSEs and smaller correlation coefficients. M6 and M7 have a lower correlation coefficient as well. M2 (RGB with fixed weights), M8 (Lab with fixed weights) and M10 (Best linear combination) have the highest success rates, smallest standard deviations and RMSEs.

Table 5.6: Statistics for the 10 heart rate measurement methods used on 30-second moving windows.

Statistic	M1	M2	M3	M4	M5
Mean error (BPM)	0.04	-0.07	-0.10	0.03	0.09
Mean absolute error (BPM)	0.85	0.69	1.19	0.83	0.79
Standard devation of error (BPM)	1.92	1.53	2.64	1.84	1.71
RMSE (BPM)	1.91	1.53	2.64	1.84	1.71
Correlation coefficient	$0.978^{\dagger}$	$0.986^{\dagger}$	$0.960^{\dagger}$	$0.980^{\dagger}$	$0.983^{\dagger}$
Success rate	87.9%	89.5%	83.1%	87.9%	86.3%
Success rate ICA	93.3%	—	—	93.1%	—
	M6	M7	M8	M9	M10
Mean error (BPM)	-1.04	-0.37	0.04	-0.10	0.10
Mean absolute error (BPM)	2.40	2.67	0.85	0.72	0.66
Standard devation of error (BPM)	4.99	5.83	1.92	1.69	1.54
RMSE (BPM)	5.09	5.84	1.91	1.69	1.54
Correlation coefficient	$0.848^{\dagger}$	$0.834^{\dagger}$	$0.978^{\dagger}$	$0.983^{\dagger}$	$0.986^{\dagger}$
Success rate	75.2%	71.0%	87.9%	89.7%	89.9%
Success rate ICA	—	—	93.3%	—	97.6%*

† indicates significance at $p < 10^{-128}$.

* This is the percentage that the correct heart rate is the maximum amplitude for one of the linear combinations and not "Success rate ICA".

Fixed weights

For the three methods which employ fixed weights these were found, as described in Section 4.1 on page 21. The found weights are

$$\begin{bmatrix} \mathrm{R}_{\mathrm{weight}} \\ \mathrm{G}_{\mathrm{weight}} \\ \mathrm{B}_{\mathrm{weight}} \end{bmatrix} = \begin{bmatrix} -0.250 \\ 0.764 \\ -0.285 \end{bmatrix}, \quad \begin{bmatrix} \mathrm{H}_{\mathrm{weight}} \\ \mathrm{S}_{\mathrm{weight}} \\ \mathrm{V}_{\mathrm{weight}} \end{bmatrix} = \begin{bmatrix} -0.202 \\ 0.790 \\ -0.145 \end{bmatrix}, \quad \begin{bmatrix} \mathrm{L}_{\mathrm{weight}} \\ \mathrm{a}_{\mathrm{weight}} \\ \mathrm{b}_{\mathrm{weight}} \end{bmatrix} = \begin{bmatrix} -0.217 \\ 0.7638 \\ -0.276 \end{bmatrix}. \quad (5.1)$$

It is seen that the green channel is weighted higher than the red and blue channels, saturation is weighted higher than hue and lightness and that a is weighted higher than Lightness and b.

Equality of variance

In Table 5.7 on the next page the p-values for pairwise Brown-Forsythe tests (Brown and Forsythe 1974) are shown. For method method 6 (Hue) and method 7 (Saturation) we can reject the null hypothesis of equal variance with all other methods (but not with each other) at a significance level of $\alpha = 0.01$. For method 3 (Green) we can reject the null hypothesis of equal variance at a significance level of $\alpha = 0.05$.

Table 5.7: P-values for pairwise Brown-Forsythe tests.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
M1	1	0.14	0.01	0.82	0.67	0.†	0.†	0.99	0.22	0.07
M2	—	1	0.†	0.21	0.27	0.†	0.†	0.14	0.83	0.71
M3	—	—	1	<0.01	<0.01	0.†	0.†	0.01	<0.001	0.†
M4	—	—	—	1	0.84	0.†	0.†	0.83	0.31	0.11
M5	—	—	—	—	1	0.†	0.†	0.67	0.40	0.15
M6	—	—	—	—	—	1	0.34	0.†	0.†	0.†
M7	—	—	—	—	—	—	1	0.†	0.†	0.†
M8	—	—	—	—	—	—	—	1	0.22	0.07
M9	—	—	—	—	—	—	—	—	1	0.58
M10	—	—	—	—	—	—	—	—	—	1

† indicates $p < 0.0001$.

5.5 Short time heart rate estimation

In Figure 5.8 on page 37 results from the experiment described in Section 4.2 on page 24 are shown. Figure 5.8(a) on page 37 shows the mean absolute error made by the RGB and ICA method as a function of the window length used. Figure 5.8(b) on page 37 shows the RMSE as a function of the window length. Figure 5.8(c) on page 37 show the success rate as a function of the window length. In Figure 5.9 on page 38 one can see the heart rate found from the finger PPG sensor, as a function of window size. This is to visualize the bias introduced by the Fourier transform for shorter windows.

5.6 Component Selection Comparison

We have compared the different selection methods identifying the correct PPG signal among the ICA components. The success rates for the different methods can be seen in Table 5.10 on the following page. It can be seen that "Most green" achieves similar performance to "Max peak", while the others perform considerably worse. In Tables 5.11 to 5.13 on the next page, we present confusion matrices that compare "Max peak" to each of the other methods. We have only tested them on the cases where the correct heart rate is present in the ICA components. It can be seen that "Max peak" and "Most green" fail on slightly different cases, while the other methods fail more on the same cases as "Max peak".

Table 5.10: Success rates for different component selection methods.

