Contact-Free Heart Rate Measurement From Human Face Videos and its Biometric Recognition Application

Muhammad Waqar

Department of Computer Science
Aberystwyth University

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

April 2019

I would like to dedicate this thesis to my loving parents.
(Mr. & Mrs. Tila Muhammad)

Declaration and Statement

DECLARATION

This thesis has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date .....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where **correction services**[1] have been used, the extent and nature of the correction is clearly marked in a footnote(s).
Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (candidate)

Date .....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date .....

[1] This refers to the extent to which the text has been corrected by others.

Acknowledgements

All the praises to the Almighty Allah (the God) who is the most merciful and beneficent. I am grateful for the opportunity and financial support from the Aberystwyth University, which made my Ph.D. study possible. I would like to express my greatest gratitude to my supervisor Dr. Bernard Tiddeman, who gave me enormous help from the very beginning of my Ph.D. application, all the way to conducting scientific research and writing up my thesis. He has always been patient to discuss with me about my project ideas, listen to my doubts and provide insightful suggestions. My appreciation also goes to my second supervisor Prof. Reyer Zwiggelaar, for his helpful comments and feedback on my project.

I would like to acknowledge the former and present members in the Vision, Graphics and Visualisation Group, for their inspiring talks and discussions in research. My appreciation also extends to all my laboratory colleagues, technical support officers especially Mr. Sandy Spence, who provided me with all the experimental equipment, and academic operations administrator Mrs. Michelle Symes, who is the symbol of kindness and support for all of us.

I have no words to express my thank to my loving parents, whose care and prayers brought me to this day of success. I salute their efforts and encouragement, they gave me at every stage of my life. I am also grateful to all my family members and friends for their moral support and encouragement. Finally, I would like to express my deepest gratitude to my lovely wife Hadia, and little angels Wafa and Aleeza, for their unconditional love throughout this long journey.

Abstract

The heart is the most important muscular organ of a human body and its strength can be assessed by the rate at which it beats [1]. Heart rate measurement plays an important role in human health assessment and there have been a number of methods suggested in order to monitor it remotely with more ease and comfort. Contact-free heart rate measurement is one way to make it user-friendly and can also be used for covert surveillance. The performance of previously developed touch-free methods was found to be efficient and accurate under controlled conditions. In realistic and more challenging scenarios, their performance is degraded, as each method has its limitations. Typically, the performance is dependent on controlled lighting and limited subject movement. More realistic situations require more robust contact-free ways to measure the heart rate. Our work aims to obtain a good understanding of the underlying problems of non-contact pulse estimation. We have proposed a method that can overcome many of the problems of light reflection and subject's movement while measuring the heart rate from human faces remotely using an ordinary webcam. We use Saragih face tracker [2] to track the faces from the recorded videos which provides more reliable extraction of a region of interest (ROI) than the simple face detection. We find the robust mean of the skin pixel's color values of the selected ROI which can improve the measurement of the skin color in each frame, reducing the impact of facial hair, wrinkles, and lighting. In addition, we calculate the least-squares error optimal filter using our training dataset to estimate the heart rate more accurately from the measured color changes over time. These methods not only improved the accuracy of heart rate measurement but also resulted in extraction of a cleaner pulse signal, which can be integrated into many other useful applications. We explore whether this technique has the potential for advancing human biometric systems employing the pulse signals obtained. We have presented a method for the biometric recognition application of the pulse signals obtained from the facial videos. The Radon transform images of the pulse signals obtained from facial imagery were used for the extraction of distinctive features, and the decision tree and ANNs machine learning techniques were employed to classify those features for identification purposes. Results obtained

show that our proposed contact-free heart rate measurement method has significantly improved on existing methods, and that biometric recognition systems can be improved with the use of the techniques we propose.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

The heart is the most important muscular organ of the human body that controls the flow of blood through all parts of the body. The heart rate assesses the strength and performance of a heart and the continuous monitoring of heart rate is vital for adequate human health treatment [1]. Heart rate can be measured by different sensory equipment such as an ECG machine [27], pulse oximeter [28], and Doppler probe [29] etc. The performance of all these types of machinery is quite satisfactory but the way in which they are used is quite uncomfortable for the patient.

Contact-free measurement of heart rate can ease the discomfort of subjects and make it user-friendly. Many different attempts have been made in the literature to measure the heart rate contact-free using a thermal camera [9] and/or a webcam [10]. All methods have reliable performance in controlled conditions, but the accuracy remains a question mark if they are exposed to more challenging and realistic capturing scenarios. The subject's motion artifacts and illumination variations can be expected in more challenging conditions, which require a robust approach to measure the pulse rate accurately and extract the pulse signal cleanly. Therefore, there is still lots of room to improve the existing methods with respect to robustness, automation, accuracy, algorithm efficiency, and the characteristics of existing methods for specific applications still can be more thoroughly investigated. The two main problems that affect the performance of contact-free heart rate measurement are subject's movement and illumination variations. The extracted pulse signal, if obtained cleanly from the subject's facial imagery, could have several useful applications such as human biometric recognition, lie or stress detection, sports analysis, etc.

1.2 Research Aims and Objectives

Our research aims to investigate contact-free heart rate measurement methods, more specifically, the main problems of subject's motion and illumination noise in estimating heart rate from human face videos. We try to obtain a good understanding of these problems, the existing methods, their performance under realistic and more challenging conditions, and the possible applications of the contact-free pulse extraction from facial videos. We look into the most popular and/or the most promising methods, identify their limitations and propose possible improvements.

The main objectives of our research can be listed as follows:

- Conducting a thorough literature review of existing non-contact heart rate measurement methods, and the biometric recognition application of the extracted pulse signals. We re-implement the most promising methods for the extraction of heart rate from facial videos using an ordinary webcam in a more challenging capturing scenario. We analyse these methods using our locally captured dataset and the publicly available dataset to draw a better conclusion.

- Improving the existing methods for contact-free heart rate measurement by addressing their limitations and introducing additional parameters to overcome these weaknesses. The main problems of almost all the previous methods are subject's movement and illumination variations. We address these problems in our work and propose a solution to minimize the effect of each. We introduce Saragih *et al.* [2] face tracking technique, detrending, and robust mean color pixels algorithm into non-contact heart rate measurement system to minimize the effect of these problems and to obtain an improved accuracy of the estimated heart rate.

- Developing a new method to achieve robustness in the non-contact pulse rate measurement system under more challenging and realistic capturing scenarios. We propose a method that can overcome many of the problems of light reflection and subject's movements while measuring the heart rate from human faces remotely using an ordinary webcam. We find the robust mean of the skin pixel's color values and calculate the least-squares error optimal filter using our training dataset to estimate the heart rate. This method can not only improve the efficiency of heart rate measurement systems but can also result in extraction of a cleaner pulse signal, which can have many other useful applications such as human biometric recognition.

- Comparing the results of our proposed method with different approaches of contact-less heart rate measurement from human facial imagery by using the locally created datasets and the publicly available dataset to draw a conclusion about the best possible outcome.
- Using the extracted pulse signals obtained from facial videos for other useful applications. We improve a method for the biometric recognition application of the pulse signals obtained from facial videos.

1.3 Thesis Outline

The remaining chapters of this thesis are organised as follows:

- Chapter 2 provides a detailed background of the heart rate measurement methods. The importance of contact-free heart rate measurement methods is highlighted. The thorough literature review of the existing non-contact heart rate measurement methods based on facial videos using a thermal camera and a webcam is conducted. The main limitations of existing methods are addressed such as subject's movement and illumination noise. The existing challenges and the key areas of the non-contact pulse estimation which requires further improvement are reported. The possible application of human biometric recognition based on the pulse signals extracted from facial imagery is also described in detail.
- In Chapter 3, we present a robust contact-free heart rate measurement from human face videos using a webcam in a challenging capturing scenarios. The main problems of the previous work are highlighted. The processing steps involved in the development of proposed method are described. The datasets used in the experiments are defined. The results are presented, which are obtained by comparing the proposed method with the ground truth, and the state-of-the-art methods using both local and public datasets. The chapter ends with the discussion and conclusion sections.
- Chapter 4 presents the human biometric recognition application of the proposed pulse signal extraction method. The methodology is explained in detail and the results are compared to the existing method to evaluate the performance of the proposed system.
- In Chapter 5, the work presented in the thesis is summarised. The contributions of the work are listed and future work is also discussed.

Chapter 2

Literature Review

This chapter provides the background of the heart rate measurement systems for human subjects. A wide range of heart rate estimation technologies and the terminologies used in them are introduced. Existing challenges in the area of non-contact heart rate measurement from human face videos are highlighted, which we have attempted to overcome in our work. Finally, the human biometric recognition application of the contact-free pulse signal from facial videos is also discussed.

