



Remote Heart Rate Extraction from Near Infrared Videos

An Approach to Heart Rate Measurements for the Smart Eye Head Tracking System.

Master's thesis in Biomedical Engineering and Systems, Control and Mechatronics

ANNA LINDQVIST MARKUS LINDELÖW

MASTER'S THESIS 2016:EX031

Remote Heart Rate Extraction from Near Infrared Videos

An Approach to Heart Rate Measurements for the Smart Eye Head Tracking System

ANNA LINDQVIST MARKUS LINDELÖW



Department of Signals and Systems Division of Signal Processing and Biomedical Engineering CHALMERS UNIVERSITY OF TECHNOLOGY Gothenburg, Sweden 2016 Remote Heart Rate Extraction from Near Infrared Videos An Approach to Heart Rate Measurements for the Smart Eye Head Tracking System ANNA LINDQVIST, MARKUS LINDELÖW

© ANNA LINDQVIST, MARKUS LINDELÖW, 2016.

Supervisor: Björn Lindahl, Smart Eye AB Academic Supervisor: Lars Hammarstrand and Karl Granström, Department of Signals and Systems Examiner: Tomas McKelvey, Department of Signals and Systems

Master's Thesis 2016:EX031 Department of Signals and Systems Division of Signal Processing and Biomedical Engineering Signal Processing Group Chalmers University of Technology SE-412 96 Gothenburg Telephone +46 31 772 1000

Typeset in LATEX Printed by Chalmers Reproservice Gothenburg, Sweden 2016 Remote Heart Rate Extraction from Near Infrared Videos An Approach to Heart Rate Measurements for the Smart Eye Head Tracking System ANNA LINDQVIST, MARKUS LINDELÖW Department of Signals and Systems Chalmers University of Technology

Abstract

Recent research has shown that the heart rate of a person can be measured from face videos by analysing the small changes in intensity of the pixels in the face. Being able to remotely measure the heart rate from videos can be very useful in situations where the physiological state of the filmed person is of interest; for example when monitoring drivers in vehicles, knowing the heart rate of the subject will be a strong indicator for the emotions of the person. This information could then be used to adjust the collision avoidance system and entertainment system accordingly.

The company Smart Eye develops near infrared cameras with advanced technology for real-time gaze, eyelid and head tracking. One of the main applications for the cameras of Smart Eye is driver monitoring and hence, knowing the heart rate of the driver is of great interest. In this thesis, a new algorithm for remote heart rate extraction from near infrared videos for the camera system of Smart Eye is presented.

The developed algorithm consists of several signal processing steps such as various filters and transforms. The periodicity of the heart rate is found with Welch's method, which is a modified periodogram with a high noise robustness. Also, a Kalman smoother and a Bayes filter is applied to remove outliers. The head tracking data from the Smart Eye system is used for motion robustness, where the measurement region of the face is continuously updated for each frame by looking at the position of the head tracking data points.

The algorithm showed very good results for stationary subjects with a mean error rate around 1% when compared to a reference signal obtained with an electrocardiography. For moving subjects, the algorithm was robust for motion of approximately 6 cm/s in velocity for scaling and translational movement.

It was also shown that the available head tracking data from the Smart Eye system contained information of the heart rate. A modified version of the algorithm for heart rate extraction was developed that instead of a video had head tracking data as an input. The algorithm resulted in a mean error rate of 11.25% for stationary subjects when compared to an electrocardiography.

Acknowledgements

Firstly, we would like to thank our supervisor Björn Lindahl for his time, commitment and support. For always being there with good inputs and we have always returned to our work with a warm feeling inside after every meeting.

This thesis would never had been made if it wasn't for Per Sörner, who introduced us to Smart Eye, its technology and this master thesis. It is amazing to observe how you humbly share your knowledge, from abstract theoretical concepts and how they concretely effect our work to specific differences between implementations and their effects. We are deeply grateful for having you around.

To our academic supervisors at Chalmers, Lars Hammarstrand and Karl Granström, thank you for your dedication and support. For caring about that this thesis should have the highest quality as well as organising peer-reviews and discussion groups. We have a lot to thank you for.

Many thanks to all the people at Smart Eye for the warm welcome, showing interest in our project and for sharing your knowldege. A special thanks to Oscar Järrebring who has been truley helpful throughout the project with set-up of the system and various technical challenges. In addition, we want to express our gratitude to the persons that volunteered as test persons for the study.

Lastly we would like to thank our nearest and dearest who have supported us during this period of time.

Markus Lindelöw and Anna Lindqvist, Gothenburg, June 2016

Contents

Li	st of	Figures	xi
Li	st of	Tables	xiii
A	crony	yms	xv
1	Intr	coduction	1
	1.1	Aim	2
		1.1.1 Research Questions	2
	1.2	Limitations	3
	1.3	Thesis Outline	3
2	The	eory	5
	2.1	Physiology of the Heart Rate	5
	2.2	Photoplethysmography	5
		2.2.1 Remote Photoplethysmography	6
	2.3	Ballistocardiography	6
		2.3.1 Remote Ballistocardiography	6
	2.4	Measuring Heart Rate from Near Infrared Videos	7
	2.5	The Technology of Smart Eye	8
3	The	eory of Methodology	11
	3.1	Welch's Method	11
	3.2	Bayes Filter	12
	3.3	Rauch-Tung-Striebel Smoother	13
		3.3.1 Forward Pass - Kalman Filter	13
		3.3.2 Backward Recursion Pass	15
	3.4	Accuracy Measurements	15
	3.5	Error Matrix	16
4	Met	thods	19
	4.1	Equipment	19
	4.2	Data Collection	20
	4.3	Heart Rate from Video Algorithm	20
		4.3.1 Region of Interest	21
		4.3.2 Pre-Filtering	$\frac{1}{22}$
		4.3.3 Heart Rate Calculations	22

 4.3.4 Post-Filtering	 23 23 23 25 27 							
 4.3.5 Motion Robustness	23 23 25 27							
 4.3.6 A Confidence Measure	23 25 27							
 4.4 Heart Rate from Head Tracking Data Algorithm	25 27							
 5 Results 5.1 Heart Rate from Video Algorithm	27							
5.1 Heart Rate from Video Algorithm								
5.1.1 Pro Filtoring	27							
0.1.1 1 Ie-r means 	27							
5.1.2 Welch's Method	27							
5.1.3 Bayes Filter	29							
5.1.4 Evaluation of the Time Window for Heart Rate Calculations .	29							
5.1.5 Post-Filtering	30							
5.1.6 Evaluation of Regions of Interest	30							
5.1.7 Stationary Subjects	31							
5.1.8 Motion Robustness	33							
5.1.9 Evaluation of the Confidence Measure	34							
5.2 Heart Rate from Head Tracking Data Algorithm	35							
5.3 The Effect of the Ballistocardiographic Signal	36							
6 Discussion	37							
6.1 Discussion of Heart Rate from Video Algorithm	37							
6.1.1 The Effect of the Time Window of Heart Rate Calculations	37							
6.1.2 Regions of Interest	38							
6.1.3 Motion Robustness	38							
6.1.4 Confidence Measure	39							
6.2 Heart Rate from Head Tracking Data Algorithm	40							
6.3 The Effect of the Ballistocardiographic Signal	41							
7 Conclusion	43							
8 Future Work	45							
Bibliography 47								

List of Figures

2.1	Light absorption for oxyhemoglobin and de-oxyhemoglobin at different wavelengths. The gram molecular weight of hemoglobin is approximated to 64500 g/mole and the hemoglobin concentration to 150 g/L .	8
2.2	Screen shot of the Smart Eye Pro software. Three cameras are shown in the upper part of the program where the gaze direction and the head pose are visualized. In the lower left window a 3D model of the environment and the tracking subject can be seen. The lower middle image shows the current estimated values such as head pose, gaze direction and eye positions etc.	9
3.1	Power spectral density estimate of a sinusoid signal with noise using a) standard periodogram b) Welch's method	11
3.2	The recursive process of the Bayes Filter. A prediction is made from the last estimate. This prediction is combined with a measurement to get the new estimate	12
5.5	FIR filter for generated example data when the true state makes a sudden change of slope.	13
4.1 4.2 4.3	Overview of the equipment set-up during data collection Data processing pipeline for the HR from video algorithm The 5 different evaluated ROIs; (a) static ROIs (b) the grid for the	19 20
4.4	dynamic ROI where three parameters were evaluated for each square. Data processing pipeline for the HR from head tracking data algorithm.	$\begin{array}{c} 21 \\ 25 \end{array}$
5.1	A 10 seconds segment of the obtained a) raw signal b) detrended signal c) moving average filtered signal d) band-pass filtered signal.	28
5.2	Power spectral density for one stationary subject. a) power density over frequency at the specific time of 40 s b) power spectral density in dB over time where (a) can be found as a vertical cross-section at	
5.3	40 s	28
	always estimates the most powerful peak as the HR	29

5.4	The oscillations from Welch's Method is removed with an RTS-smoother	
	when estimating the HR for one stationary subject.	30
5.5	The dynamic ROI where the green areas represent the obtained ROI	0.1
-	for one example subject.	31
5.6	Comparison between the ECG signal and the obtained signal from a	~ ~
	video of a stationary subject.	32
5.7	The HR from ECG and the estimated HR from a video of a stationary	
	subject.	32
5.8	The absolute error of the estimated HR in bpm for the stationary	
	data-set.	33
5.9	MAE for different velocities from the 16 measurements of the mo-	
	tion robustness data-set for (a) translational movement (b) scaling	
	movement	33
5.10	Scatter plot of absolute error for each estimated HR from the video	
	algorithm against its confidence for the stationary data-set. A con-	
	fidence threshold of 0.86 can be identified where the estimated HR	
	that has a higher confidence value can be classified as reliable	34
5.11	The absolute error against time for one stationary subject where	
	each estimated HR is classified as reliable or unreliable depending	
	on whether they fulfil the confidence threshold of 0.86	35
5.12	The absolute error in bpm for HR calculations from head tracking	
	data, derived from the stationary data-set.	36
5.13	Comparison of HR for the video of the BCG-PPG dataset calculated	
	by different methods; ECG, PPG signal from skin region, PPG signal	
	from tape region and from movement of head tracking data points.	36