Max peak	87.9%
Most green	86.1%
Kurtosis	77.4%
SNR	73.2%

Table 5.11: Confusion matrix for max peak and most green PPG signal identification methods

Max peak\Most green	True	False
True	384	18
False	9	10

Table 5.12: Confusion matrix for max peak and kurtosis PPG signal identification methods

Max peak\Kurtosis	True	False
True	376	60
False	8	19

Table 5.13: Confusion matrix for max peak and SNR PPG signal identification methods

Max peak\SNR	True	False
True	359	77
False	4	23

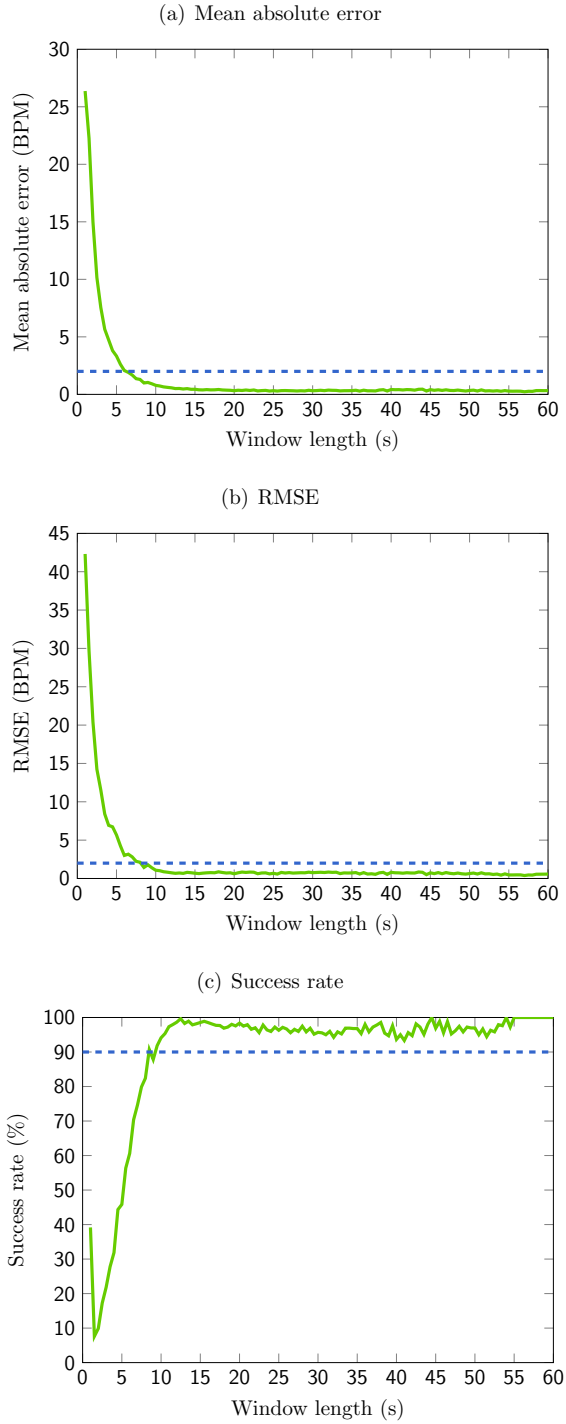


Figure 5.8: Quality measures as a function of the window length.