2.1 Vital Signs

The physiological measurements that assess the basic functions of a human body and are vital for the human health monitoring are called vital signs [30]. They are good indicators of the general physical health of a person, which help diagnose the possible diseases and trace the progress towards recovery. There are four main vital signs which are generally monitored by a healthcare professional which includes: body temperature, respiration rate, blood pressure, and pulse rate or heart rate.

2.1.1 Heart

The heart is a muscular organ that lies in the center of our chest. It pumps oxygenated blood to all parts of the body and receives less oxygenated blood from the body to pass on to the lungs to absorb more oxygen and supply it back to the body. The structure of the heart comprises four main chambers shown in Fig. 2.1. The two chambers on the left side are called the left Atrium and the left Ventricle, which supply oxygenated blood to the body through the Aorta. The other two on the right side are the right Atrium and the right Ventricle, which receive less oxygenated blood through the Vena

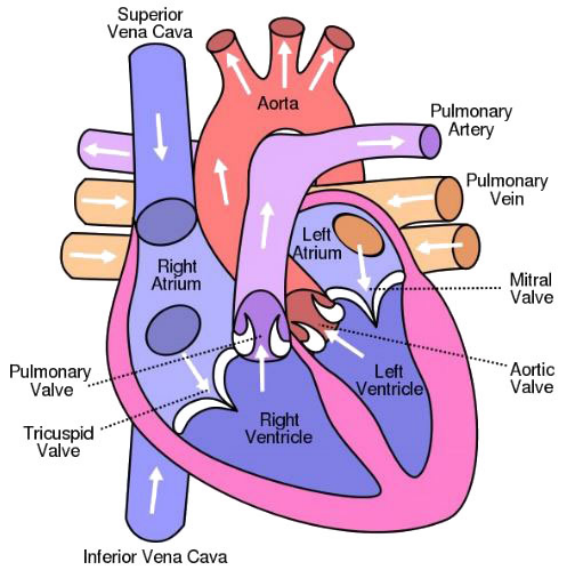


Fig. 2.1 Basic structure of a human heart [3].

Cava and pass it to the lungs to absorb oxygen. These chambers are connected by valves, which ensure the one-way flow of blood through them. The veins carry the blood towards the heart and arteries take the blood away from it. The flow of blood through the heart does not happen automatically but it is driven by contractions of the heart muscle which are triggered by electrical impulses. The journey of the blood from entering the heart and leaving it is called one cardiac cycle or pulse and the heart is said to beat once [31].

2.1.2 Heart Rate

The Heart Rate (HR) refers to the number of times a heart beats in one minute [32]. The normal resting heart rate of a healthy person ranges between 60 and 80 beats per minute (bpm) [33]. As the heart is a muscle and like all other muscles, it can also strengthen by doing exercise. People in a good physical condition, for example, a marathon runner or a swimmer can have a resting heart rate as low as 35 bpm. The maximum heart rate of a person is highly correlated to the age of a person [34]. The equation to estimate it is: HR_{max}= 208 – 0.7 $\times$ Age. For a person of age 30, this works out as 187 bpm. According to this theory, the maximum heart rate of children is higher than for adults. To cover a wide range of acceptable heart rate values, we have chosen to limit the range to 48-240 bpm for the subjects age range of 25-40 years in our study. In order to make it possible to estimate the heart rate of children using our method, we set the maximum threshold to 240 bpm because the heart rate of kids can reach up to 240 bpm.

2.2 Heart Rate Measurement Methods

The continuous monitoring of heart rate is very important to assess human health. Several different methods are currently in practice to measure the heart rate. The most common methods in practice are described in this section.

2.2.1 Sphygmology

Sphygmology is a traditional way to measure heart rate and is in practice from over two thousand years ago. A famous physician Bian Que (C. 500 B.C.E) claims that he examined the pulse of the prince of Guo for the first time ever when people actually thought that he was dead, determined that he was in a state of coma [35]. The heartbeat can be counted by placing the index and middle fingers on the subject's radial arteries located near the wrist. When the fingers are pressed down, we can feel a throbbing effect, which is caused by the pulse. Unique patterns in the subject's body can be discovered by measuring the pulses, which can help diagnose illness [35]. Fig. 2.2 shows a person counting the pulses of another person. This method of estimating the heart rate is still in practice because it is the cheapest and quickest way to assess a subject's health condition. However, there are some limitations in this method, for example, cold hands and illness can cause errors. In today's technological world, such a traditional way of estimating the heart rate has a lower accuracy when compared to the gold standard of ECG [36].

Fig. 2.2 An indian doctor takes the patient's pulse. Image by a Delhi painter, name unknown, circa 1825. Image credit: Welcome library, London [4].

2.2.2 Phonocardiogram (PCG)

PCG (Phonocardiogram) is a heart rate measurement technique that detects and register the heart sounds which can be visually represented as a graphical chart. The heart sounds are produced due to several cardiac activities which include; opening and closing of four values of a heart, the flow of blood through these values, the blood flow through the ventricular chambers, and rubbing of heart surfaces [5]. The generated heart sound is referred to as "lub-dub". Fig. 2.3 shows the signal produced as a result of heart sounds, where s1 represents the sound 'lub' and s2 referred to as 'dub'. The other two sounds s3 and s4 are the gallop rhythms of the heart. PCG estimate the heart rate effectively but the touch-based application of this method make it uncomfortable for the subject. The external noise interference in the noisy environment is the main cause of the noisy signal generation, which could result in the incorrect estimation of heart rate.

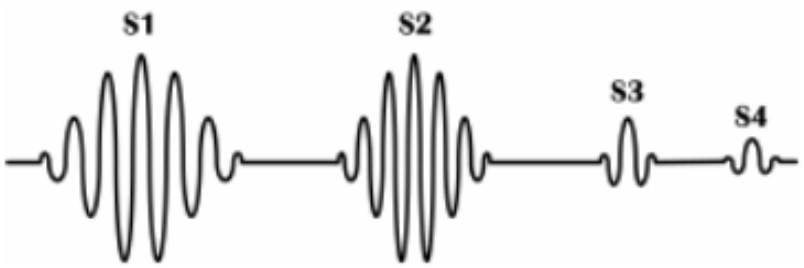
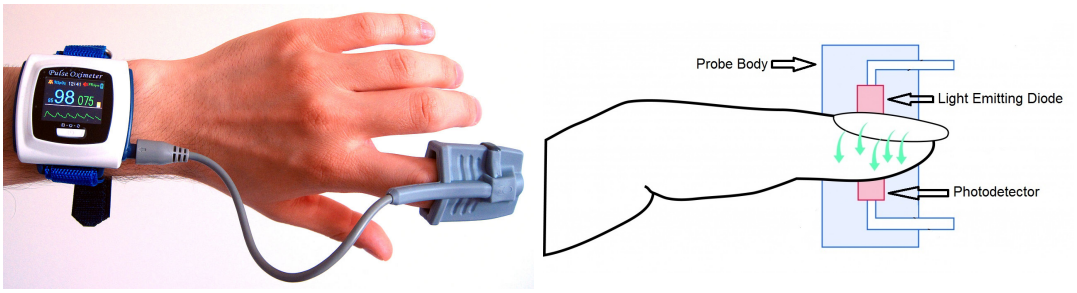


Fig. 2.3 PCG heart sounds [5].

2.2.3 Pulse Oximetry

Pulse oximetry is a non-invasive method to measure the oxygen saturation level of blood in our body and estimates the heart rate. It is widely used due to its easy and relatively comfortable setup. A pulse oximeter is a small microprocessor unit comprising a peripheral probe and a display monitor, shown in Fig. 2.4 (a). The probe which consists of a photodetector, and two light emitting diodes can be clipped to a finger, earlobe or foot of a subject, and the estimated heart rate can be displayed on the monitor screen [28]. The two light emitting diodes emit light of different wavelengths, which is absorbed by the tissues, and the amount of absorption is determined by the photodetector [37] (see Fig. 2.4 (b)). This method is easy to use but there are some limitations that affect its accuracy. Examples include: when a subject is shivering causing the occurrence of movement artifacts; a bright light such as an operating room lamp may cause light interference; electromagnetic radiation such as emitted from



(a) (b)

Fig. 2.4 a) A wrist mounted remote sensor pulse oximeter with a probe [6].
b) A pulse oximeter non-invasively measures oxygen saturation by shining light through a finger.

an MRI (Magnetic Resonance Imaging) scanner also affect pulse oximetry; and nail polish on the subject's finger could result in incorrect measurements [38]. Although it is a comfortable way to measure the heart rate, there are some side effects, for example Ischemic pressure necrosis may occur if the probe is placed too tightly on the subject's finger, and prolonged use of the oximeter may result in skin irritation from the adhesive on the probe. In our experiments, we have used pulse oximeter Contec CMS50E [39] to measure the ground truth heart rate. We asked the participants to avoid the finger movement so that the risk of movement artifacts can be minimized. The effect of subject's finger movement noise can be viewed in Fig. 2.5.