List of Tables

3.1	Overview of an error matrix	16
5.1	Error measurement rates for different time windows for HR calcula- tions, derived from the stationary data-set.	30
5.2	Error measurement rates for different regions of interest, derived from	
	the stationary data-set	31
5.3	Error matrix of reliable/unreliable classification	34
5.4	Error measurement rates for HR measurements from head tracking	
	data, derived from the stationary data-set	35

Acronyms

Acronym	Expansion
BCG	Ballistocardiography
bpm	beats per minute
ECG	Electrocardiography
FIR	Finite impulse response
HR	Heart rate
M_{eRate}	Mean error rate of percentage
MAE	Mean absolute error
NIR	Near infrared
NPV	Negative predictive value
PCA	Principal component analysis
PPG	Photoplethysmography
PPV	Positive predictive value
RMSE	Root mean square error
ROI	Region of interest
rPPG	remote photoplethysmography
RTS	Rauch-Tung-Striebel
SD	Standard deviation

1 Introduction

Recent research has shown that the heart rate of a person can be measured from facial regions of persons filmed with a camera system [1]. Conventional methods for measuring heart rate involve electronic or optical sensors that have to be connected to the body. Being able to remotely measure the heart rate from videos can be very useful in situations where the physiological state of the filmed person is of interest. For example in a vehicle when the driver is monitored, knowing the heart rate of the object is a strong indicator for the emotions and the stress level of the person. This information could then be used to adjust the collision avoidance system and entertainment system accordingly.

The company Smart Eye develops cameras with advanced technology for real-time gaze, eyelid and head tracking. Two of the main market segments are the automotive industry and within research. When monitoring a driver, the monitor system should not be dependent on light sources of the visible light spectrum since they could disturb the driver at night. Therefore, the systems of Smart Eye use frequencies in the near infrared light spectrum which is invisible for the driver. The camera systems of Smart Eye use a wavelength of either 850 nm or 940 nm, which means that the systems are monochrome. One of the outputs from the software of Smart Eye is the position of facial feature points such as eyes, mouth and nostrils. [2]

Previous methods for extracting heart rate from videos have mostly covered RGBvideos [3, 4] and thermal imaging [5]. Near infrared cameras have been used to make the measurements more robust but not as stand-alone monochrome measurement systems [6, 7, 8]. Several studies have used the technology of head tracking to compensate for motions of the subject [1, 6]. By making sure that the studied facial region always stays the same despite movement, a motion robust method can be developed.

The majority of the previous works have implemented methods that measure the variation in light absorption due to blood flow in the face related to the heart rate, referred to as photoplethysmography, [1, 3, 4, 6, 9]. The absorption variation is much stronger in the color spectrum than in the near infrared spectrum and hence the effect is not seen as clearly in a near infrared camera system [10]. However, there are other works that have extracted the heart rate by measuring the small head motions caused by the pulsatile blood flow to the head, referred to as ballistocardiography [11]. This method is assumed to be indifferent to whether the video system is of the color spectrum or the near infrared spectrum.

As the heart rate estimate may not always be accurate, a confidence measure can be developed to indicate how much one can trust the estimated heart rate value. The confidence measure should be able to classify the reliability of the heart rate estimate by analysing the signal properties of the measured input signal.

The purpose of this project is to find an approach for remote heart rate extraction that also work for the near infrared system of Smart Eye. As stated, the previous methods have used RGB-videos and hence, the method for this study must be adapted for near infrared videos, especially the fact that a monochrome system is used and that the light absorption variation is much lower than for the color light spectrum. Due to this, the possibility to measure the heart rate from head movements is also investigated.

1.1 Aim

The aim of the master thesis is to create an algorithm that extracts the heart rate of a subject from the near infrared system of Smart Eye. Firstly, an algorithm is created to work for videos with standard environments, meaning no changes in illumination and no movements from the subjects. This is done by finding the optimal parameters for the algorithm, such as what region of the face to extract the signal from, filter parameters and what time window to use for the heart rate calculations. The parameters are evaluated by comparing different accuracy measurements. A confidence measure for the heart rate estimate is also developed.

Secondly, the algorithm is developed to work for moving subjects with the implementation of a moving region for heart rate extraction with the help of the head tracking data of the Smart Eye system.

Lastly, the possibility to extract the heart rate by analysis of small head movements due to the blood flow to the head is investigated. For this purpose, a modified algorithm is developed where the head tracking data of the Smart Eye system are used as input. The movement of the feature points in the face, given by the Smart Eye system, is measured.

1.1.1 Research Questions

The following questions are addressed in the report:

- Can a signal for the heart rate be extracted from near infrared videos, when the subject is still and the lighting conditions are controlled?
- What are the optimal parameters for the developed heart rate algorithm?
- Can a signal for the heart rate be extracted from near infrared videos when the subject is moving, with support from the head tracking data of Smart Eye?

- Can a confidence measure be created for the algorithm that can classify the reliability of the estimated heart rate?
- Can a signal for the heart rate be extracted from a subject by analysing the head tracking data of Smart Eye?

1.2 Limitations

The goal of the study is to extract heart rate from the Smart Eye system for a data-set of a limited size and not to create a fully functional application. The data-set is recorded in a controlled environment with a camera setup used regularly by Smart Eye. The developed algorithms are developed independently as stand-alone programs in MATLAB without any software or code developed by Smart Eye. In addition, all code is strictly developed by the authors or contained as built-in functions in MATLAB unless stated otherwise.

The heart rate algorithms are developed and evaluated using the camera system of Smart Eye and hence, only for the properties of this specific system, such as sampling rate, wavelength of cameras. An assumption is made regarding the availability of accurate data. It is assumed, regarding the videos, that the faces of the subjects are visible during the entire time of recordings. In addition, it is assumed that the head tracking output from the software of Smart Eye contains the relevant head tracking data feature points for the entire time of the recordings.

Motion robustness is only investigated for the video based algorithm and for a limited set of motions, independent translational and scaling movements. For a full investigation of motion robustness, rotational movements as well as mixed movements would have to be investigated.

The accuracy of the developed algorithms is not compared to the accuracy of reference works. This is due to that the purpose of the study is to develop an algorithm that works for the system of Smart Eye which is of a very different set-up than other systems.

1.3 Thesis Outline

Next chapter, Chapter 2, explains the theory behind how the heart rate can be measured, both in a conventional and in a non-contact manner. The chapter also introduces the technology of Smart Eye. In Chapter 3, theory of methodology, the theory of the applied filters and the transforms needed for heart rate extraction is presented. Chapter 4 explains in detail the methodology for the work, how various filters and transforms were used and with which parameters, how the evaluation of the algorithms was constructed and how the experiments were conducted. The results of the study are presented in Chapter 5, for both algorithms. The results are followed up with a discussion in Chapter 6. The conclusion of the study is presented in Chapter 7 and, lastly, in Chapter 8 some recommendations for future work are given.

Theory

2.1 Physiology of the Heart Rate

The heart rate (HR) is a measure of how often the heart contracts in order to supply the rest of the body with blood, normally measured with the unit beats per minute (bpm) but can also be expressed in Hertz (Hz). A healthy human adult has a HR between 45 - 240 bpm (0.75 - 4 Hz) [1]. There are many factors influencing the HR of a person, both modifiable (e.g. physical activity, mental stress, smoking) and non-modifiable (e.g. age, gender) [12]. The HR can change rapidly from activities such as postural changes [13] and also from uncontrolled events such as cardiac arrhythmia and sudden cardiac arrest [14, 15].

The most accurate method to measure the HR is electrocardiography (ECG), where the electrical impulses generated by the cardiac muscle fibers is measured with the help of three electrodes on the skin. It is these electrical impulses that cause the heart to contract, the blood to be dispersed out into the body and hence creates a pulse [16]. There are also other simpler methods to measure the HR, such as photoplethysmography (described in more detail below), piezoelectric transducer or Doppler ultrasound [12]. Another method that is not used in clinically settings anymore is ballistocardiography, which is described in detail in Section 2.3 [11].

2.2 Photoplethysmography

Photoplethysmography (PPG) is a non-invasive optical technique for measuring vital rates, such as HR and respiration rate, by looking at volume changes in microvascular tissue. Due to the difference in light absorption of blood and surrounding tissue, there will be small changes in intensity values when investigating a part of the skin, normally the finger, ear or toe. The changes result in a periodic waveform related to the HR and one more slowly varying signal related to the respiration rate [17]. PPG was first described in the 1930s by Hertzman and Spealman and the only equipment needed for PPG is a light source to illuminate the skin and a photodetector to measure the variation in light absorption [18]. The system operates at a red or near infrared (NIR) wavelength [17].

2.2.1 Remote Photoplethysmography

In recent years, the possibility to extract the HR with the help of a remote photodetector has been established, so called remote PPG (rPPG) [1, 3, 4, 6, 9]. The new technique offers a measurement system for the HR that does not need to have contact with the studied object, a valuable feature for both medical and surveillance purposes [6]. In 2005, Humphreys et. al [9] showed that one can extract the HR with the help of rPPG when studying a finger at a 40 cm distance. They used a NIR light source with a wavelength of 800 nm. However, the method could not handle any motions and the scenario was not much more flexible than conventional PPG [9].