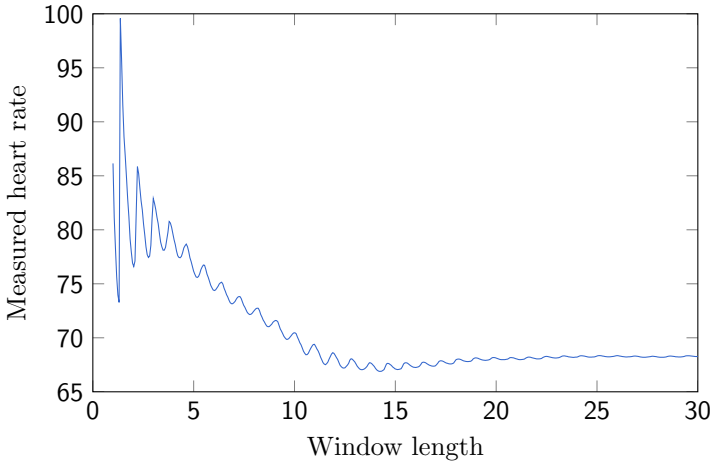
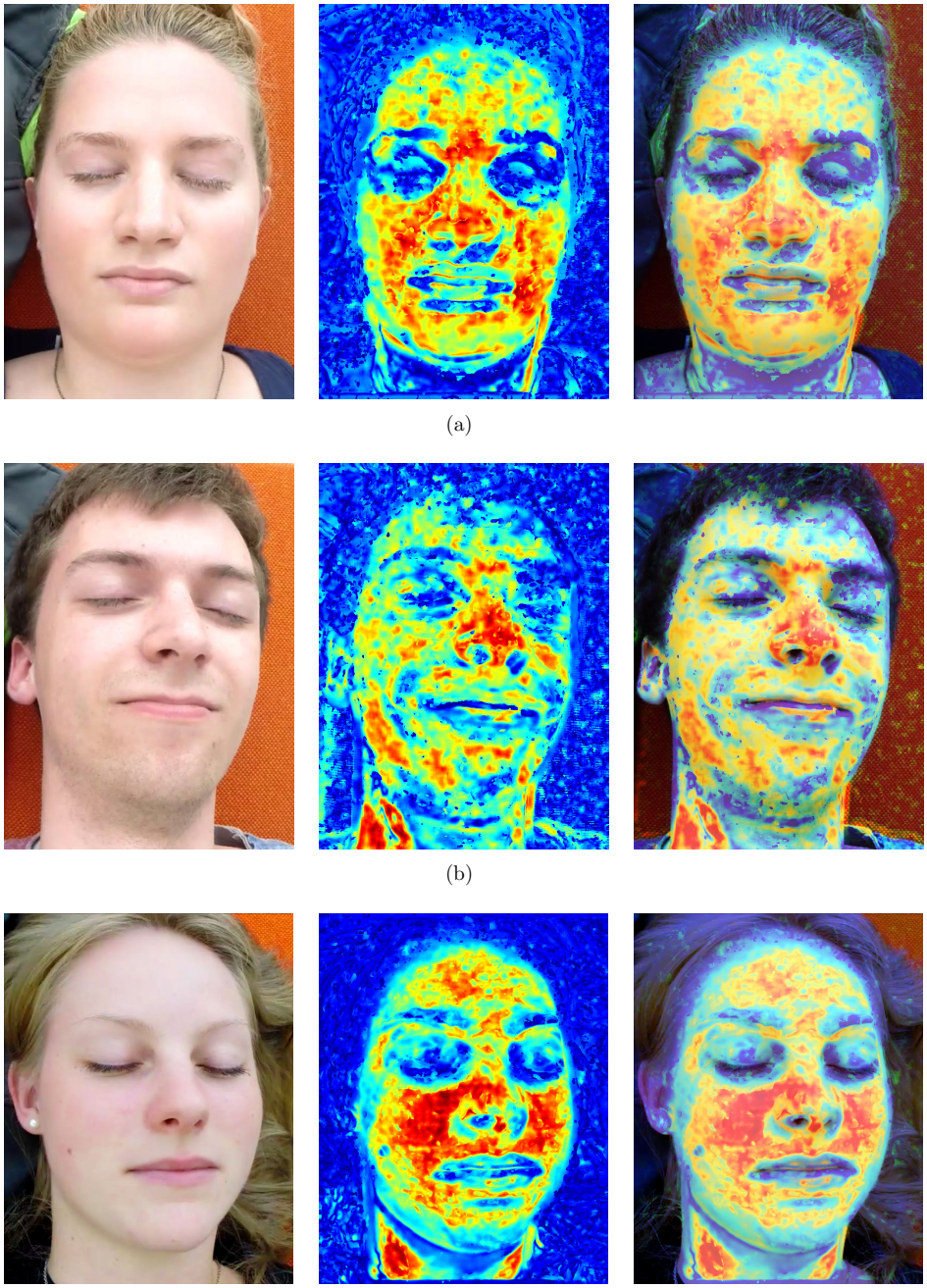


Figure 5.9: The heart rate measured from a finger PPG signal as a function of window length.

5.7 Heart rate intensity map

The method described in Section 4.3 on page 24 have been used for calculating HRIMs for a hand and the face of the test subjects described in Section 3.2 on page 15. Figure 5.14 on the following page and Figure 5.15 on page 41 shows examples of the HRIMs for the faces and Figure 5.16 on page 42 shows examples for the hands. Red colours describes high intensities and blue colours describe low intensities. It should be noted that the colour scale is not absolute and HRIMs for different persons are not immediatetly comparable. In the hands there are not any immediate visual pattern in the HRIMs. In the faces a pattern however can be seen, namely high intensities in the cheek, nose and lower forehead. It should be noted that these are relatively good examples but the tendency in the rest are the same.



(a)

(b)

(c)

Figure 5.14: Three examples of a heart rate intensity map for a face overlayed on an image of the subjects face. Red colours indicates higher intensities while blue colours indicate smaller.