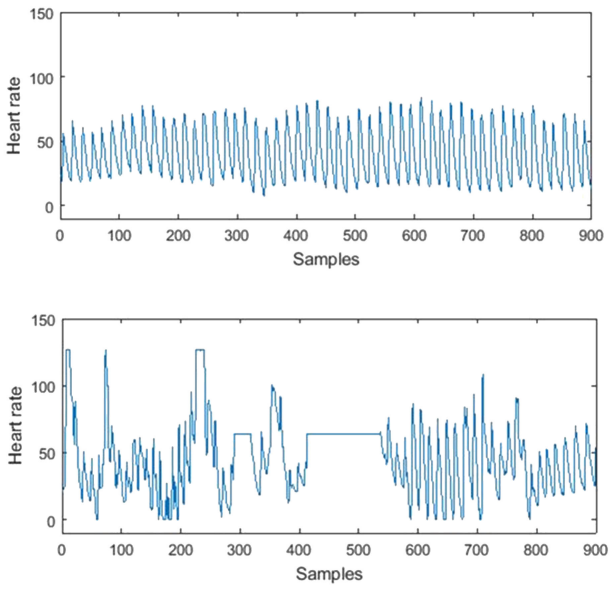


Fig. 2.5 Pulse signal cleanly obtained from pulse oximeter on the top and the pulse signal which is affected by the subject's finger movement on the bottom.

A number of fitness trackers are also available in the market that works on the principles of pulse oximetry to estimate the heart rate. One of those trackers is Fitbit [40], which uses an optical sensor to shine a light on our skin and measures the rate at which the blood is pumped to reveal the passing of pulse through it. Such fitness trackers are lightweight, non-invasive, comfortable, and less expensive, but they are not very accurate specifically while we are doing an exercise.

2.2.4 Electrocardiogram (ECG)

ECG is the current gold standard for the accurate measurement of heart rate [36]. As mentioned earlier, the flow of blood through the heart does not happen automatically but is driven by contractions of the heart muscle which are triggered by electrical impulses. This electrical activity of the heart with respect to each heartbeat is measured by ECG. Electrodes are placed at different locations of the subject's body such as limbs and chest to monitor such electrical activity over a period of time, and the output can be seen in the form of an electrical signal display on the monitor or printed on the grid paper. Fig. 2.6 presents the basic structure of the electrical signal obtained by the process of ECG.

The P wave is created by the contraction (or depolarization) of the left and right Atriums, QRS complex belongs to contraction of the left and right Ventricles, and T

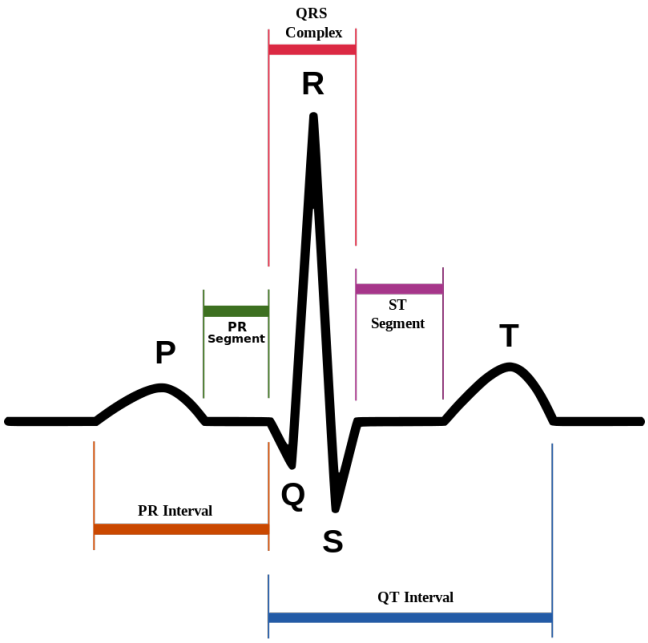


Fig. 2.6 Schematic diagram of normal sinus rhythm for a human heart as seen on ECG [7].

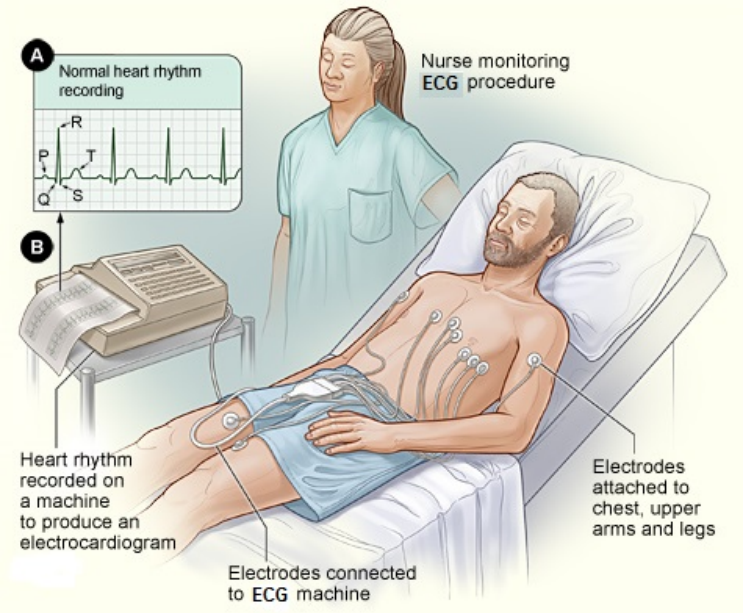


Fig. 2.7 ECG setup and ECG trace [8].

wave represents relaxation (or re-polarization) of both left and right Ventricles. The relaxation (or re-polarization) of left and right Atrium also occurs but the signal is obscured by the large QRS complex [41]. When reading the ECG signal, the peak R counts as one beat and the number of R peaks in the signal over a period of one minute calculates the heart rate. The limitations of ECG are that it is quite uncomfortable as the subject must undress and wear several electrodes and associated cables (see Fig. 2.7). It also restricts movements and is not applicable to outdoor environments as the subject performs walking or sports activities. ECG equipment is also very expensive [42] and is mostly used by healthcare professionals only at the clinical practices, and is therefore not affordable to install at home for daily use.

2.3 Non-Contact Heart Rate Measurement

We have discussed the heart rate measurement methods in practice in the previous section 2.2. As described above, they are accurate and efficient but there are certain limitations associated with them and they are quite uncomfortable with the subject. It is therefore required to find contact-free ways for the estimation of the heart rate and it has been an interesting area for the researchers to work on over the past decade. In this section, we will describe the most popular and/or the most promising contact-free heart rate measurement methods, discuss their performance accuracy, and identify their limitations.

Photoplethysmography (PPG), introduced in 1937 [43], is the main idea of most of the previous work related to the non-contact heart rate measurement. The term photoplethysmography is composed of two words 'photo' and 'plethysmography'. Plethysmography refers to a Greek word 'plethymos' which means increase, and describes the measurement of volumetric changes in different parts of the body [44]. These changes are caused by the change of blood volume triggered as a result of a specific event occurring in a body. The word 'photo' refers to the use of light for plethysmography. This phenomenon states that when a heart beats, it produces a pulse wave called a cardiovascular pulse or the blood volume pulse (BVP), which travels through the whole body's vascular system. When this wave reaches the face, it leaves a small variation in the volume of blood, which causes the skin color to change [45]. The naked eye cannot view such subtle changes but we can measure the intensity of the absorbed or reflected light by means of images taken from a camera sensor.

The cameras used in the previous work related to pulse extraction from videos are the thermal camera, CCD camera, and a simple computer webcam. The face of the subject is focussed by the sensors of each of these cameras.