Since the study of Humphreys, many methods have been developed for rPPG of the face with RGB-cameras as photodetectors [3, 4]. The use of three color channels with multiple wavelengths gives the methods the possibility to be robust to motion of the subject since principal component analysis (PCA) can be applied [6]. PCA separates the HR signal from the motion noise by expressing them as a linear combination. Though, PCA uses the assumption that the motion noise is not of a periodic character [6].

One study worth highlighting is of Li et al. [1], since a single color channel is used for rPPG, namely the green channel. Even though the study uses RGB cameras, the fact that they only use one of the color channels makes the set-up similar to the camera system of Smart Eye due to the monochrome character. In the method of Li et al., PCA is not performed but instead an advanced filtering procedure is applied.

2.3 Ballistocardiography

To evaluate cardiac parameters by measuring the periodic motions created from the ejection of blood into the great vessels with every heart contraction is referred to as ballistocardiography (BCG). The technique was first described in the 1930s. There are several ways to measure BCG motions, either by placing the test person on a low-friction platform and measure the motion of the platform, or by placing the test person on a scale and measure the recoil of the person's body on the scale [19].

2.3.1 Remote Ballistocardiography

The conventional BCG methods all require large stationary equipment, however there are some group of researchers that have obtained a BCG signal from measuring head movements only. When blood flows periodically to the head through the abdominal aorta and the carotid arteries, small oscillatory motions of the head are generated [11]. He et al. have developed a wearable wireless BCG device where the BCG signal is obtained through measuring the head movements with an accelerometer worn at the ear similar to a hearing aid device. According to the study, the BCG signal is the strongest in the vertical direction and is in the range of 10 mG in acceleration [19]. Another research group, Balakrishnan et al., have developed an approach to measure the BCG signal remotely, similar to rPPG measurements. The BCG signal is obtained through tracking the movement of facial feature points in a video of the head. Balakrishnan et al. states that the method is complementary to the rPPG methods and similarly PCA is used for data processing of the feature point motion signals [11]. The BCG approach has several advantages compared to PPG when measuring HR, since the method is not dependent on the visibility of skin. However, according to Irani et al., the BCG method is very sensitive to head movements and facial expressions [20].

The small head motions that give rise to the BCG signal can assumingly also be seen as intensity changes in a video. When the head moves in relation to a fix point light source, the intensity of the skin changes, resulting in a higher intensity level when the head is closer to the light source and vice versa. Hence, the variation in intensity of the skin related to the HR is both due to absorption changes (PPG signal) but also due to head motion (BCG signal). However, it is difficult to separate the two signals and it is therefore also difficult to know the contribution of the two signals to the intensity changes.

2.4 Measuring Heart Rate from Near Infrared Videos

One way to measure the HR from NIR videos is to use rPPG. The absorbed light of a human in NIR consists of two components: one static absorption constant from muscles, fat, bones and venous blood (de-oxyhemoglobin) and one varying component from arterial blood (oxyhemoglobin). As the heart beats, the concentration of oxyhemoglobin will fluctuate in the body and effect the light absorption subsequently [21]. The red color of the blood is caused by the fact that absorption coefficient is low for the red wavelengths (620 - 700 nm) and high for the other wavelengths of visible light, meaning that only the red light is reflected. The light absorption coefficient in NIR (700 - 1400 nm) is very low: $5.7 \,\mathrm{cm}^{-1}$ in 850 nm. The wavelength of 540 nm (green light), which Li et al. have used, has around a 50 times higher absorption coefficient than what NIR has [10].

Another difference between RGB-videos and NIR-videos are the number of wavelengths captured. An RGB-camera records three wavelengths simultaneously while a monochrome NIR-camera only records one. The benefit of having more than one frequency is to use the other wavelengths that have a lower oxyhemoglobin absorption as reference to cancel out ambient light and noise as Verkruysse et al [3].

When it comes to measuring the HR via BCG, it is assumed that the method is indifferent to what light source is used. Hence, using BCG for a NIR camera system would produce similar results as the results of the study of Balakrishnan et al. where a RGB-camera system was used [11].



Figure 2.1: Light absorption for oxyhemoglobin and de-oxyhemoglobin at different wavelengths. The gram molecular weight of hemoglobin is approximated to 64500 g/mole and the hemoglobin concentration to 150 g/L.

2.5 The Technology of Smart Eye

Smart Eye was founded in 1999 and released their first non-intrusive eye tracking system in 2001. In 2005, they released an anti-sleep system which monitor drivers if they fall asleep. Today Smart Eye have a wide range of solutions targeting auto-motive, aviation and aerospace research and manufacturing industries.[2]

Smart Eye also provides a software named *Smart Eye Pro* which comes with a state of the art non-intrusive head tracking system which automatically detects the head and tracks its 3D pose in six dimensions of freedom. Over time, the system learns increasingly more about the face of the subject and the tracking is improved continuously. The head tracker is very robust and handles for example partially cloaked faces [22]. As an output from the Smart Eye Pro, one can obtain the 3D position of several feature points of the face, such as the eye corners, the nose tip and the mouth position. With the information, it is possible to study the head movement of the studied subject. A printscreen of the software can be seen in Figure 2.2.



Figure 2.2: Screen shot of the Smart Eye Pro software. Three cameras are shown in the upper part of the program where the gaze direction and the head pose are visualized. In the lower left window a 3D model of the environment and the tracking subject can be seen. The lower middle image shows the current estimated values such as head pose, gaze direction and eye positions etc.

2. Theory

3

Theory of Methodology

Several signal processing techniques such as filters and transforms are used in the developed HR extracting algorithms. This section explains the purpose and the theory behind different techniques. For Section 3.2 and 3.3, the information is taken from *Bayesian Filtering and Smoothing* by Särkkä (2013) unless otherwise stated [23].

3.1 Welch's Method

In order to determine the HR from a pulsatile signal, the period of the signal is estimated. To do this, the frequency content of the signal is investigated by looking at the power spectral density. One method for this is Welch's method which is a modified version of the standard periodogram. While it has lower frequency resolution, it benefits from noise reduction as seen in Figure 3.1. [24]



Figure 3.1: Power spectral density estimate of a sinusoid signal with noise using a) standard periodogram b) Welch's method.

The method starts by dividing the signal into n number of segments with d number of samples overlap between each segment. Often d is expressed in percentage of the segment length with 50% as the most common value. Each segment is then windowed with a Hamming window before calculating the discrete Fourier transform of the segment. The Hamming window is a type of window which is designed to minimise the sidelobe level. The last step averages the square magnitudes of the transforms and produces the final modified periodogram which shows the power spectral density estimate of the signal. [24]

3.2 Bayes Filter

When estimating the HR, physiological behaviour of the heart can be used to improve the accuracy of the estimated signal and reject outliers. As described in Section 2.1 the HR has constraints in upper and lower values and can sometimes be highly volatile, however it is assumed that the HR at one time instance is similar to the HR at the previous time instance. This knowledge can be implemented when filtering the HR data in order to bias towards consistent measurements, i.e. if two equally strong signals occur the one in the neighbourhood of the previous value is the most likely true signal.

One method that takes advantage of known behaviour of measured states is Bayes filter, also known as *Recursive Bayesian estimation*. It is a solution to the Bayes filtering problem, i.e. to estimate some state, x_k , given all measurements up until now, $y_{1:k}$ where k is the current step. The Bayes filter is an iterative and recursive process where new measurements, y_k , are feeded into the filter.

The filter uses the last estimate, $p(x_{k-1}|y_{1:k-1})$, and some knowledge of the behaviour of what is measured, $p(x_k|x_{k-1})$, to predict the next estimate, $p(x_k|y_{1:k-1})$, as following:

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1}) p(x_{k-1}|y_{1:k-1}) \,\mathrm{d}x_{k-1}$$
(3.1)

If one for example observes a car that moves in a straight forward direction with a constant velocity, \dot{x}_{k-1} , the next estimate, x_k , is probably $x_k = x_{k-1} + \dot{x}_{k-1}\Delta t$, where Δt is the elapsed time since the last estimation.

The prediction is combined with the likelihood of the measurement, $p(y_k|x_k)$, in an update step to get the new estimate, $p(x_k|y_{1:k})$:

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k) p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})}$$
(3.2)

This process, as seen in Figure 3.2, is repeated over and over again for all k.



Figure 3.2: The recursive process of the Bayes Filter. A prediction is made from the last estimate. This prediction is combined with a measurement to get the new estimate.

3.3 Rauch-Tung-Striebel Smoother

When the HR is estimated some noise can be present in the measurement. By combining previous and future measurements with probability of how the HR changes in a Rauch-Tung-Striebel (RTS) smoother the noise can be decreased. The usage of a smoother, where future data is also used, compared to using a filter the smoother will have a shorter delay and yield more accurate estimates when measurement losses occur. Rauch-Tung-Striebel is a forward-backward style of smoother where the data is processed in two passes. The first pass is standard Kalman filter and the second pass is of a backward-recursion-pass type.

3.3.1 Forward Pass - Kalman Filter

The first pass in the RTS-smoother is a standard Kalman filter. The idea of the Kalman filter is to combine noisy measurements with a prediction by a weighted sum to estimate state more robust to noise. For example if the prediction of the state is the previous state the Kalman filter acts like a low-pass filter. However a more complex prediction, but still simple, e.g. a first order Taylor expansion of the state can decrease the time delay between the estimate and the true state. This can be seen in Figure 3.3 (a) where the low pass filter model has a higher delay than the Taylor expansion model represented as Kalman.