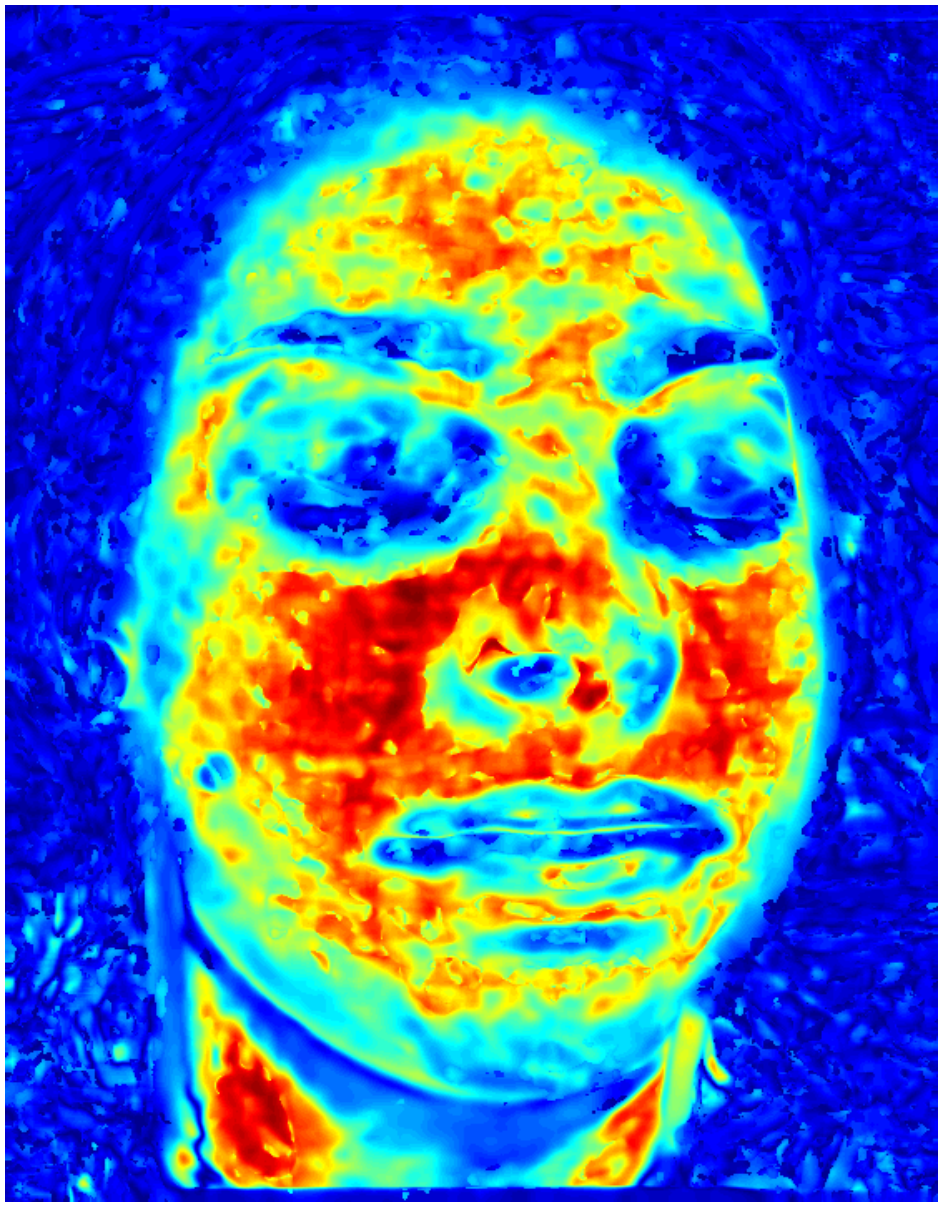


Figure 5.15: Example of heart rate intensity map for a face. Red colours indicates higher intensities while blue colours indicate smaller.

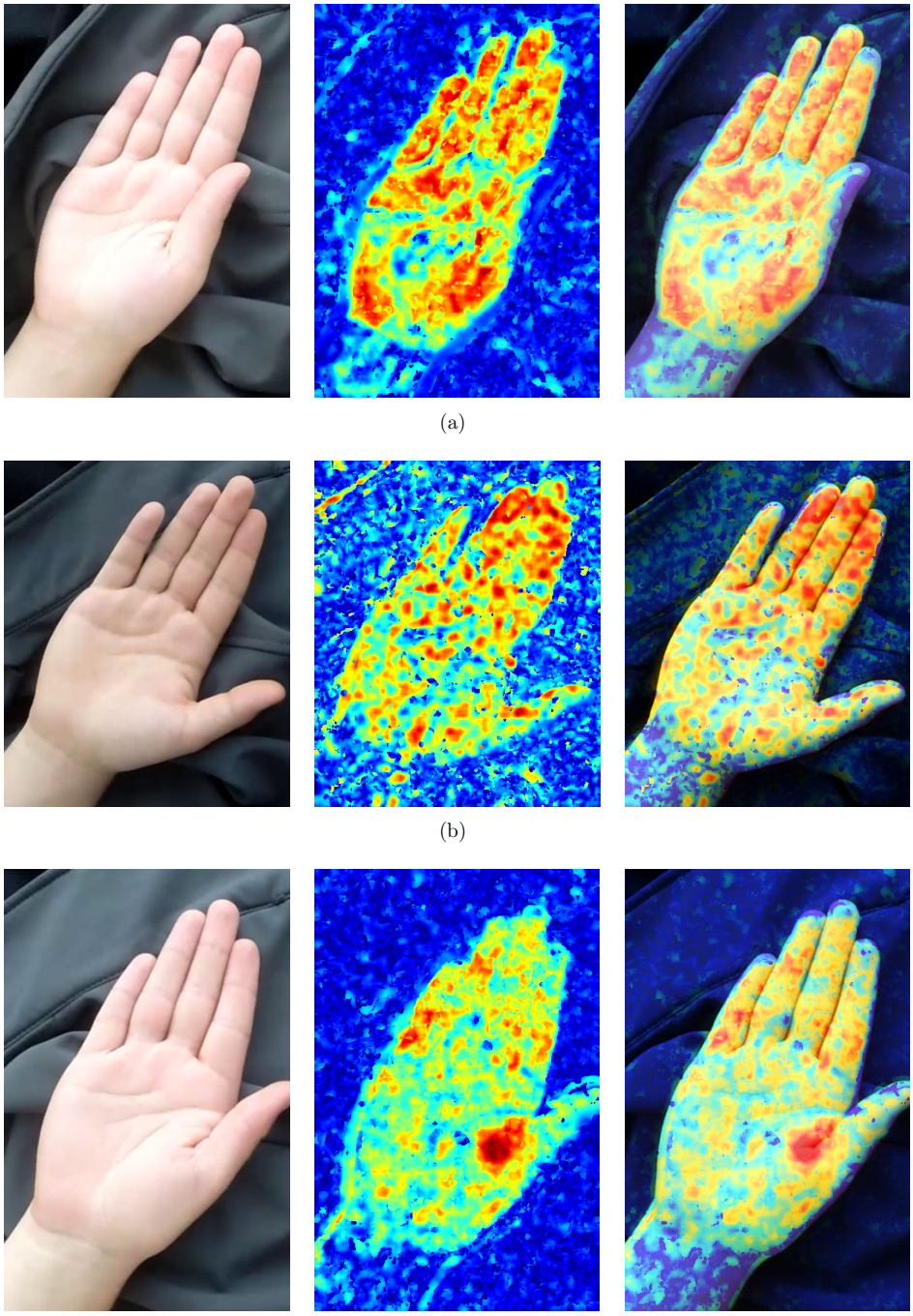


Figure 5.16: Three examples of a heart rate intensity map for a hand overlayed on an image of the subjects hand. Red colours indicates higher intensities while blue colours indicate smaller.