2.3.1 Non-Contact Heart Rate Measurement Using Thermal Camera

Non-contact heart rate measurement was first introduced by Pavlidis in 2003 [46], and later demonstrated by analyzing thermal videos of the front face [9, 47]. They uncover the information of blood flow, cardiac pulse, and breath rate from a bioheat modeling of facial imagery using a sensitive thermal camera. Thermal imaging is a passive modality, as it depends on the heat radiations emitted from different parts of the subject's body. Fig. 2.8 shows an overview of their pulse measurement methodology. The face was targeted as the region of interest (ROI) because it is the most sensitive and exposed part of a body. The ROI was further narrowed down to the superficial blood vessel, which they assumed to be the best-suited area for the extraction of the strongest thermal signal. The face is always in constant motion and a robust tracking technique is required to accommodate such movements. They proposed a tandem tracker (TT) to track the face from the thermal image and the measurement tracker (MT) to detect a specific vessel tissue within the tracked face. Fourier transform is applied to the extracted thermal signal to obtain dominant pulse frequency. In order to minimize the effect of noise from the pulse signal, adaptive estimation function is used, which takes into account a specific threshold in each successive heart rate measurement [9, 47]. The foundation for contact-free heart rate measurement was laid

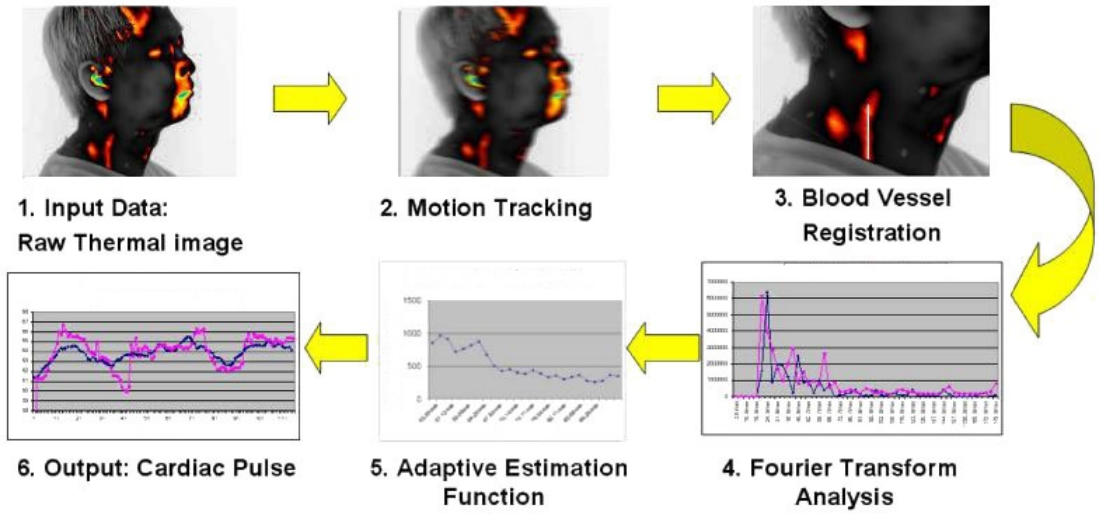


Fig. 2.8 An overview of non-contact heart rate estimation using a thermal camera [9].

down as they reported a good accuracy of 88% of their novel methodology but still the need for improvement in the face tracking was addressed by them. The dataset was collected in a controlled lighting condition and little subject movement was permitted, indicating that more noise could be expected when their method is applied in realistic capturing scenarios, which could affect the performance. The use of an expensive thermal camera for the data acquisition also raises concerns.

Gatto [48] also used a thermal camera to measure the heart rate by monitoring the flow of blood on the forehead area and analyzed the fluctuations in the thermal energy released with respect to each heartbeat. They used Autoregressive Moving Average (ARMA) models [49] to extract heartbeat information from the thermal video signal. Results obtained are quite satisfactory but there are some limitations that could affect the system performance. The thermal signal extracted from the video contains sufficient information to form a time series of inter-beat intervals but as the subject tilts the head in either direction the tracker fails to localize the region of interest, and therefore it requires more robust tracking techniques to avoid such noise. The difference between the subject's body temperature and the environmental temperature also generates noise and decreases the strength of the thermal signal extracted. The region of interest (forehead) can be easily covered by long hair of the subject, which can lead to an incorrect measurement of the heart rate. Finally, the use of an expensive thermal camera raises a question on its suitability for daily life activities.

2.3.2 Non-Contact Heart Rate Measurement Using CCD Camera

Takano and Ohta [50] measured the pulse rate using 30s (s means seconds) time-lapse images captured by means of a Charge Coupled Device (CCD) camera. CCD is a small circuit etched on to a silicone surface embedded in the digital cameras to transfer electrical charge to the memory in the form of a digital value [51]. They analyzed the variations in the average brightness of a subject's facial skin images. An ROI of 30×40 pixels was extracted from the subject's cheek images and a series of processing steps were performed on the average brightness of each image, which included a first-order derivative, a low pass filter of 2 Hz, and Auto-Regressive spectral analysis. The heart rate peak was chosen as the highest peak in the auto-regressive spectrum of the processed signal [50]. The results show high accuracy in their experimental setup with the controlled illumination settings but in realistic scenarios, where rapid light variations could be expected, the performance may degrade. Such noise elimination must be considered as a challenge. The size and cost of a CCD camera also make this method less applicable to the daily life activities.

De Haan and Jeanne [26] highlighted the problems of subject's motion and illumination noise, while measuring the heart rate using a CCD camera, and presented a chrominance-based remote pulse rate method to overcome such problems. The RGB color channels were extracted from the ROI and normalized by dividing their samples by their mean over a specific period of time, which must cover at least one heartbeat time interval. The combination of red and green color channels made a pulse signal. This worked well enough for subject's motion noise but the illumination variation was still present in the signal. The specular reflection component, which was directly reflected from the skin surface, was eliminated by using chrominance (color difference) signals. They proposed different algorithms by combining different proportions of RGB color channels and were tested on stationary and moving subjects. The results show better accuracy and signal to noise (SNR) ratio when compared to other state-of-the-art heart rate estimation methods. The use of CCD camera makes this method non-affordable to be implemented for the real-life applications.

2.4 Non-Contact Heart Rate Measurement Using an Ordinary Webcam

The performance of estimating the heart rate from human face videos using a thermal camera and a CCD camera was found quite accurate under controlled conditions. The

performance could be affected by the subject movement and illumination variations, and the resulting noise in the extracted pulse signal can be considered as the main problem in the accurate measurement of heart rate from facial imagery. As the CCD and thermal cameras can be very expensive, in the past few years, researchers have reported on the estimation of pulse rate using an inexpensive computer webcams [10–21, 24, 25, 52].

2.4.1 Color Based Methods

A simple webcam was used by Poh *et al.* [10] in 2010 for the first time to estimate the heart rate from human face videos. They presented a novel method for the non-contact automated heart rate measurement using the facial video imaging and blind source separation (BSS). This novel approach was applied to the color images of a human face tracked by an automatic face tracker, and, by decomposing the color channels into independent components using blind source separation, estimated the heart rate. The problems of non-automation, motion artifacts, and expensive sensing device for the estimation of the heart rate were addressed and an automated, motion-tolerant, and cheaper solution was proposed. Fig. 2.9 shows an overview of their method. Color videos of 12 subjects (10 males, 2 females) of different skin origin (Asians, Caucasians, and Africans) and age group (18-31 years) were recorded using a basic webcam embedded in a laptop at the rate of 15 frames per second (fps). The faces from the video were tracked using the OpenCV (Open Computer Vision) face detection

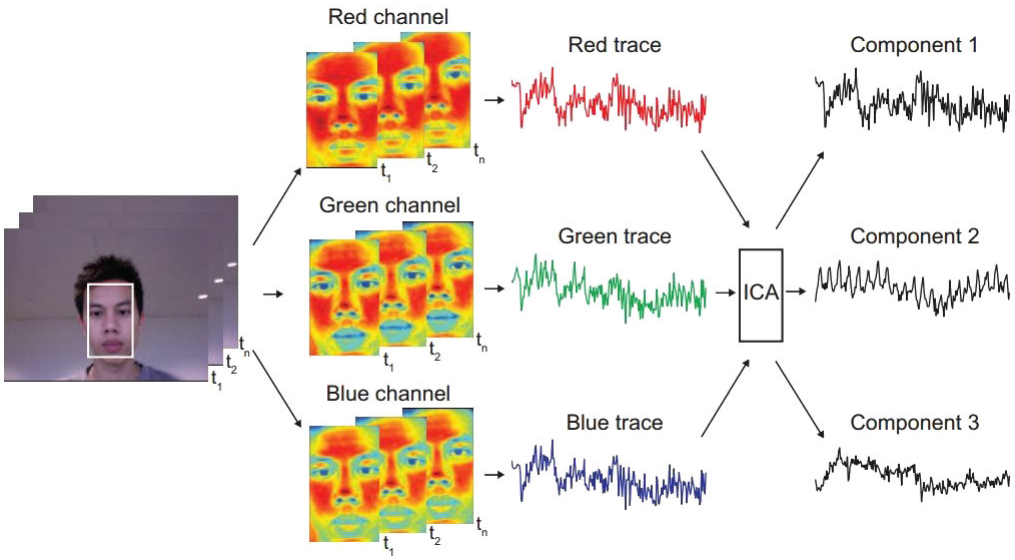


Fig. 2.9 An overview of the Poh *et al.* method for the estimation of heart rate from human face videos [10].