Figure 3.3: Visualisation of the differences between RTS, Kalman and a low-pass FIR filter for generated example data when the true state makes a sudden change of slope.

The Kalman filter achieves this by two steps: a prediction step and an update step. The prediction step describes the change of the state by a linear mathematical model given previously estimated state. This prediction model is called the motion model and is described in equation 3.3.

$$\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{x-1} + \mathbf{q}_{k-1} \tag{3.3}$$

Where \mathbf{x}_k is the state at step k, \mathbf{A}_{k-1} a transition matrix from step k-1 to step k and \mathbf{q}_{k-1} is Gaussian process noise.

The motion model is not only used to calculate a new predicted state, equation 3.4, but also an uncertainty of this state, equation 3.5. The uncertainty is later used as part of the weight when averaging with the measurements.

$$\mathbf{m}_{k}^{-} = \mathbf{A}_{k-1}\mathbf{m}_{k-1} \tag{3.4}$$

$$\mathbf{P}_{k}^{-} = \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^{T} + \mathbf{Q}_{k-1}$$
(3.5)

Where \mathbf{m}_k^- is the predicted mean of the filter at step k and \mathbf{m}_{k-1} the mean of the filter at step k-1. Observe that \mathbf{m}_k^- is before the update and \mathbf{m}_{k-1} is after the last update. In Equation 3.5 \mathbf{P}_k^- is the predicted covariance matrix at step k, \mathbf{P}_{k-1} the covariance matrix after the update at step k-1 and \mathbf{Q}_{k-1} is the covariance of the process noise.

The Kalman filter can estimate states by measuring other states. For example can it estimate the speed by measuring the traveled distance. The mapping between the state domain and the measurement domain is done by a measurement model:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_x + \mathbf{r}_k \tag{3.6}$$

Where \mathbf{y}_k is the measurement at step k, \mathbf{H}_k is the measurement model matrix at step k and \mathbf{r}_k is a vector of Gaussian measurement noise. Observe that it describes the measurement of a state and not the other way around.

By subtracting the measurement of the predicted mean from the measurement, the innovation vector is obtained, \mathbf{v}_k , at step k, equation 3.7. This vector describes not only the error between the prediction and the measurement but also in which direction the prediction should be updated.

$$\mathbf{v}_k = \mathbf{y}_k - \mathbf{H}_k \mathbf{m}_k^- \tag{3.7}$$

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_k^- \mathbf{H}_k^T + \mathbf{R}_k \tag{3.8}$$

Equation 3.8 gives the innovation covariance matrix, \mathbf{S}_k . This is the sum of the predicted covariance and the measurement covariance, \mathbf{R}_k , in the measurement domain. The measurement covariance is most of the time established from experiments.

In equation 3.9 the Kalman gain is calculated, \mathbf{K}_k , based on how big part the predicted covariance is of to the innovation covariance. The measurement matrix in the equation will later be used in the transformation of the innovation vector, \mathbf{v}_k , from measurement domain to state domain.

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}_k^T \mathbf{S}_k^{-1} \tag{3.9}$$

Note that $\lim_{\mathbf{R}_k\to 0} \mathbf{K}_k = \mathbf{H}_k^{-1}$ and $\lim_{\mathbf{P}_k^-\to 0} \mathbf{K}_k = 0$.

The last step in the update step is to use the Kalman gain to update the mean and the covariance matrix.

$$\mathbf{m}_k = \mathbf{m}_k^- + \mathbf{K}_k \mathbf{v}_k \tag{3.10}$$

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T \tag{3.11}$$

To summarise: given a motion model, measurement model, measurement noise and process noise the Kalman filter can optimally estimate the current state given previous measurements.

3.3.2 Backward Recursion Pass

The second pass is a backward pass where the step, k, is iterated backwards. Currently each state is only estimated by previous measurements, by using future measurements one can get a better estimate with a lower time lag.

The gain matrix, \mathbf{G}_k for time step k, is given by:

$$\mathbf{G}_{k} = \mathbf{P}_{k} \mathbf{A}_{k}^{T} \left[\mathbf{P}_{k+1}^{-} \right]^{-1}$$
(3.12)

The new smoothed mean, \mathbf{m}_k^s , for the time step k can be calculated as following:

$$\mathbf{m}_{k}^{s} = \mathbf{m}_{k} + \mathbf{G}_{k} \left[\mathbf{m}_{k+1}^{s} - \mathbf{m}_{k+1}^{-} \right]$$
(3.13)

and the estimated smoothed covariance, \mathbf{P}_k^s , as:

$$\mathbf{P}_{k}^{s} = \mathbf{P}_{k} + \mathbf{G}_{k} \left[\mathbf{P}_{k+1}^{s} - \mathbf{P}_{k+1}^{-} \right] \mathbf{G}_{k}^{T}$$
(3.14)

As seen in Figure 3.3 (a) a low pass filter will always have a delay. The Kalman filter, with a constant velocity model, will only have a delay when the measurements do not comply with the motion model, which is the case when the true state makes a distinct change in step k = 50. The Kalman filter will however find its way back to the true state when the measurements start to comply with the motion model. The RTS-smoother can use its future measurements to decrease this error time and only differ around the step which do not obey the model, this can be seen in Figure 3.3 (b) where the RTS error have a small peak around step k = 50.

3.4 Accuracy Measurements

In order to compare and evaluate results from different methods of HR calculations, several accuracy measurements can be used to measure the error. The measurements are all based on a comparison between the calculated HR and a reference ground-truth HR, e.g. obtained from an ECG. The measure error HR_{error} can be calculated as following: $HR_{error} = HR_{video} - HR_{ECG}$. Thereafter, different accuracy measurements can be obtained [1]:

- Mean of the measure error : $M_e = \sum H R_{error} / n$
- Standard deviation of the measure error : $SD_e = \sqrt{\sum \left[(M_e HR_{error})^2 \right] / (n-1)}$
- Root mean square error: $RMSE = \sqrt{\sum HR_{error}^2/n}$
- Mean of error-rate percentage : $M_{eRate} = \frac{1}{N} \sum HR_{error} / HR_{ECG}$

3.5 Error Matrix

To evaluate a classification, such as a reliable or unreliable HR estimation, an error matrix may be used. This matrix segments the classified data into four categories: true positive, true negative, false positive and false negative. True positive and true negative contains the data which are classified correctly as positive and negative respectively. False positive and false negative contains the data which are wrongly classified as positive and negative. These values are presented in a matrix where each cell contains the number of data points, or sometimes the percentage of total number data points, as seen in Table 3.1. Together with the matrix often sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are featured. [25]

Table 3.1: Overview of an error matri	ix
---------------------------------------	----

	Positive classification	Negative classification	
Condition positive	True positive	False negative	Sensitivity
Condition negative	False positive	True negative	Specificity
	Positivo prodictivo valuo	Nogativo prodictivo valuo	

Positive predictive value Negative predictive value

Sensitivity explains how much data is positive classified of the data which should have been classified positive. The specificity is the same for negative classifications, i.e. how much data is negative classified of the data which should have been classified negative. For a specific data point which already is classified, PPV and NPV describes the probability that the point is classified correctly. PPV describes the probability for a positive classified data point to be correct and NPV describes the same for a negative classified data point. The precise equations for the different values are:

$$Sensitivity = \frac{True \text{ positive}}{True \text{ positive} + False \text{ negative}}$$
(3.15)

$$Specificity = \frac{True negative}{False positive + True negative}$$
(3.16)

$$PPV = \frac{True \text{ positive}}{True \text{ positive} + \text{False positive}}$$
(3.17)

$$NPV = \frac{True negative}{False negative + True negative}$$
(3.18)

3. Theory of Methodology

Methods

This chapter explains the design of the experiments, the details of the developed algorithms and how they were evaluated. Firstly, the video recordings of three data-sets that were used to create and evaluate the HR algorithms are explained. Thereafter, the several steps of the data processing of the HR from video algorithm is presented as well as methods of the different features of the algorithm such as motion robustness and confidence measure. Lastly, the method of the modified algorithm for head tracking data is presented.

4.1 Equipment

Three cameras of the type *Basler acA640 - 100gm* [26] equipped with F1.4/8mm optics were used during the recordings and placed 30 cm apart from each others as seen in Figure 4.1. The cameras were fitted with 850 nm frequency band-pass filter to allow only recordings of this frequency. To light up the scene two flashes from Smart Eye were used where each of them has an output of 5 Watts.



Figure 4.1: Overview of the equipment set-up during data collection.

An ECG was used to measure the reference HR of the subject and used as a groundtruth. The ECG was collected with an Olimex ECG-shield equipped Arduino and to log the 6 dimensions of freedom head pose the software Smart Eye Pro was used [22, 27, 28].

4.2 Data Collection

Three different data-sets were collected for development and analysis of the HR algorithms. For each data-set, the recordings from the centre camera of the camera system was used whereas the others were only there to improve the head tracking of the system. During the recordings, the ECG was sampled at 125 Hz and to sync the ECG with the video the Arduino sent a start signal to the recording PC. The start signal was used to get a time-stamp of when the ECG started recording in the video-time domain. This allowed for a time difference offset between the reference and the video of less than 50 milliseconds.

Stationary data-set

Ten videos were recorded, in an uncompressed format, for investigation of stationary subjects. The ten subjects were sitting still in front of the camera for 60 seconds.

Motion robustness data-set

Videos of four subjects were recorded in order to investigate motion robustness. Each subject performed scaling and translational movements in four different velocities with a duration of 30 seconds, resulting in 8 measurements for each subject and in total 32 measurements. The translational scenario consisted of repetitive horizontal translational head movements and the scaling scenario consisted of repetitive horizontal movements towards and away from the cameras.