5.8 Blood pressure estimation

The blood pressures of the subjects changed slightly before and after recording the video. The correlation plot can be seen in Figure 5.17, and overall the results seem slightly lower for the after measurement. The PTTD has been calculated for the videos of the 10 subjects from Section 3.2 on page 15 with the method described in Section 4.4 on page 25. The result for cross correlation can be seen in Figure 5.18 on the following page and the Fourier transform angle method is in Figure 5.19 on the next page. The observed correlations are presented in Table 5.20. We have also calculated the correlations without observation 8, which has an abnormally large PTTD.

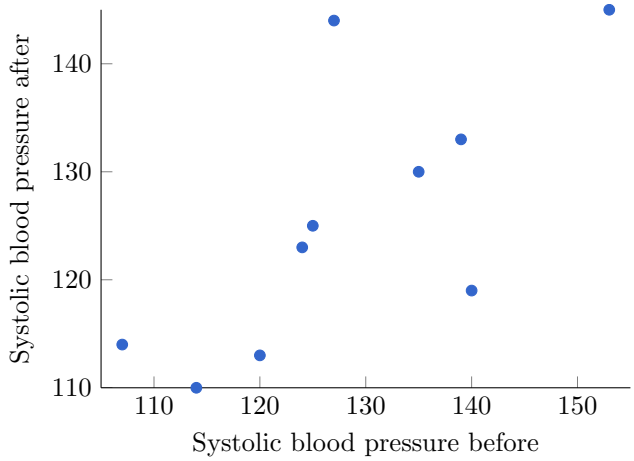


Figure 5.17: Correlation plot of systolic blood pressure before and after recording the video.

Table 5.20: Correlations of PTTD with blood pressure after.

Method	Blood pressure	All observations		Excluding observation 8	
		Corelation	p	Correlation	p
Cross correlation	Systolic	-0.337	0.340	-0.794	0.011
Fourier angle difference	Systolic	-0.543	0.105	-0.807	0.009
Cross correlation	Diastolic	0.067	0.850	-0.239	0.537
Fourier angle difference	Diastolic	0.161	0.657	0.040	0.919

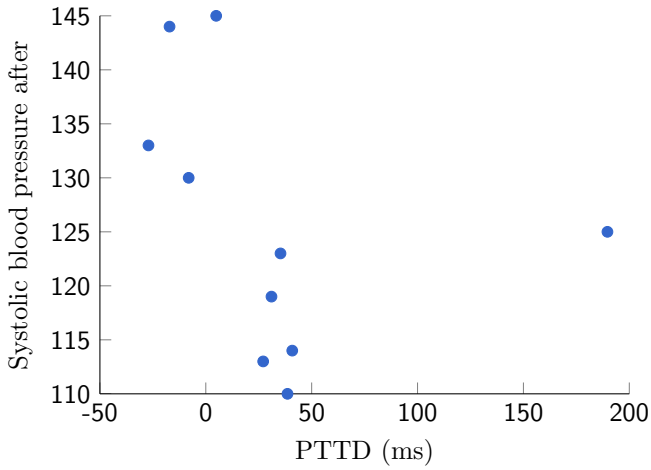


Figure 5.18: Correlation plot of PTTD using cross corelation vs systolic blood pressure.

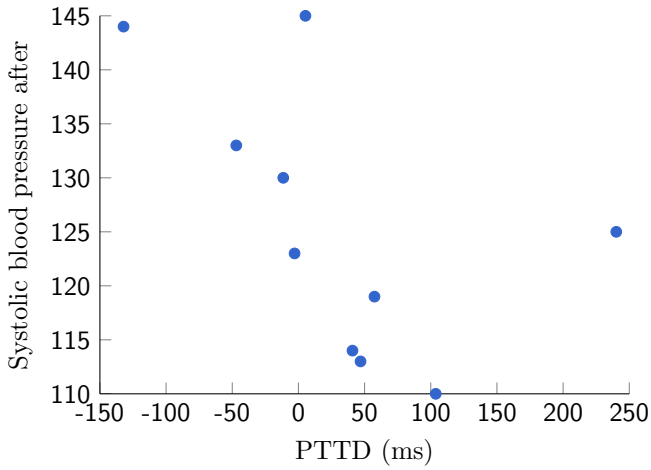


Figure 5.19: Correlation plot of PTTD using Fourier transform angles vs systolic blood pressure.

CHAPTER 6

Discussion

6.1 Face detection accuracy

The face detection we have used is very good at detecting faces. The position and scale of the detected face are however not very accurate. This cause large changes from frame to frame in our detected bounding boxes. This is the reason we chose to smooth the detected bounding boxes with a moving average filter.

We considered using an active appearance model of the face (Edwards et al. 1998) to improve the positioning of the face while simultaneously correcting partially for perspective distortions. This was not pursued since the active appearance model available from DTU only maps the lower part of the face, and none of the forehead, where a significant portion of the heart rate resides as seen in Figure 5.14 on page 40. In order to train a model including the forehead, we would have needed to annotate large amounts of data, and we decided to spend our time otherwise.

An alternative to improve the position accuracy could be to run a Viola-Jones eye detector (Viola and Jones 2004) on the area where the face has been detected. This would hopefully result in an improvement in position accuracy and would even make it possible to correct for small rotations of the face.

A third alternative to these inaccuracies would be to use KLT feature tracking. We would only perform face detection on the first frame, and use the feature tracker to follow the position of the face through the rest of the video. This was not done due to time constraints, and would be a nice to have in future works.