algorithm [53] which is based on the work of Viola and Jones [54]. The full face was chosen as the region of interest (ROI), which was then separated into three RGB (red, green, and blue) color channels and spatially averaged over all pixels in the ROI for each frame of the video to yield red, blue and green raw signals. The RGB raw signals were then normalized by subtracting each sample of a signal from the mean of all samples and divided by the standard deviation of all samples. The normalized raw signals were the mixture of a pulse signal and other noise sources. In order to separate the pulse signal, the three RGB signals are decomposed into three independent source components using Independent Component Analysis (ICA) technique of blind source separation [55]. Joint approximate diagonalization of eigenmatrices (JADE) algorithm of ICA, developed by Cardoso [56] was used for this purpose. Conventionally, the ICA returns components in random order and does not reveal any information relating to identification of the expected pulse signal among all of three extracted source components. Poh *et al.* [10] suggested that the second component typically contained a strong plethysmographic signal and was therefore selected as the desired pulse component for the purpose of automation. The fast Fourier transform (FFT) was then used to obtain the power spectrum of the desired pulse component and the frequency corresponding to the highest peak, within a valid frequency range, was designated as the heart rate frequency. The frequency band was set between [0.75-4] Hz, which can cover a wide range of heart rates i.e. [45-240] bpm. The ground truth heart rate was obtained by means of a blood volume pulse sensor clipped to the subject's finger to evaluate the results. The experiments were conducted in two different states i.e. sitting still with no movement and with the slight movement permitted with the ambient sunlight as the only source of illumination in each case. In case of sitting still with no movement allowed, the results obtained before ICA had a mean absolute error of 2.79 bpm and a root mean square error (RMSE) of 6 bpm, which were improved to mean absolute error of 0.91bpm and RMSE of 2.29 bpm after the use of ICA. In case of slight movement, a mean absolute error of 8.16 bpm and RMSE of 19.36 bpm before ICA were improved to the mean absolute error of 2.44 bpm and RMSE of 4.63 bpm after ICA. The results obtained were reasonably accurate, but there are some limitations that need improvement. The experiments were conducted indoors with a varying amount of sunlight as the only source of illumination, the performance may be affected if the videos are captured in the presence of fluorescent light bulbs. The face tracker used suffered from false positives (multiple faces detected) and false negatives (no face detected), and the previous image (face) was used in such cases to detect and compensate for these problems. In case of more realistic and challenging scenarios, such problems may result in completely different results. The second component was

always selected as the desired pulse component for the purpose of automation, but as ICA returns the output components randomly and in no specific order, the choice of selecting the second component may result in an increased overall error. ICA is not capable of removing all the noise from the RGB signals and the resultant independent source component may still contain noise, which could result in the prediction of the wrong pulse frequencies. The threshold of 12 bpm was set between two successive heart rate readings to overcome this problem, but further improvement is needed to eliminate noise from the desired pulse component.

Poh *et al.* [24] in 2011 modified their existing work [10] by introducing additional parameters to further refine the extracted pulse signal from the recorded face videos. The raw color traces were detrended in the pre-ICA stage. The smoothness priors approach [57] with the smoothing parameter $\lambda = 10$ (cutoff frequency of 0.89 Hz) was used for detrending. The ICA source component with the highest peak in its FFT power spectrum was selected for the expected pulse component. This component was then smoothed by a five-point moving average filter and bandpass filter (128-point Hamming window, 0.7-4 Hz). The limitations of the face detector to tolerate head motions and illumination variations were reported, which require further improvement.

Lewandowska *et al.* [52] measured the heart rate from human face videos using a webcam by applying Principle Component Analysis (PCA) [58] to the color channels. PCA reduces the computational complexity as compared to ICA. All experiments were performed indoors in daylight and with very limited subject movement. The 30 seconds facial videos of 10 subjects were recorded using a webcam at 20 fps and the ground truth ECG signal was obtained using electrocardiograph. Two different ROI's, which includes face and forehead, were extracted from each frame of the color video and decomposed into three RGB channels. The proposed algorithm was tested on a combination of different color channels: RGB, RG, GB, and RB. The ICA and PCA were both performed on different combinations of color. The power spectral density (PSD) of the desired pulse components was obtained and similar results were found for different ROI's (face and forehead) when compared to the ground truth. The RG color channel combination produced better results. The combination of different color channels was found effective for the estimation of pulse rate using a webcam. The reduction of color channels from three (RGB) to two (RG) decreases the computational time but resulted in an increase of noise in the pulse signal. The performance of this simple algorithm was quite accurate for a still subject but if applied to a more noisy and moving subject in different illumination setup could be a challenge. The problems of motion artifacts and lighting noise require a more robust way to measure the heart rate. The selection of forehead for the ROI can be affected by the subject's hair or

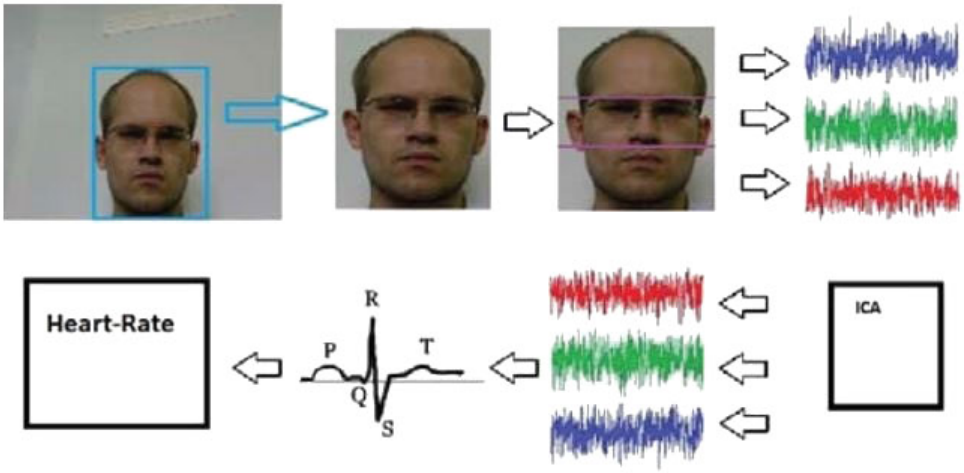


Fig. 2.10 An overview of the Pursche *et al.* method for the estimation of heart rate from human face videos [11].

head covering, it is therefore not the best suitable area for the heart rate measurement to choose from the facial image.

Pursche *et al.* [11] also investigated the contact-free heart rate measurement from human faces using an ordinary webcam. They have extracted three ROIs from the face which includes the forehead, an area around the eyes and nose, and mouth area. Color videos of 2 minutes were recorded using a webcam at 30 fps under the controlled experimental setup with indirect sunlight as the only source of illumination. The subject movement was also restricted. They reported noise in the RGB color channels and found ICA as the best solution to remove such noise. Two algorithms were implemented to estimate heart rate from the time signals. First, a peak detection algorithm was applied to the time signals to find peaks in the signal, and by counting those peaks, obtained the heart rate. Secondly, Fourier transform of the time signal was calculated to obtain the related power spectrum, and the highest power in the spectrum was nominated as the heart rate frequency. Fig 2.10 shows the overview of their method. They have reported that the performance of FFT was more effective than the peak detection algorithm. The results obtained were quite satisfactory but in controlled conditions, and more improvement would be required in the measurement of heart rate in realistic scenarios where illumination variations and subject movement such as facial expressions can be expected. The real-time estimate of heart rate was not possible with their experimental system, but it could be automated by transferring their code on to the hardware and firmware.

Kwon *et al.* [25] measured the heart rate using the built-in camera of a smartphone. They recorded a color video of 10 subjects using a front facing-camera of an iPhone

4 at the speed of 29.99 fps in the presence of natural sunlight as the only source of illumination. Subjects were told to stare on to the middle of the phone screen and remain still during the recordings. They implemented Poh *et al.* [10] methodology to measure the heart rate. It was reported that instead of using the second source component at the post-ICA stage as the desired pulse component, the raw green channel can be used as the desired pulse signal. The green color signal contains a strong plethysmographic signal due to its high absorptivity feature [59]. The power spectrum of the raw green signal predicted the heart rate. The raw green signal was found cleaner than the second source ICA component. The results of the green color channel were reported the best under controlled experimental conditions but the variations of light and subject's movement may not generate better results. The RGB color sensor of the phone camera may not be able to pick up the clean RGB color channels due to such illumination and motion artifacts noise.