BCG-PPG data-set

To determine the relative contribution of BCG and PPG to the intensity changes in the video, one video was recorded for evaluation. In the video, the subject was stationary and its nose was covered with non-transparent adhesive tape, in order to eliminate the effect of PPG. The duration of the video was 60 seconds.

4.3 Heart Rate from Video Algorithm

The data processing pipeline for the HR from video algorithm can be seen in Figure 4.2. The data processing applied in this project was a version of the method from [1]. The steps are described in more detail below. The first step was that the videos, head tracking data from Smart Eye Pro and ECG signals from Arduino were all imported into MATLAB.



Figure 4.2: Data processing pipeline for the HR from video algorithm.

4.3.1 Region of Interest

Five regions of interest (ROI) were tested for the stationary data-set, both automatically chosen from the head tracking data obtained from Smart Eye Pro, as well as manually chosen ROIs. Also, a combination referred to as semi-automatical was used, where the head tracking data was used to find a region which size had been manually selected. The ROIs were chosen from the head tracking data from the first frame of each video for the stationary videos. The raw signal was calculated as the mean value of the intensity of the pixels in the ROI of each frame of the video, $r_{ROI} = [r_1, r_2, \ldots, r_n]$ where *n* is the frame index and *r* is the raw signal. The five different ROIs are described below and can bee seen in Figure 4.3.



Figure 4.3: The 5 different evaluated ROIs; (a) static ROIs (b) the grid for the dynamic ROI where three parameters were evaluated for each square.

1. Face

Semi-automatically chosen ROI where the distance between the tracking data points for eyes and the mouth at the first frame of the video were used to approximate the face region.

2. Cheek

Semi-automatically chosen ROI based on the approximated face ROI and the location of the tracking data point of the nose tip at the first frame.

3. Forehead

Manually selected for each subject by assigning a rectangular ROI over the forehead at the first frame.

4. Nose Region

Semi-automatically chosen rectangular ROI from the head tracking data from the first frame of each video. The width of the ROI was the same as between the feature points of the eyes and the height was the middle third of the distance between the feature points of the eyes and the mouth.

5. Dynamic

Automatically chosen ROI where the image was divided into a grid. Three parameters were evaluated for each square of the grid; maximal temporal derivation, cross-correlation and signal to noise ratio, as in the article from Feng

et al. [29]. The grid squares with lowest maximal temporal derivation, high cross-corelation and high SNR-ratio were chosen as the ROI. The evaluated grid can be seen in Figure 4.3 (b).

4.3.2 Pre-Filtering

Before calculation of the HR, the raw signal was pre-filtered with a three step filtering sequence consisting of three filters; a detrending filter, a moving average filter and lastly a band-pass finite impulse response (FIR) filter.

Detrending filter

Removed trends such as small head movements and small light changes. The filter was taken from the article An Advanced Detrending Method With Application to HRV Analysis by Tarvainen et al [30]. It was a time-varying high pass FIR filter with a cut-off frequency of approximately 0.38 Hz for the HR system.

Moving average filter

Removed high frequency dynamic noise which occurred during the recording. The time window for the filter was 0.125 seconds.

Band-pass filter

Removed frequencies which physiologically could not correspond to the HR. It was a Hamming window band-pass FIR filter with cutoff frequencies [0.67 4] Hz, corresponding to 40 - 240 bpm.

4.3.3 Heart Rate Calculations

To obtain a varying HR, the HR was calculated over a sliding time window centered at each time sample. The chosen default value of the time window was 15 seconds, however, an evaluation of different time window lengths was also performed, see below. In order to find the most occuring frequency for the time window of each sample, Welch's method and a Bayes filter was applied, before extracting the maximal frequency component of each sample.

Welch's Method

When calculating the HR obtained from the videos, the HR likelihood was estimated with Welch's method, described in Section 3.1. An overlap of 50% was used for each time segment.

Bayes Filter

A Bayes Filter (see Section 3.2) was applied after Welch's method to select the most probable HR peak in the power spectral density domain. Since it was a little higher probability that the next peak was close to the previous peak, a model of the type random walk was used to predict the next peak. The random walk was of a Gaussian type with a standard deviation of 15 bpm [23]. This permissive value was chosen from the fact that the HR can change rapidly as mentioned in Section 2.1. The ground truth ECG signal was calculated by evaluating the distance between the periodic peaks of the ECG. For each sample, the mean value of the peak distances within the time window was chosen as the ground truth HR.

4.3.3.1 Time Window for Heart Rate Calculations

In order to decide the optimal size for the time window used for HR calculations, six different sized time windows, $L = [4 \ 8 \ 12 \ 16 \ 20 \ 24]$ seconds, were evaluated. Corresponding error rate variables were calculated by comparing the calculated HR with the ground truth HR from the ECG for the stationary data-set.

4.3.4 Post-Filtering

Because of the properties of the signal together with Welch's Method's window segmentation, some oscillations were present in the filtered signal. To remove the effect, an RTS-smoother was used, described in Section 3.3. The smoother used a motion model of the constant velocity type. For example if the HR was 60 bpm and was increasing with 3 bpm/s, the next predicted HR would be 60 + 3/fs where fs is the sampling frequency. The standard deviation for the motion noise and the measurement noise were tuned to 0.1 and 10 respectively by trial and error.

4.3.5 Motion Robustness

In order to evaluate the motion robustness of the developed HR from video algorithm, the motion robustness data-set was recorded and evaluated. For both types of motions, scaling and translational, a threshold for the maximal velocity the algorithm could handle was found. With the help of the head tracking data feature points, the ROI was updated for each frame and thereby the ROI included the same facial region despite movement. As a comparison, the ROI for the stationary case was the same for all frames. Due to the noisy character of the head tracking data, the feature point position data was temporally filtered with a moving average filter with a time window of 1 second, as in [31].

The motion robustness was evaluated by calculating the mean absolute error (MAE) for each velocity measurement:

$$MAE = \frac{1}{N} \sum |HR_{video} - HR_{ECG}| \tag{4.1}$$

where N is the number of samples of the measurement.

4.3.6 A Confidence Measure

To be able to determine whether the obtained HR signal was accurate or not, a confidence measure was developed by the authors. The measure ranged between 0 and 1 where a value of 0 corresponded to a poor HR signal and a value of 1 corresponded to a very reliable HR signal.

An assumption that the most accurate estimate of the algorithm would occur when the raw signal after the detrending filter was most sinusoidal-like was made. The confidence value could now be represented by the similarity between a sinusoidal model with a periodicity corresponding to the calculated HR and the obtained raw signal after the detrending filter. A close match would mean that the obtained signal from the video would give a high confidence, meanwhile a big difference would mean that the obtained signal from the video would give a low confidence.

First, calculations were performed of the first order of Fourier series coefficients given the measured bpm, HR, and the detrended signal segment, \mathbf{y} , which has n numbers of measurements.

$$\begin{bmatrix} a_0 \\ a_1 \\ b_1 \end{bmatrix} = \mathbf{A}^{-1} \mathbf{y}$$
(4.2)

where

$$\mathbf{A} = \begin{bmatrix} 1 & \cos(\omega \cdot 1) & \sin(\omega \cdot 1) \\ \vdots & \vdots & \vdots \\ 1 & \cos(\omega \cdot n) & \sin(\omega \cdot n) \end{bmatrix}$$
(4.3)

and the frequency in radians is:

$$\omega = \frac{HR}{60 \cdot fs} \cdot 2\pi \tag{4.4}$$

The signals were normalised in order to make the confidence measure indifferent to the amplitude of the signals and only care about the periodicity. Otherwise signals of low amplitude would yield a higher similarity. The normalisation were the following steps:

$$\bar{y}_i = \frac{y_i - a_0}{s} \forall i \in \{1 \dots n\}$$

$$(4.5)$$

$$\hat{y}_i = \frac{a_1 \cos(r \cdot i) + b_1 \sin(r \cdot i)}{s} \forall i \in \{1 \dots n\}$$

$$(4.6)$$

where s is the amplitude:

$$s = \sqrt{a_1^2 + b_1^2} \tag{4.7}$$

The similarity between the sinusoid signal and the detrended signal of the video was compared by looking at the root mean square error (RMSE):

$$e = \left[\sum_{i=1}^{n} \frac{1}{n-1} \left(\bar{y}_i - \hat{y}_i\right)^2\right]^{\frac{1}{2}}$$
(4.8)

The last step was to transform the RMSE to the confidence domain of the interval [0-1]. This was done to be consistent with Smart Eye's other systems. In this domain 1 was the most confident value and 0 represented the lowest confidence.

The RMSE has no upper constraint but it can be assumed that above some RMSE value the estimated HR will not contain any information of the cardiac pulse, therefor an upper limit of the RMSE was set to ten times the amplitude of the normalised theoretical signal, i.e. an RMSE limit of 10. With this limit the RMSE error could be transformed to the Smart Eye domain by:

$$conf = \begin{cases} 1 - \frac{e}{10} & \text{if } e \le 10\\ 0 & \text{if } e > 10 \end{cases}$$
(4.9)

In order to evaluate the developed confidence measure, it was compared to the absolute error of the HR estimate at each time stamp for the data-set of stationary subjects. A threshold for the confidence was determined for which the HR estimate would have a lower absolute error than 2.5 bpm. The accuracy of the confidence threshold was evaluated with the help of an error matrix and its related statistical parameters explained in Section 3.5.