6.2 Detrending parameter selection

We used cross validation to select the detrending parameter. When cross validation is used, there is always the possibility of overfitting. In our case this makes the detrending parameter optimised for the heart rates actually present in our dataset. Because of the way our detrending subtracts a signal smoothed to this exact parameter, it is most suitable for removing noise for exactly one heart rate:

$$\frac{2n+1}{FPS} \cdot 60 = \frac{37}{30} \cdot 60 = 74 \text{ BPM}.$$

The further a heart rate lies away from this value, the less effective will the detrending be. This problem could be resolved by using a different detrending method, for

example the method described by Tarvainen et al. (2002). We have experimented briefly without success with this method, but not enough to bring a comparison.

6.3 PPG signal identification

From the confusion matrices in Tables 5.11 to 5.13 on page 36 we know that the different methods fail on different cases. Especially "Most green" and "Max peak" both have high performance and different cases where they fail. Because of this it is possible to create a combined classifier that utilises one or more of these measures to obtain a higher correct classification rate. One could also simply make up a large number of possible measures and apply feature selection and use these to train a better classifier. From Table 5.6 on page 34 we know that the upper limit on the success rate is 93.3%, if we stick to RGB and ICA. If our PPG signal identification method is good enough, we may even be able to achieve a success rate close to 97.6%. This would be possible by using the "Best linear combination" method, coupled with this new superior PPG signal identification method.

It may also be possible to increase this number further, by introducing a possible outcome to be that no heart rate could be found. This would especially be useful in situations involving an end user, where it is very easy to ask them to conduct a new experiment, hopefully yielding cleaner data.

6.4 Heart rate measurement methods

All 10 heart rate measurement methods have been shown to have high correlation with a reference heart beat. An important remark is that the different samples are not independent as we have used moving windows as samples, and as such the correlation coefficients are largely influenced by this. The methods based on linear models (all except M3, M6 and M7) all had success rates close to 90% and with very similar standard deviations and RMSEs Table 5.6 on page 34. The best of the lot proved to be M10 (best linear combination) which the highest success rate, the smallest mean absolute error, the smallest standard deviation an RMSE. Furthermore it has the greatest promise for improvement with an improved PPG signal identification method. Due to the linearity of the Fourier transform is it also feasible for a real-time implementation. The RGB and ICA has a RMSE of 1.91. This is less than the the RMSE of 2.25 found by Poh et al. (2010). The RMSE of M10 was even better at 1.54, so we were able to not only achieve similar results but in fact slightly better results.

Linear combinations

The 10 proposed methods for measuring heart rate are all capable of measuring the heart to some extent as evident from Table 5.6 on page 34. However 3 of the methods M3 (Green), M6 (Hue) and M7 (Saturation) all have greater variances than the rest of the methods which is evident from Table 5.7 on page 35. The rest of the

methods employ some sort of linear combination of the colour components to capture the underlying PPG signal. Thus performance is improved when considering linear combination of the colour channels instead of the individual components. The true underlying mixture is perhaps not linear but given the relatively short time windows used we expect a linear combination to be reasonable approximation. This is supported by the low standard deviations (< 2 BPM) of all the methods using a linear model.

Best linear combination optimisations

The method M10 "Best linear combinations" is currently calculated by evaluating fourier transform of 250 linear combinations of RGB and normalising with the standard deviation of the linear combinations. Hopefully this can be changed to finding the linear combination for each value of the Fourier transform which maximises the current value, after normalisation with standard deviation. This should be possible since we can calculate the standard deviation of any linear combination of RGB using standard rules of calculation for variances and the Fourier transform can easily be calculated as a linear combination.

ICA assumptions

In order to use the ICA we assume that we have exactly 3 sources, and that the RGB captured by the camera is a linear mixture of these 3. In the majority of our experiments the 3 recovered sources are like Figure 4.3 on page 20, where one component corresponds to the PPG signal, one to light changes in the scene and the last one roughly being Gaussian noise. This violates the assumption that the source signals are non-Gaussian, but our ICA algorithm is evidently able to gather all Gaussianity in one component.

Although we assume that there are exactly 3 underlying sources the true number could possibly be higher. This is an issue if the aim is to completely separate the PPG signal from the rest of the signals. However, in the context of extracting only the heart rate from the PPG signal it is sufficient that it is separated enough so that the PPG signal is the most dominant periodic signal, which the "Success rates ICA" in Table 5.6 on page 34 for the ICA methods (M1, M4 and M8) shows is in fact the case most of the time.

Possible issues with fixed weights

In Tables 5.4 and 5.6 on page 30 and on page 34 it can be seen that fixed weighting provides slightly better performance compared with the ICA counterpart in all colour spaces. We believe it performs better than the ICA because of the following reasons. The fixed weights learned by averaging many ICA components describe the specific camera and lighting conditions of the videos nicely. There are also no issues with PPG signal identification, since there is only one signal to choose from. There might

also exist cases the ICA has issues with separating the components due to very noisy conditions. This would be fixed by fixed weights as they would average these issues out.

We do however think that these weights are very specific to the lighting conditions of the recorded videos and possibly also to the skin colour of the subjects, and further experiments need to be done to determine how dependent these weights are on camera and lighting conditions. A solution to this may be to have a large library of fixed weights to choose from, which could be selected by the level of ambient light and current white balance settings of the camera.

Colour space comparisons

We introduced different colour spaces with the hope that the nonlinear transforms would separate the signals in ways the ICA would not be able to. We would also expect to find the strongest signal in changes of colour and not in brightness, which is why we chose colour spaces where most components describe colour. It was also our hope that because the components in HSV and Lab are less corelated, convergence of the ICA would be more likely. Our pairwise Brown-Forsythe tests in Table 5.7 on page 35 show no significant differences between the variances of the errors for either of RGB, HSV or Lab with ICA, since the lowest p-value for these pairs are 0.82. This leads us to the conclusion that the nonlinearities introduced by the change in colour space is not significant when it comes to finding the PPG signal.