Lee *et al.* [12] addressed the problem of non-skin pixels of the face used in the estimation of heart rate from videos captured by means of a webcam. The face contains many meaningless areas which contain no information about the heart rate such as the eyes, eyebrows, beards, etc., and these areas affect the accuracy of the heart rate measurement system. The skin color classifier based on neural networks [60] was used

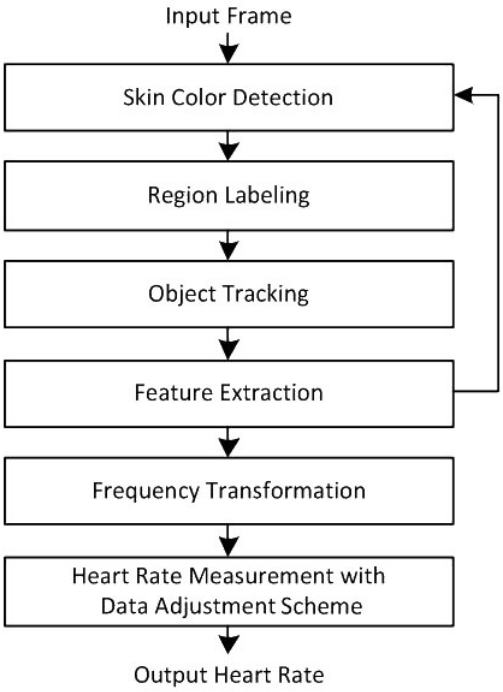


Fig. 2.11 Workflow diagram of the Lee *et al.* method for the estimation of heart rate from human face videos [12].

to identify skin color pixels from the images. They labeled the skin pixel regions on the face and tracked them using the mean-shift tracker [61, 62]. The RGB values of the ROI were calculated and the raw signals were analyzed using discrete-time Fourier transform (DFT) [63]. The heart rate was then measured using a data adjustment scheme and they proposed a data counting technique which transforms a small amount of data in the early stage of video capture resulting in measurement of the heart rate in a very short time period. Fig. 2.11 presents the workflow diagram of their proposed method. The results show the effectiveness of their proposed method but the problem of light effects and body movements while capturing videos was also highlighted. The possible solution to these two problems can be the color enhancement algorithms such as Retinex [64], and more robust face tracking algorithms which can accommodate the subject movements and stabilize the feature values.

Wei *et al.* [13] presented an automatic pulse rate measurement system based on the facial videos captured by means of a webcam. They used Laplacian Eigenmap (LE) [65] and inter-beat intervals (IBIs) along with some data processing steps to uncover the pulse signal from the three RGB channels of a webcam. LE is a manifold learning technique for uncovering the inner structure of data. It was addressed that according to the Beer-Lambert law, light absorbed by facial tissue varies non-linearly with distance,

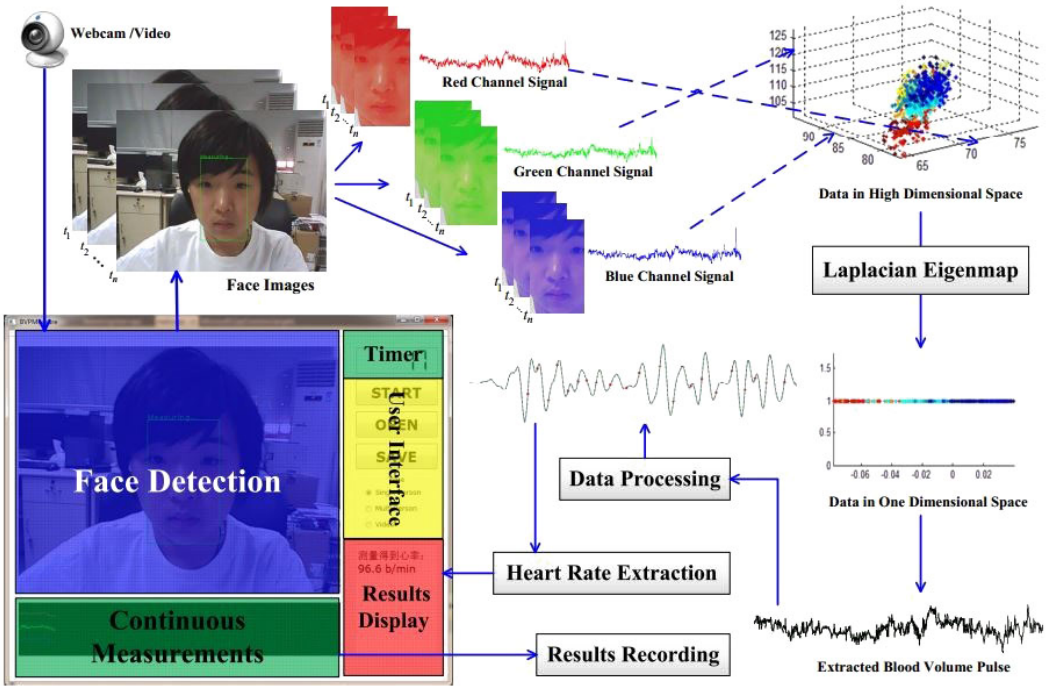


Fig. 2.12 Workflow diagram of the Wei *et al.* method for the estimation of heart rate from human face videos [13].

therefore the linear approaches of ICA and PCA could not extract the pulse signal cleanly. They proposed a novel non-linear approach of LE to extract the heart rate. Experiments were performed on 20 subjects and an ordinary webcam was used to capture a 30 seconds color video at the speed of 30 fps. Subjects were advised to sit still with very little movement allowed. Data was collected in two lighting conditions. For the first set of data, natural sunlight was used and a spotlight was used for the other dataset. Fig. 2.12 describe the different processing steps of their method. The OpenCV face detector [53] which is based on the work of Viola and Jones [54] was used to identify the face in the images, and the RGB color channels were extracted. The RGB values were averaged for all pixels in each frame of the video and were supplied to LE as an input, and the algorithm output the mapping data in a one-dimensional space. The one-dimensional output of LE, which should contain the desired pulse signal, was processed through some additional processing steps to smoothen it and estimate the heart rate. The processing steps included moving average filter, band-pass filtering, and IBIs. Results presented showed a better performance as compared to the use of other dimensionality reduction methods for the estimation of the heart rate. The experimental setup would need to be more challenging so that their method can be assessed in realistic capturing scenarios. The main problems of heart rate estimation i.e illumination variations and subject's movement still remain unsolved.

Wu *et al.* [14] revealed subtle changes in the skin color of human subjects using Eulerian Video Magnification, and measured the heart rate from facial videos. Fig. 2.13 shows the workflow diagram of the system. The system takes a color video sequence as an input and spatially decomposes it into different frequency bands. Temporal filtering was applied to all frequency bands depending on the application needs. The filtered

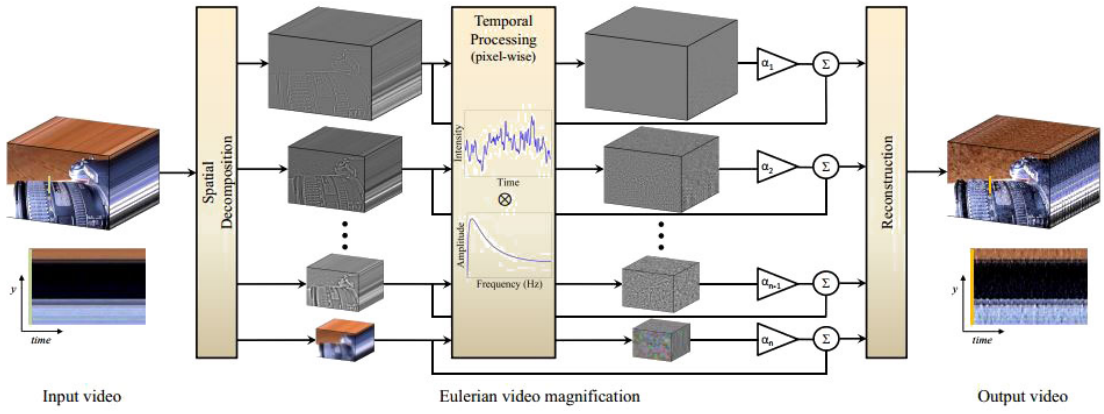


Fig. 2.13 Workflow diagram of the Wu *et al.* method for the estimation of pulse rate using Eulerian Video Magnification [14].

spatial bands were then amplified by a factor (α). The magnified signal was added back to the original signal to produce the output video. The variations in the skin color can be viewed in the output video in real time. This method was implemented on a dataset of largely stationary subjects which includes a sleeping baby, a motionless face, a still wrist etc., which means that video magnification cannot be applicable to real-life scenarios where motion noise can be expected. The variations in the amplitude of the desired signal of interest can be much smaller than the noise affecting the video in such real-case scenarios, it is therefore expected that the extraction of a clean desired signal is not possible.