4.4 Heart Rate from Head Tracking Data Algorithm

A modified version of the HR from video algorithm was developed in order to use the head tracking data from Smart Eye Pro for motion measurement of the facial feature points. The method was developed with inspiration from the article of Balakrishnan et al. (explained in more detail in Section 2.3), [11]. The modified algorithm is referred to as HR from headtracking data algorithm and its data processing pipeline can be seen in Figure 4.4. The same pre-filtering process as described in Section 4.3.2 was used. Thereafter, the average of the positions of the head tracking data points were calculated for each time sample. The resulting signal was once again band-pass filtered before the power spectral density was estimated over a sliding time window of 15 seconds. The HR was extracted with a Bayes filter and then post-processed with an RTS-smoother as described in Section 4.3.3 and 4.3.4. The developed HR from head tracking data algorithm was applied to the stationary data-set.

Figure 4.4: Data processing pipeline for the HR from head tracking data algorithm.

4. Methods

5

Results

In this chapter, results regarding the developed HR algorithms are presented. In addition, results on the effect of the BCG signal are shown.

5.1 Heart Rate from Video Algorithm

In this section, properties of the HR signal at varying stages of the data processing process of the HR from video algorithm are shown. The results shown are from one example measurement of a stationary subject except for in Section 5.1.4 and 5.1.9 where the measurements are derived from the whole stationary data-set, as well as in Section 5.1.8 where results from the entire motion robustness data-set are shown.

5.1.1 Pre-Filtering

For each time sample, the mean value of the intensity levels within the ROI was calculated and the raw signal was obtained, see Figure 5.1 (a) for an example of a 10 seconds time segment. The raw signal was thereafter detrended, filtered with a moving average filter and lastly band-pass filtered, see Figure 5.1 (b) - (d) for an example of a 10 seconds time segment after each step respectively. The signal strength of the HR signal can be seen in Figure 5.1 (b), where the top-to-top amplitude of the periodic signal is around 0.8 intensities, the intensity range is between 0 and 255.

5.1.2 Welch's Method

Welch's Method, which was applied after the band-pass filter, estimates the power spectral density of the signal. The spectra contained information of the HR which was represented as a peak, as seen in Figure 5.2 (a) where the HR peak is at -20 dB for 80 bpm. One can also see the first harmonic of the signal at 160 bpm. This HR information was continuously present across time as seen in Figure 5.2 (b).

Figure 5.1: A 10 seconds segment of the obtained a) raw signal b) detrended signal c) moving average filtered signal d) band-pass filtered signal.

Figure 5.2: Power spectral density for one stationary subject. a) power density over frequency at the specific time of 40 s b) power spectral density in dB over time where (a) can be found as a vertical cross-section at 40 s.

5.1.3 Bayes Filter

As seen in Figure 5.2 (a), two peaks were detected. This is a phenomena which occurred for all subjects. The two peaks were related to each other with a ratio of 2 in frequency and appeared unconditioned of whatever the real peak was the upper or the lower one. For most of the time the stronger peak was the correct one and since the Bayes filter used prior measurements it could filter out the other outliers quite well when the peak was weaker for a short time. An example of this can be seen in Figure 5.3 where the Bayes filter rejects the upper outlier correctly when estimating the HR for a subject performing a translational movement.

Figure 5.3: The effect of the Bayes filter for one moving subject. The Bayes filter uses the knowledge of previous estimates when selecting the most prominent frequency in the power density spectra and thus do not always estimates the most powerful peak as the HR.

5.1.4 Evaluation of the Time Window for Heart Rate Calculations

As presented in Section 4.3.3, six different sized time windows were evaluated for each video of the stationary data-set. The ROI used for the calculations was the nose region. The calculated error rates, described in Section 3.4, can be seen in Table 5.1. One can see that the mean of error-rate percentage (M_{eRate}) is already below 2% for a time window of 8 seconds, and time windows with a size larger than 12 seconds have a similar error rate around 1%.

Time window (s)	$M_{e} (bpm)$	$SD_{e} (bpm)$	RMSE (bpm)	M_{eRate} (%)
4	2.17	4.42	4.92	5.37
8	-0.02	1.37	1.37	1.50
12	-0.32	0.87	0.92	0.95
16	-0.38	0.93	1.00	1.01
20	-0.43	0.98	1.07	0.97
24	-0.45	1.03	1.12	1.00

 Table 5.1: Error measurement rates for different time windows for HR calculations, derived from the stationary data-set.

5.1.5 Post-Filtering

ī

When the HR was extracted from the power spectral density, a small oscillation occurred as mentioned in Section 4.3.4. This can be seen in Figure 5.4 where the red plot, before RTS, has an oscillation of around 0.7 Hz and a top-to-top amplitude of around 1 bpm. An RTS-smoother was applied to this extracted signal to remove this phenomena and make the signal oscillation free. This was successful as the oscillations disappeared completely as seen as the blue line in the figure.

Figure 5.4: The oscillations from Welch's Method is removed with an RTS-smoother when estimating the HR for one stationary subject.

5.1.6 Evaluation of Regions of Interest

Five different ROIs were evaluated, described in more detail in Section 4.3.1 and can be seen in Figure 4.3. The automatically derived dynamic ROI can be seen in Figure 5.5 for one subject. As one can see, the dynamic ROI rejected the eyes, the nostrils, some hair and most of the regions close to the border of the face.

Figure 5.5: The dynamic ROI where the green areas represent the obtained ROI for one example subject.

The corresponding error measurements for the five different ROIs can be seen in Table 5.2, derived from the entire stationary data-set.

Table 5.2:	Error	measurement	rates f	for	different	regions	of	interest,	derived	from
		heta	e static	ona	ry data-s	set.				

ROI	M_e (bpm)	$\rm SD_e~(bpm)$	RMSE (bpm)	${ m M}_{ m eRate}~(\%)$
Face	0.01	5.71	5.71	4.20
Cheek	-0.56	1.37	1.47	1.38
Forehead	11.95	22.71	25.65	24.47
Nose Region	-0.43	0.10	1.09	1.03
Dynamic	-0.39	1.43	1.48	1.27

When looking at the error rates, it is evident that the forehead as a region produces the worst result with a M_{eRate} of 24.5%. The best result is achieved by the nose region with a M_{eRate} close to 1%. The cheek and the dynamic ROI gives a M_{eRate} around 1.4% and 1.3% respectively. The complete face as a region, with eyes and partly hair as seen in Figure 4.3 (a), gives a M_{eRate} of around 4.2%.

5.1.7 Stationary Subjects

HR calculations were performed for the stationary data-set, with the nose region as ROI and a sliding time window of 15 seconds. Following, one example of an obtained signal is presented before the absolute error for all 10 videos is presented.

The correlation between the periodic intensity signal from the video of a stationary subject and its ECG can be seen in Figure 5.6. As one can see, the periodicity of the two signals is very correlated.

Figure 5.6: Comparison between the ECG signal and the obtained signal from a video of a stationary subject.

The calculated HR signals from both the video and the ECG signal, over a time period of 60 seconds, can be seen in Figure 5.7. The estimated signal follows the ECG signal closely even when there are subtle changes in HR.

Figure 5.7: The HR from ECG and the estimated HR from a video of a stationary subject.

The estimated HR signal for most subjects are close to the HR observed with the ECG and the absolute error is less than 2 bpm for most of the videos, as seen in Figure 5.8. However, there are a few videos where the error is above 2 bpm and for one video the maximum absolute error peaks at 12 bpm. Overall, the measurements for the entire stationary data-set, when using a sliding time-window of 15 seconds and the nose region as ROI, resulted in a M_{eRate} of 1.03 % and a RMSE of 1.09 bpm.

Figure 5.8: The absolute error of the estimated HR in bpm for the stationary data-set.

5.1.8 Motion Robustness

To evaluate the motion robustness of the HR from video algorithm, different velocities of both translational and scaling movements were recorded, see the motion robustness data-set in Section 4.2. The results for translational and scaling movement can be seen in Figure 5.9 (a) and (b) respectively, where the mean absolute error (MAE) of the HR measurements for different velocities is plotted.

Figure 5.9: MAE for different velocities from the 16 measurements of the motion robustness data-set for (a) translational movement (b) scaling movement.

One can see that for low velocities, below 6 cm/s, MAE is low for both scaling and translational movements. However, for scaling movements with a velocity above 6 cm/s, the algorithm does not seem to work resulting in high MAE. For translational movement, MAE is low, below 10 bpm, up until approximately 8 cm/s where the algorithm can not handle any faster motions. For both types of motion, it looks like MAE decreases for very high velocities, over 25 cm/s, however this is due to the subject moving with a motion periodicity similar to its own HR.

5.1.9 Evaluation of the Confidence Measure

The objective of the confidence algorithm is to classify the estimated HR from the HR from video algorithm as reliable or unreliable. By analysing how the absolute error of the estimated HR depends on its corresponding confidence value for each time stamp a region can be identified, seen in Figure 5.10, where the estimated HRs with a higher confidence than 0.86 have a low absolute error, most of them beneath 2.5 bpm.

Figure 5.10: Scatter plot of absolute error for each estimated HR from the video algorithm against its confidence for the stationary data-set. A confidence threshold of 0.86 can be identified where the estimated HR that has a higher confidence value can be classified as reliable.

The confidence value can be used as a threshold to indicate if the estimated HR is reliable. The sensitivity and specificity of this classification is 87.8% and 93.9% respectively. Table 5.3 shows an overview of the result.

	Inside Threshold	Outside Threshold	
Error ≤ 2.5 bpm	83.1 %	11.6%	Sensitivity: 87.8 %
Error > 2.5 bpm	0.319%	4.95%	Specificity: 93.9%
	PPV: 99.6 %	NPV: 29.9 %	

 Table 5.3:
 Error matrix of reliable/unreliable classification

As seen in Figure 5.11, high absolute error values are classified as unreliable but some estimated HR which have a low absolute error are also falsely classified as unreliable.