From the fixed weights in Equation (5.1) on page 34, it can be seen that there still is some PPG signal in the component of the colour spaces meant to describe brightness. For example we have a weight of $-.217$ for the lightness component in the Lab colour space, and for HSV the value component weighs almost as much as the hue. This weight for hue is also surprisingly low, leaving almost all of the signal in the saturation channel, but this may be very dependent on the camera and lighting of the video.

The fixed weight of the RGB colour space has the majority of its weight in the green channel, which agrees nicely with the absorbtion spectra described in Section 2.2 on page 4. Because a large part of the PPG signal is present in the green channel, the method only using the green channel also performs acceptably. It does however work with less information and has worse performance as evident in Table 5.6 on page 34. The variance of the errors is also significantly different between the only green and RGB with ICA at $\alpha = 0.01$ as seen in Table 5.7 on page 35.

Pixel noise

Any time a measurement is performed, there is bound to be some error committed. This is especially true for pixels in cheaper video cameras and under poor lighting conditions. In our method we average over a large amount of pixels, thus reducing the influence of all this noise by a large factor. We also know that the signal is best in the green channel. This channel is less prone to noise due to most image sensors

using a Bayer filter with twice as many green photosensitive areas, compared to red and blue.

FPS choice

We have used 30 FPS for all of our projects measurements. According to the Nyquist-Shannon sampling theorem (Shannon 1949) we need at least 7 FPS in order to be able to recognize heart rates up to 210 BPM. Our 30 FPS lies comfortably above this, which reduces the importance of each individual observation thus increasing the robustness with regard to noise.

Changes in heart rate

The problem illustrated in Figure 5.3 on page 29 can arise any time the heart rate changes during the recording of the video and not only in long videos. One could argue that the most correct heart rate to assign such a video would be the average heart rate observed over the duration of the video. There is no simple way to handle this issue while using the Fourier transform . However by measuring directly on the times between peaks in the video and taking the average, one would be able to alleviate this issue mostly. This method has been applied by Poh et al. (2011).

6.5 Short time heart rate estimation

Figure 5.8 on page 37 shows several measures of the quality of our heart rate methods as a function of the window length used to calculate the heart rate. The general tendency is that windows longer 10 seconds of length do not provide better performance than their shorter counterpart. For a possible mobile app for measuring the heart rate, we would thus suggest using a window length of 10 seconds for measuring the heart rate. It also seems like a reasonable amount of time to make a person wait for a measurement.

The bias introduced by the Fourier transform visualized in Figure 5.9 on page 38 shows that it is not recommendable using windows of length shorter than 10 seconds as the bias significantly increases for under 10 seconds. This is in line with the prior suggestion of using a window length of 10 seconds, but not any shorter.

6.6 Heart rate intensity map

The HRIMs shown in Figures 5.14 to 5.16 on pages 40–42 shows the distribution of the amplitude of the heart rate in the face and hand. It is not unreasonable to think that these are in some way related to the underlying microvascular structure, but we do not have the anatomical knowledge to delve further into this. However it should be noted that the amount of light, the angle of the light and shadows all influence the actual appearance of the HRIM.

However in the context of non-contact photoplethysmography it is an interesting result. It can be utilised, as we have done for our blood pressure investigation, to get a better PPG signal. The HRIMs of the faces shows that the amplitude of the heart rate is high in the upper cheek area. This could perhaps be combined with the Viola-Jones eye detection previously discussed to ROIs under the eyes instead of using the entire face.

6.7 Blood pressure estimation

The correlation plots in Figures 5.18 and 5.19 on page 44 shows the systolic blood pressure measured after the recording of the video plotted against the PTTD measured using cross correlation and Fourier transform respectively. The readings after the video recording are chosen since the we expect the blood pressure to be more stable after the subjects have lain down for the video recording.

The two plots are very similar. If the sample with the largest PTTD (much larger than the rest) is excluded, both plots have correlation coefficients around -0.8 which with p-values 0.011 and 0.09 respectively. This suggests a strong correlation between the systolic blood pressure and the PTTD. However we have no satisfactory explanation as to why we should exclude that sample, and thus the correlation coefficients are much smaller and not very significant. Because of this we can not reject the null hypothesis of zero correlation. A larger dataset is necessary to determine if there is correlation or no correlation between the PTTD and the systolic blood pressure.

From the results in Table 5.20 on page 43 it is seen that a linear relationship between the PTTD and the diastolic blood pressure is doubtful, but due to the small dataset we will not completely rule out some correlation.

6.8 Real time applications

To demonstrate the feasibility of our method working in real-time, we have also implemented a very simple real-time version of our method. It only uses the green channel, a fixed ROI and linear resampling of the signal. It can be tried out at the following homepage:
`www.student.dtu.dk/~s113258/blood/`

6.9 Rolling shutter

For our PTTD measurements we are measuring very small differences in time. Using a video camera for this leaves us open to artefacts introduced by how the images are read off the sensor. A very common method of doing this is by reading one line at a time, which introduces a time difference between the top and bottom of the image. If our videos had been oriented with the hand placed on lines above the face, this would introduce a larger error. In our videos we have the hand and face placed

approximately the same place vertically, which means that this error is minimal. Even if we were in a worst case scenario, we are working with 120 FPS video, which means that the entire sensor is read in 8.3 ms. This means that the upper limit on the error introduced by rolling shutter is 8.3 ms, which is an acceptable error to have, since our numbers stretch over quite larger ranges.