Yu *et al.* [15] presented modifications in the Poh *et al.* [15] method by introducing the use of Short-Time Fourier Transform (STFT) [66]. The STFT can be useful in the cases where rapid heart rate variations can be expected. In order to identify the changing heart rate patterns, temporal information is required to extract the heart rate more accurately. The heart rate of a subject who was cycling during the experiments was analyzed in three different phases. In each phase, an increased effort was applied to the cycling activity to vary the heart rate. Fig. 2.14 shows the workflow diagram of the system. Results showed improvement in the heart rate measurement after the application of STFT. It was found that STFT provides better temporal information than the Fourier transform, and is therefore useful in the situations where rapid pulse rate variations are expected at regular intervals. The Fourier transforms gives the frequencies over a given time period and does not show the time localization. The

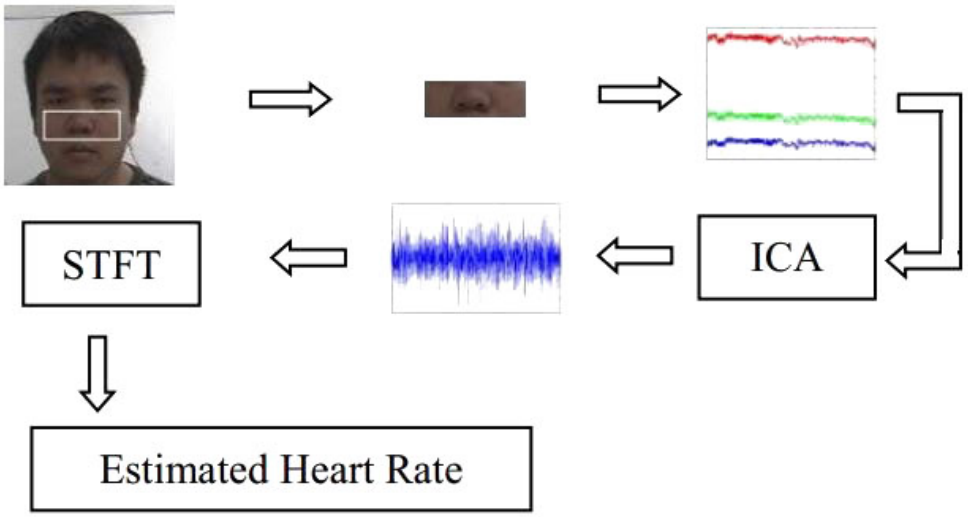


Fig. 2.14 Workflow diagram of the Yu *et al.* method for the estimation of pulse rate using STFT [15].

performance comparison of STFT and Fourier transform for a non-rapidly changing heart rate scenarios needs to be investigated.

Monkaresi *et al.* [16] reported a machine learning approach for the measurement of contactless pulse rate measurement using a simple webcam. The Poh *et al.* [10] method was modified with the advancement of machine learning algorithms to measure the heart rate in controlled and naturalistic situations. The workflow diagrams of both these methods are shown in Fig. 2.15. Poh *et al.* [10] tracked the skin color pixels changes on the face and their method worked very well under controlled capturing scenarios. The performance gets degraded for a subject's movement and lighting variations conditions. Monkaresi*et al.* [16] made the experimental setup more challenging by capturing videos under more subject motion. Fig. 2.15 (b) shows the workflow diagram of their proposed method. The face was tracked from the recorded video using OpenCV face detector [54] and the RGB values were calculated from the ROI. The RGB signals are normalized and detrended before being decomposed into three independent source

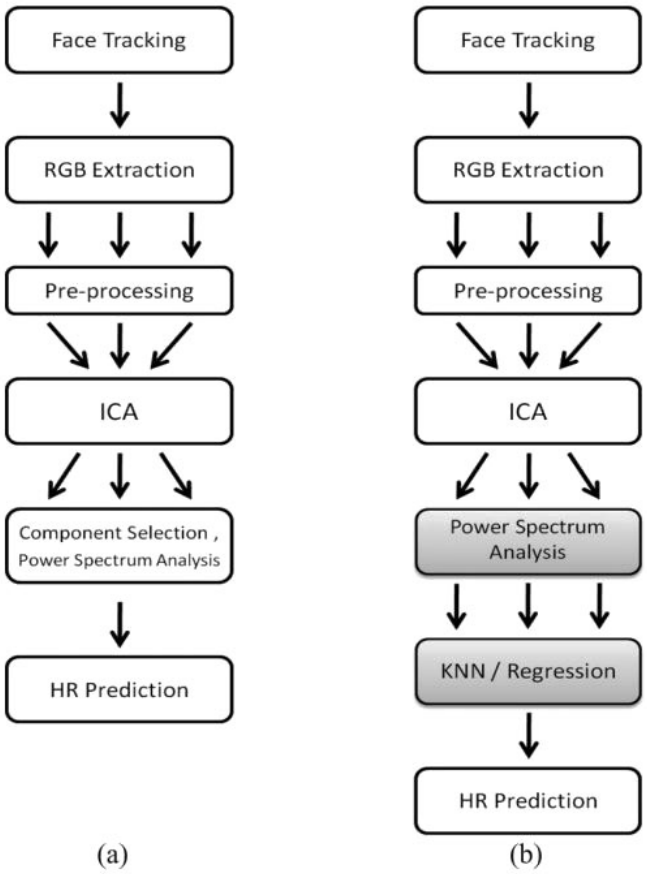


Fig. 2.15 Comparison of the workflow diagram of the (a) Poh *et al.* [10] method vs (b) Monkaresi *et al.* [16] method for the estimation of pulse rate.

components using the JADE algorithm of ICA developed by Cardoso [56]. Unlike the Poh *et al.* [24] criterion, where the source component with the highest peak in its power frequency spectrum was chosen as the pulse component, the selection of pulse component among the three source components was performed by machine learning techniques. Two algorithms of machine learning which includes linear regression and k-nearest neighbor (kNN) were applied to the three source components to extract nine different features from them for the classification purpose. Linear regression tries to find a linear relationship between a dependent variable and one or more explanatory variables. The kNN classifier is based on the simple distance measure to match the training and testing instances and make the decision if the testing instance belongs to the same class as training instance [67]. The results obtained show that the kNN-based technique performed better than the linear regression algorithm for the selection of pulse component to estimate the heart rate. The limitation of this work is the generalization of the model, which cannot be achieved using a user-dependent model and the proposed model may behave differently if exposed to new users without training. The experiments were performed under controlled lighting conditions, it is, therefore, obvious for the performance to be degraded if the capturing conditions are more realistic and challenging.

Li *et al.* [17] addressed some of the existing problems of contact-free heart rate measurement using a webcam. The problems of illumination variations and subject's motion are highlighted. They reported that most of the existing methods perform accurately in controlled lighting conditions and limited subject's movement but if they are exposed to the more realistic scenarios, where illumination variations and motion noise is expected, their performance degrades significantly. They proposed a method based on the face tracking and Normalised Least Mean Square (NLMS) adaptive filtering [68] to overcome the influences of these problems. To compensate for the problem of rigid head movements, face tracking was used to localize the ROI. Viola-Jones face detector of OpenCV [54] was used to detect the face. Fig. 2.16 shows the workflow diagram of the system. Discriminative Response Map Fitting (DRMF) method [69] was used to identify landmarks on the face and a mask of ROI (an area around the mouth) was created in the first frame. They then employed Kanade-Lucas-Tomasi (KLT) algorithm [70] to keep tracking the ROI. The green channel of the ROI was chosen as the raw pulse signal. In order to rectify the effect of illumination variations, NLMS adaptive filter was employed on the raw signal. Temporal filters (which include the detrending filter, moving-average filter, and bandpass filter) are then applied to the signal to smoothen it and eliminate frequencies which fall outside the range of interest. Power Spectral Density (PSD) of the pulse signal was calculated

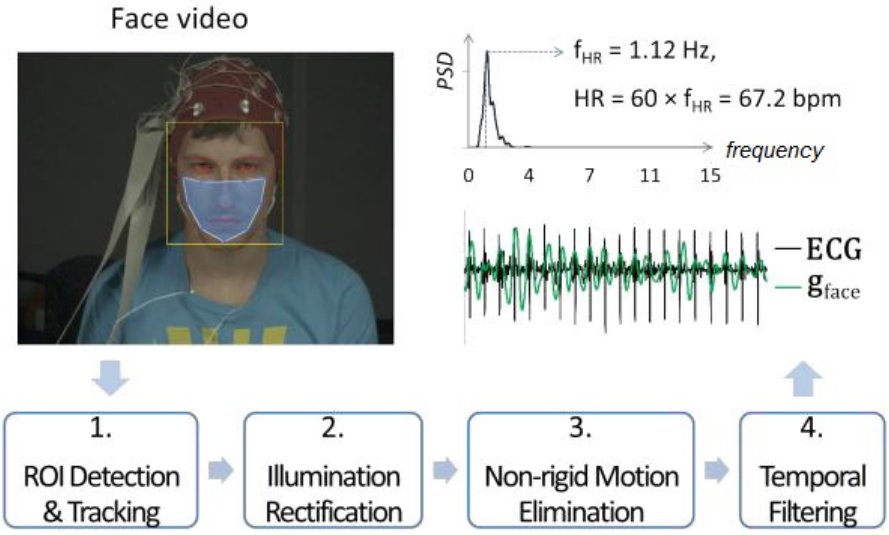


Fig. 2.16 Workflow diagram of the Li *et al.* method for the estimation of pulse rate [17].

using Welch's method [71] to find the heart rate frequency. The publicly available database of MAHNOB-HCI [72] was used to test their method, which contains color videos of different subjects recorded under HCI scenarios. The results show that their method substantially outperformed most of the previous methods. The limitation of their work is that it failed to track the ROI when the subject moved their head to a larger angle and therefore ROI tracking needs improvement to tolerate such non-rigid movements. The facial expressions of the subject caused enormous noise to the signal, for example, if the subject started to smile from a neutral phase, dramatic fluctuations were observed in the signal. The solution to this problem could be to segment the pulse signal and eliminate the segment affected by those uncertain changes and re-concatenate the signal but this could result in the loss of information. The MAHNOB dataset [72] was used in their experiments, which is still recorded in controlled lighting condition which consists of only two spotlight lamps. The subjects also moved very little and the facial expressions did not force them to create much movement.