Figure 5.11: The absolute error against time for one stationary subject where each estimated HR is classified as reliable or unreliable depending on whether they fulfil the confidence threshold of 0.86.

5.2 Heart Rate from Head Tracking Data Algorithm

Extracting the HR from the head tracking data was investigated for the videos of the stationary data-set. The accuracy measurements of the results can be seen in Table 5.4. When comparing the results with the error rates from the HR calculated from intensity changes of an ROI, see Section 5.1.7, one can see that the accuracy is not as good, M_{eRate} of 11.25% compared to 1.03%.

 Table 5.4: Error measurement rates for HR measurements from head tracking data, derived from the stationary data-set.

Measurements	M_e (bpm)	$\rm SD_e~(bpm)$	RMSE (bpm)	M_{eRate} (%)
Tracker Data	3.31	9.65	10.19	11.25

However, if one looks at the absolute error for each of the video, see Figure 5.12, one can see that for some of the videos and some parts of the recordings, the results are very accurate (absolute error below 5 bpm).

Figure 5.12: The absolute error in bpm for HR calculations from head tracking data, derived from the stationary data-set.

5.3 The Effect of the Ballistocardiographic Signal

As mentioned in Section 2.3.1, both head motions (BCG signal) and absorption changes (PPG signal) contribute to intensity changes in a video. To investigate the contribution of the BCG signal to the HR signal obtained with the HR from video algorithm, the BCG-PPG data-set was evaluated. In the video of the data-set, the PPG signal had been eliminated from the nose region with non-transparent adhesive tape. The results can be seen in Figure 5.13, where the HR from video algorithm are represented twice: once with PPG and BCG information with an ROI from the forehead and once with only BCG information from the nose region. This is compared to the HR obtained from the ECG and from the HR from head tracker data algorithm.

Figure 5.13: Comparison of HR for the video of the BCG-PPG dataset calculated by different methods; ECG, PPG signal from skin region, PPG signal from tape region and from movement of head tracking data points.

Discussion

In this chapter, the results presented in the previous chapter are discussed in terms of algorithm accuracy, reliability and computational complexity.

6.1 Discussion of Heart Rate from Video Algorithm

Below follows a discussion about the different parts of the HR from video algorithm.

6.1.1 The Effect of the Time Window of Heart Rate Calculations

If one only looks at the RMSE data in Table 5.1 and interpolates the data, the conclusion would be that a time window of around 13s would yield the most accurate HR estimation. This might be surprising since a longer time window should be more robust against noise and outliers and therefore yield a more accurate estimation. The problem might be that the reference HR from ECG is the mean HR for the period while the estimated HR from video is the strongest frequency during the time window. With a longer time window the signal strength can fluctuate more within the time period and therefore trick the Welch's method to bias against short strong signals. This can explain why the error, such as RMSE and SD_e , increases for longer time windows.

However, the time period of 13 s might still not be suitable since it is a matter of what kind of information one is interested in. The accuracy is not the only important parameter when deciding the length of a time window. A longer time window will have a longer delay, approximately half of the window length, before HR changes will be present in the estimated HR. Another downside to a longer time window is that it needs longer time before the first HR value can be estimated, i.e. if a time window of 24 seconds is used it will take 24 seconds before the first HR estimation is available.

In other words, there is a trade-off between signal delay and accuracy when determining the length of the time window. Some users, for example ones who would like to measure resting HR during sleep, may be more interested in a high accuracy rather than to detect rapid changes in HR as early as possible. Meanwhile the opposite may be true for users who are interested in monitoring car drivers where it may be for example important to detect sudden cardiac arrest as quickly as possible. Hence a perfect general time window length does not exist, it is application based.

6.1.2 Regions of Interest

One can see in Section 5.1.6 that the ROIs nose region, cheek and dynamic all produce an HR estimate with low error rates, where the nose region yields the most accurate estimation. This result combined with other advantages makes the nose region to the most appropriate ROI. For example, the dynamic region consists of complicated time-consuming calculations where the values for the grid squares are averaged over the entire video length. The cheek ROI and nose region ROI are overlapping a great deal in space, however the nose region also contain the nose ridge and the nose tip. In addition, the nose region is rarely cloaked by e.g. glasses or facial hair.

The nose region ROI is also believed to be more generic and robust to head motions than the other regions since its position is in the middle of the face and therefore more robust to changes in position without reaching the area outside the facial skin area.

6.1.3 Motion Robustness

The developed HR from video algorithm has a limited robustness for motion. As the results for translational and scaling movements show, see Section 5.1.8, the algorithm can handle movements with a velocity of 6 cm/s and 8 cm/s respectively for the two movements. However, movements in real-life scenarios such as for the case of driver monitoring are very different to the studied movements. In addition, the data-set is of a limited size with only 16 measurements from 4 different subjects for each movement. Though, the results presented on motion robustness is an indication on what the HR from video algorithm can handle when it comes to movement but for a full analysis, motion robustness has to be studied in more depth with more subjects and measurements.

For driving situations, fast and short movements are common which the HR from video algorithm will not be able to track. However, due to the short time period of the movements, the algorithm could be implemented so that it looks past this movements and work for the longer time periods where the driver is relatively still. One idea would be to integrate the velocities calculated from the head tracking data into the confidence measure of the signal, where one would classify time samples with too high velocity to be unreliable. For measurements as the driving scenario, a short time window for HR calculations is also preferable since the effect of the short fast motions then will be smaller in time.

The change in intensity of the image depends a lot on the motion in relation to the flashes used in the system. In the system used for the recordings of this study, two flashes are used that are connected to the outer cameras, see Figure 4.1. One thought is that the direction of movements in relation to the flashes will effect the motion robustness of the HR from video algorithm. For example, the algorithm is more robust to translational movements than scaling movements which might be due to the different directions of the movements compared to the flashes.

The camera system setup resulted in two clear shadows from the nose at each side and as the head moved the shadows changed shape, resulting in a BCG signal. For other camera systems setups, this might not be the case and hence the BCG signal would not be as strong, which could be a cause of concern. The shadows could also disappear if the head pose in relation to the flashes changes which would need to be further investigated.

6.1.4 Confidence Measure

As mentioned in Section 5.1.9 the classifications from the confidence measure have some false negatives but very few false positives. This results in a high PPV and a low NPV which means that the classification can be used as an indicator of when the HR estimation is reliable but not as a good indicator of when the estimation is unreliable. The classification is also too strict for many real world applications. An absolute error of 2.5 bpm for some applications may be more than acceptable, even absolute errors of up to 10 bpm may be acceptable when one is not interested in the exact HR but one rather wants to detect if the subject suddenly got stressed. The confidence measurement lack this permissiveness ability and can only detect if the error rate is very low.

The confidence measurement uses a sinusoidal signal to measure how the detrended signal deviates from a theoretical noise free signal, as mentioned in Section 4.3.6. However a perfect theoretical HR signal is probably not sinusoidal since, see Figure 5.1 (b), the detrended signal is more of a sawtooth wave. Even small individual variations may occur. This can explain why most confidence values for the stationary data-set that yields in a good estimated HR have a confidence of 0.94 and no measurement has a confidence over 0.96. An evaluation against a sawtooth wave would probably give a little more precise confidence measurement.

As seen in Figure 5.10 most confidence values are present roughly in the interval from 0.5 to 1. This is due to the RMSE cut-off above 10, mentioned in Section 4.3.6. A lower cutoff value would yield in a more even distribution of confidence values over the entire interval of 0 and 1. The precision of the confidence measurement would not increase by this (neglecting the floating point precision error), but it could be more readable for humans since less decimal numbers could be used. Another way to make the confidence value easier to understand to the end user may be to present the reliable-classification: 1 for reliable and 0 for unreliable. The Smart Eye Pro software already does this for some quality measurement outputs.

Another reason for that the data is unevenly distributed over the confidence interval may be because the stationary data-set is too accurate. Since a majority of the samples in the stationary data-set gives a low error rate, the confidence also gives high values. The lack of measurements with an absolute error in the 10 bpm region makes it hard to find a threshold which detects less accurate measurements that are still useful for some applications. If more data was to be collected in more different situations another more permissive threshold may be found.

6.2 Heart Rate from Head Tracking Data Algorithm

Extracting the HR from the head tracking data of the Smart Eye Pro software is not as accurate as looking at the intensity variations of a chosen ROI, as one can see if one compares the error rates for HR from head tracking data algorithm in Table 5.4 (11.25 % M_{eRate}) with the error rates for the HR from video algorithm in Section 5.1.7 (1.03 % M_{eRate}). However, if one looks at the absolute error for the ten videos of the stationary data-sets in Figure 5.12, it is clear that for some subjects and some part of the measurements, the method results in very low absolute errors. Since the head tracking data already is an output from the software of Smart Eye, it would be simple to implement the HR from headtracking data algorithm which makes it into a very attractive method even though the accuracy is not as high as for a ROI solution.

One interesting effect of extracting the HR from measurement of motion of the head tracking data is that the method works even for videos where occlusions are present, such as protective masks, hair and makeup etc, assuming that the head tracking of feature points still is accurate. Since the HR from head tracking data algorithm does not measure the variations in intensity of a facial skin region, the method is indifferent to if the skin is visible or not. In addition, it makes it possible to also post process old logged head tracking data where the video is no longer available.

The length of time window used for HR calculations is not evaluated for the HR from head tracking data algorithm. It is assumed that the results would be similar as for the HR from video algorithm since the varying signals after the detrending filter are similar for the two methods. Hence, it can be stated that the optimal length of the time window is application based for HR from headtracking data algorithm as well.