CHAPTER 7

Conclusion

We have succesfully recreated the method described in Poh et al. (2010), and achieved slightly better but comparable results. We have also examined the use of different colour spaces and found no significant differences in the results obtained using them.

We have created an alternative to ICA which achieved even better performance on our dataset, and is feasible to implement in real-time. It works by trying a large number of combinations of RGB.

Furthermore, we have identified that the a window length of 10 seconds is reasonable for real time measurements.

We have also calculated several HRIMs of faces and hands, which can be utilised to find the heart rate more accurately and under harder conditions.

Finally we have looked at the corelation between PTTD and systolic blood pressure, but we were unable to reach a definitive conclusion without looking at more test subjects.

7.1 Future work

Several topics for future work have been mentioned in the discussion. One of them is improving the face detection location accuracy. Viola-Jones eye detection could possibly increase the position accuracy and compensate for small rotations in the face. An alternative is using the KLT feature tracker, which can easily correct for rotations as well as increasing the position accuracy.

Another topic for future work is the PPG signal identification, which can possibly be improved by training a classifier of different measures and using that to identify the correct ICA component.

Finally the relationship between the blood pressure and the PTTD could be explored further with a larger dataset.

Bibliography

Bowmaker, J. K. and H. J. Dartnall (1980). "Visual pigments of rods and cones in a human retina". In: *The Journal of physiology* 298, pp. 501–511.

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools*.

Brown, Morton B. and Alan B. Forsythe (1974). "Robust Tests for the Equality of Variances". English. In: *Journal of the American Statistical Association* 69.346, pp. 364–367.

Cardoso, Jean-François (2013). *Implementaion of JADE in Matlab*. URL: `http://perso.telecom-paristech.fr/~cardoso/Algo/Jade/jadeR.m` (visited on Mar. 21, 2014).

Cardoso, Jean-François and Antoine Souloumiac (1993). "Blind beamforming for non Gaussian signals". In: *IEE Proceedings-F* 140.6, pp. 362–370.

Cui, Weijia, Lee E. Ostrander, and Bok Y. Lee (1990). "In Vivo Reflectance of Blood and Tissue as a Function of Light Wavelength". In: *IEEE Transactions on Biomedical Engineering* 37.6, pp. 632–639.

Douniama, Christian and Robert Couronné (2007). "Blood Pressure Estimation based on Pulse Transit Time and Compensation of Vertical Position". In: *3rd Russian-Bavarian Conference on Bio-Medical Engineering*, pp. 38–41.

Edwards, G. J., C. J. Taylor, and T. F. Cootes (1998). "Interpreting Face Images Using Active Appearance Models". In: *Proceedings of the 3rd. International Conference on Face & Gesture Recognition*. FG '98. Washington, DC, USA: IEEE Computer Society, pp. 300–.

Hertzman, A. B. and C. R. Spealman (1937). "Observations on the finger volume pulse recorded photoelectrically". In: *Am. J. Physiol. Meas.* (119), pp. 334–335.

Liao, Shengcai, Xiangxin Zhu, Zhen Lei, Lun Zhang, and Stan Z. Li (2007). "Learning Multi-scale Block Local Binary Patterns for Face Recognition". In: *Proceedings of the 2007 International Conference on Advances in Biometrics*. ICB'07. Seoul, Korea: Springer-Verlag, pp. 828–837.

Nitzan, M., B. Khanokh, and Y. Slovik (2002). "The difference in pulse transit time to the toe and finger measured by photoplethysmography". In: *Physiol Meas* 23.1, pp. 85–93.

Philips (2013). *Vital Signs Camera*. URL: `www.vitalsignscamera.com` (visited on May 5, 2014).

Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard (2010). "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation". In: *Optics Express* 18.10.

Poh, Ming-Zher, Daniel J. McDuff, and Rosalind W. Picard (2011). "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam". In: *IEEE Transactions on Biomedical Engineering* 58.1.

Richert, Micah (2009). *mmread*. URL: `http://www.mathworks.com/matlabcentral/fileexchange/8028-mmread`.

Scanlon, Valerie C. and Tina Sanders (2007). *Essentials of Anatomy and Physiology*. F.A. Davis Company.

Shannon, C. E. (1949). "Communication in the Presence of Noise". In: *Proceedings of the IRE* 37.1, pp. 10–21.

Tanaka, H., K. D. Monahan, and D. R. Seals (2001). "Age-predicted maximal heart rate revisited". In: *J. Am. Coll. Cardiol.* 37.1, pp. 153–156.

Tarvainen, Mika P., Perttu O. Ranta-aho, and Pasi A. Karjalainen (2002). "An Advanced Detrending Method With Application to HRV Analysis". In: *IEEE Transactions on Biomedical Engineering* 49.2, pp. 172–175.

Tomasi, Carlo and Takeo Kanade (1991). "Detection and Tracking of Point Features". In: *Carnegie Mellon University Technical Report* 132.

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

Viola, Paul and Michael J. Jones (2004). "Robust Real-Time Face Detection". In: *Int. J. Comput. Vision* 57.2, pp. 137–154.

ViTrox (2013). *What's my Heart Rate*. URL: `www.whatsmyheartrate.com` (visited on Apr. 4, 2014).

Wu, Hao-Yu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman (2012). "Eulerian Video Magnification for Revealing Subtle Changes in the World". In: *ACM Transactions on Graphics (TOG)* 31.4.

Zijlstra, W.G., A. Buursma, and O.W. van Assendelft (2000). *Visible and Near Infrared Absorption Spectra of Human and Animal Haemoglobin: Determination and Application*. Taylor & Francis.