Motion robustness of the HR from head tracking data is not evaluated either, nor is a confidence measure developed. The development of the algorithm can be seen as a proof of concept that the HR can be extracted from the headtracking data of Smart Eye. However, as stated, the properties of the algorithm have not been fully analysed.

6.3 The Effect of the Ballistocardiographic Signal

When analysing Figure 5.13, one can see that the three different methods for HR extraction follow the ground truth HR from the ECG well. The obtained HR from the skin and the tape signals are very similar, apart from a difference at t = 6s, where the tape ROI was subjected to an external shadow. It is a strong proof that the BCG signal contributes extensively to the intensity changes, since the HR can be calculated solely from an ROI where the PPG signal has been eliminated, as in the case with non-transparent tape. In addition, it also shows that the HR signal can be extracted even though the skin is not visible. The HR calculated from the tracking data is not as correct as the others which corresponds to the findings in Section 5.2.

For further understanding, measurements where the PPG signal is present on its own and the BCG signal is eliminated needs to be studied. One drawback if the BCG signal stands for the major contribution to the HR signal is that the motion robustness is not as high as for PPG signals as Irani et al. stated.

6. Discussion

7

Conclusion

In this study, an approach for extracting HR from the system of Smart Eye is made. One main application for the systems of Smart Eye is driver monitoring in vehicles where knowing the HR of the driver is of great interest. By knowing the HR, the stress level of the driver could be monitored and the safety system of the vehicle could be adjusted accordingly.

The developed HR from video algorithm works very well for stationary subjects resulting in a M_{eRate} of 1.03 %. Different regions of the face were evaluated for HR extraction and it is concluded that the nose region is the optimal ROI, resulting in the highest accuracy and several advantages such as better motion robustness as well as easy implementation. To obtain a varying HR, the HR is calculated over a sliding time window. The choice of window size is a trade-off between HR accuracy and signal delay. Which window size to use is application based depending on what information is of most importance for the specific application.

The motion robustness of the HR from video algorithm is limited. For translational and scaling movements, the algorithm is robust for motions up to velocities of 8 cm/s and 6 cm/s respectively. Even though the results are indications on what the algorithm can handle in terms of motion, movements in real life situations such as driving are of much more complicated nature.

The confidence measure, which was introduced to be able to reject unreliable HR estimations, can only classify HR estimates as reliable when the confidence is very high. This may be acceptable when accurate measurements are significant, however, for many applications where a rough HR estimation is used to detect a stressful situation this is probably a too strict classification restriction.

The HR can also be extracted from head motion by studying the variation of position of head tracking data. However, the modified algorithm adapted to head tracking data has a lower overall accuracy (M_{eRate} of 11.25 %) than the method of studying intensity variation of an ROI (M_{eRate} of 1.03 %). Though for some parts of the measurements, the accuracy was still very high which means that the HR from head tracking data algorithm has potential. Being able to measure the HR from the head tracking data would mean an easy implementation since the data already is an output from the software Smart Eye Pro.

7. Conclusion

Future Work

For future work, some recommendations can be made. While HR estimation of stationary subjects works very well, HR estimation for moving subjects needs further investigation such as a bigger data-set where more complex movements are included. Currently, it is hard to say how accurate the HR algorithms are for real world applications.

The confidence measure could also be further incorporated in the method to remove segments with too high noise level before Welch's method. This could lead to higher robustness for noise and outliers, especially longer time windows could benefit from this. To increase the reliable-classification accuracy, a more sophisticated confidence method can be used where also the head tracking data is taken under consideration. Also, a bigger data-set with higher HR estimation errors could yield a more permissive confidence threshold for the classification, which would lead to a better classification where higher errors could be allowed.

One more recommendation for future work is to investigate the relation between the signal strengths of the PPG and BCG signals. Our findings indicate that BCG is the dominant signal but it can not be disclosed how strong the PPG signal is.

Bibliography

- X. Li et al. "Remote Heart Rate Measurement from Face Videos under Realistic Situations". In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. June 2014, pp. 4264–4271. DOI: 10.1109/CVPR.2014. 543.
- [2] Smart Eye AB. Smart eye. 2016. URL: http://www.smarteye.se.
- [3] W. Verkruysse, L. Svaasand, and J. Nelson. "Remote plethysmographic imaging using ambient light". English. In: Optics Express 16.26 (2008), pp. 21434– 21445.
- [4] M. Poh, D. McDuff, and R. Picard. "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation". English. In: *Optics express* 18.10 (2010), p. 10762.
- [5] M. Garbey et al. "Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery". In: *Biomedical Engineering*, *IEEE Transactions on* 54.8 (Aug. 2007), pp. 1418–1426. DOI: 10.1109/TBME.2007.891930.
- [6] M. van Gastel, S. Stuijk, and G. de Haan. "Motion Robust Remote-PPG in Infrared". English. In: *IEEE Transactions on Biomedical Engineering* 62.5 (2015), pp. 1425–1433.
- G. Cennini et al. "Heart rate monitoring via remote photoplethysmography with motion artifacts reduction". In: Opt. Express 18.5 (Mar. 2010), pp. 4867– 4875. DOI: 10.1364/0E.18.004867.
- [8] K. Humphreys, T. Ward, and C. Markham. "Noncontact simultaneous dual wavelength photoplethysmography: A further step toward noncontact pulse oximetry". In: *Review of Scientific Instruments* 78.4, 044304 (2007). DOI: http://dx.doi.org/10.1063/1.2724789.
- [9] K. Humphreys. "A CMOS camera-based system for clinical photoplethysmographic applications". In: *Proceedings of SPIE*. Vol. 5823. 2005, pp. 88–95.
- Scott Prahl. Molar extinction coefficients of oxy and deoxyhemoglobin. 1999.
 URL: http://omlc.ogi.edu/spectra/hemoglobin.
- [11] G. Balakrishnan, F. Durand, and J. Guttag. "Detecting Pulse from Head Motions in Video". In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. 2013, pp. 3430–3437.
- [12] M. Valentini and G. Parati. "Variables Influencing Heart Rate". In: Progress in Cardiovascular Diseases 52 (2009), pp. 11–19.
- C. Borst et al. "Mechanisms of initial heart rate response to postural change". In: American Journal of Physiology - Heart and Circulatory Physiology 243.5 (1982), H676–H681.

- [14] National Heart, Lung and Blood Institute. How Is Sudden Cardiac Arrest Diagnosed? http://www.nhlbi.nih.gov/health/health-topics/topics/ scda/diagnosis. Accessed: 2016-06-02.
- [15] National Heart, Lung and Blood Institute. What Is an Arrhythmia? http: //www.nhlbi.nih.gov/health/health-topics/topics/arr. Accessed: 2016-06-02.
- [16] A. Workers A and A. Sargent. ECG Workbook. 2nd. M&K Publishing, 2011.
- [17] J. Allen. "Photoplethysmography and its application in clinical physiological measurement". In: *Physiological Measurement* 28.3 (2007), R1–R39.
- [18] A. Hertzman and C. Spealman. "Observations on the finger volume pulse recorded photo-electrically". In: American Journal of Physics 119 (1937), pp. 334– 335.
- [19] D. He, E. Winokur, and C. Sodini. "A Continuous, Wearable, and Wireless Heart Monitor using Head Ballistocardiogram (BCG) and Head Electrocardiogram (ECG)". In: 33rd Annual International Conference of the IEEE EMBS. 2011, pp. 4729–4732.
- [20] R. Irani, K. Nasrollahi, and T. Moeslund. "Improved Pulse Detection from Head Motions using DCT". In: Computer Vision Theory and Applications (VISAPP), 2014 International Conference on. Vol. 3. 2014, pp. 118–124.
- [21] Larry J. Mengelkoch, Daniel Martin, and John Lawler. "A Review of the Principles of Pulse Oximetry and Accuracy of Pulse Oximeter Estimates During Exercise". English. In: *Physical Therapy* 74.1 (1994), p. 40.
- [22] Smart Eye AB. Smart Eye Pro. Version 6.1. URL: http://smarteye.se/wpcontent/uploads/2014/12/Smart-Eye-Pro.pdf.
- [23] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.
- [24] P. Welch. "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms". In: *IEEE Transactions on Audio and Electroacoustics* 15.2 (June 1967), pp. 70– 73. ISSN: 0018-9278. DOI: 10.1109/TAU.1967.1161901.
- [25] "Selecting and interpreting measures of thematic classification accuracy". In: *Remote Sensing of Environment* 62.1 (1997), pp. 77–89. ISSN: 0034-4257. DOI: http://dx.doi.org/10.1016/S0034-4257(97)00083-7.
- [26] acA640-100gm. 1st ed. Basler. URL: http://s.baslerweb.com/media/ documents/BD00054701_Basler_acA640-100gm_EMVA_Standard_1288.pdf.
- [27] Arduino. URL: https://www.arduino.cc/en/main/arduinoBoardUno.
- [28] Olimex Ltd. 2011. URL: https://www.olimex.com/Products/Duino/ Shields/SHIELD-EKG-EMG/resources/SHIELD-EKG-EMG.pdf.
- [29] L. Feng et al. "Dynamic ROI Based On K-means For Remote Photoplethysmography". In: *IEEE International Conference on Acoustics, Speech and Signal Processing* (2015), pp. 1310–1314.
- [30] M. Tarvainen, P. Ranta-aho, and P. Karjalainen. "An Advanced Detrending Method With Application to HRV Analysis". In: *IEEE Transactions On Biomedical Engineering* 49 (2002), pp. 172–175.
- [31] Å. Rustand. "Ambient-light Photoplethysmography". MA thesis. Norwegian University of Science and Technology, 2012.