Bousefsaf *et al.* [18] proposed a pulse estimation method based on the selection of most appropriate pixels of interest on the lightness criteria. They divided the captured face area into different sub-regions based on the face lightness distribution. The system selected the most relevant sub-regions automatically and group them together by evaluating their SNR (signal-to-noise ratio). Fig. 2.17 describes the overview of their proposed ROI segmentation method. Different experiments were conducted as the participants interacted with the computer to make the capturing scenario more

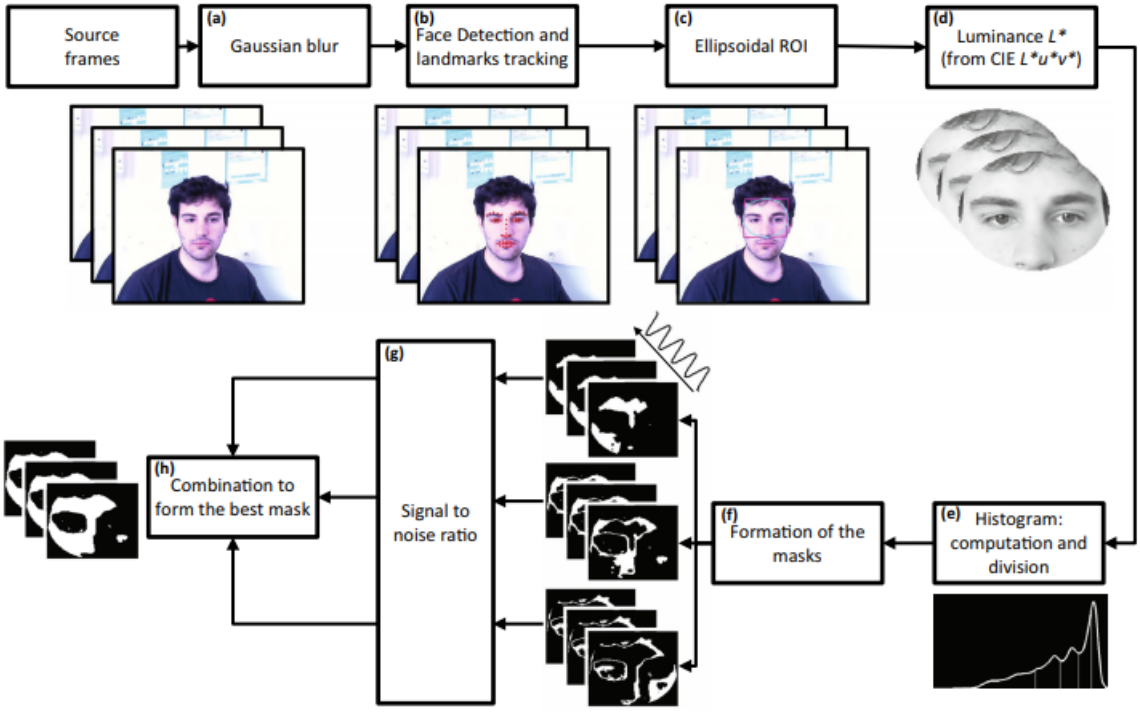


Fig. 2.17 Workflow diagram of the Bousefsaf *et al.* method for the estimation of pulse rate [18].

realistic and the results obtained were quite satisfactory with high accuracy and low heart rate error. The selection of the most suitable region of interest is very important for the accurate estimation of a pulse signal from facial imagery.

Po *et al.* [19] highlighted the importance of pulse signal extraction from human face videos but addressed the noise associated with the extracted pulse signal. They attempted to improve the quality of pulse signal by introducing the use of Block-based adaptive ROI approach to obtain the photoplethysmography details. Fig. 2.18 presents different processing steps of their proposed method. The face video was captured by a webcam and block-based spatial-temporal division was applied to it. The spatial-temporal quality distribution of the divided chunks of the video was then estimated based on their SNR (signal-to-noise ratio) feature. The mean-shift clustering was then used and adaptive thresholding in SNR maps are calculated to obtain the adaptive ROIs. The quality of the final pulse signal was therefore improved which accurately estimated the heart rate.

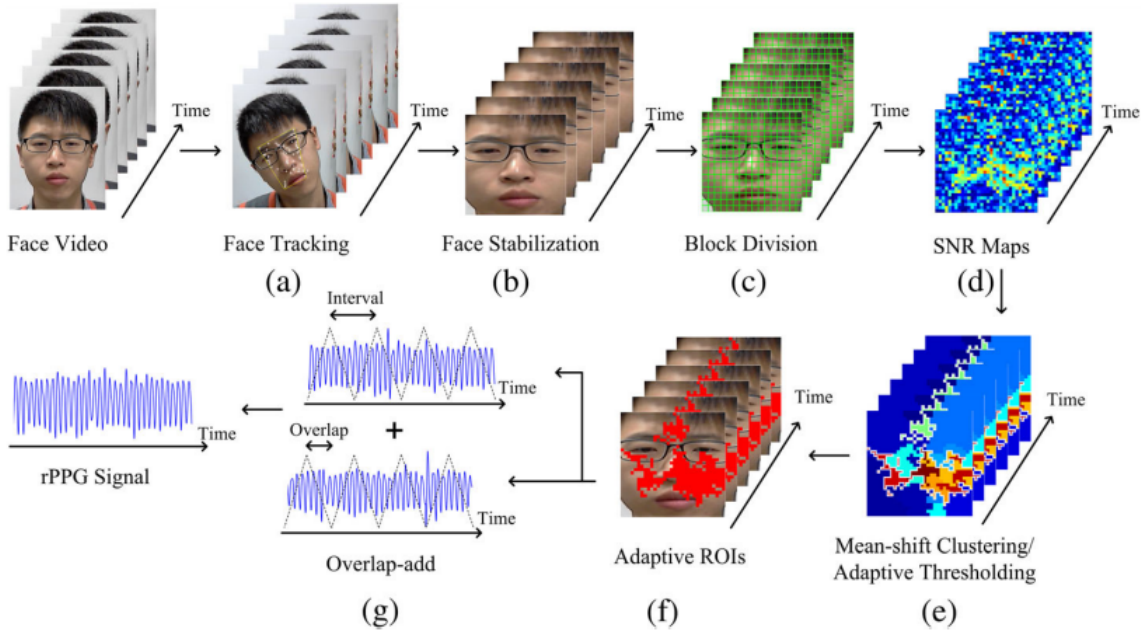


Fig. 2.18 Workflow diagram of the Po *et al.* method for the estimation of pulse rate [19].

2.4.2 Motion Based Methods

Heart rate can be estimated from the subtle head motion occurring due to blood flow with respect to every heartbeat. This concept was implemented by Balakrishnan *et al.* [20], who measured pulse rate and beat lengths from the recorded videos. The subject was assumed to be stationary to avoid noise caused by the subject's movement. The workflow diagram of their method is presented in Fig. 2.19. The face was detected using an OpenCV face detector [54], and the eyes area was eliminated from the region of interest to avoid eyes blinking noise. The feature points were selected within the ROI and tracked in each frame of the recorded video. The trajectories of the tracked features modeled the head motion and obtained the raw trajectory signals. These signals contained noise due to extraneous motions which resulted in out of range heart rate frequencies. Temporal filters were used to eliminate such unwanted frequencies. PCA [58] was applied to further reduce noise and extract the pulse source component. The source component with the clearest peak in the power spectrum was chosen as a pulse component. The heart rate was computed using the frequency corresponding to the highest power of the frequency domain of the pulse component. Results show that extracting the heart rate based on the head motion was quite effective but there are some limitations in their study as well. The varying amount of light and subject movement could affect the performance. It was suggested in their study that if we can

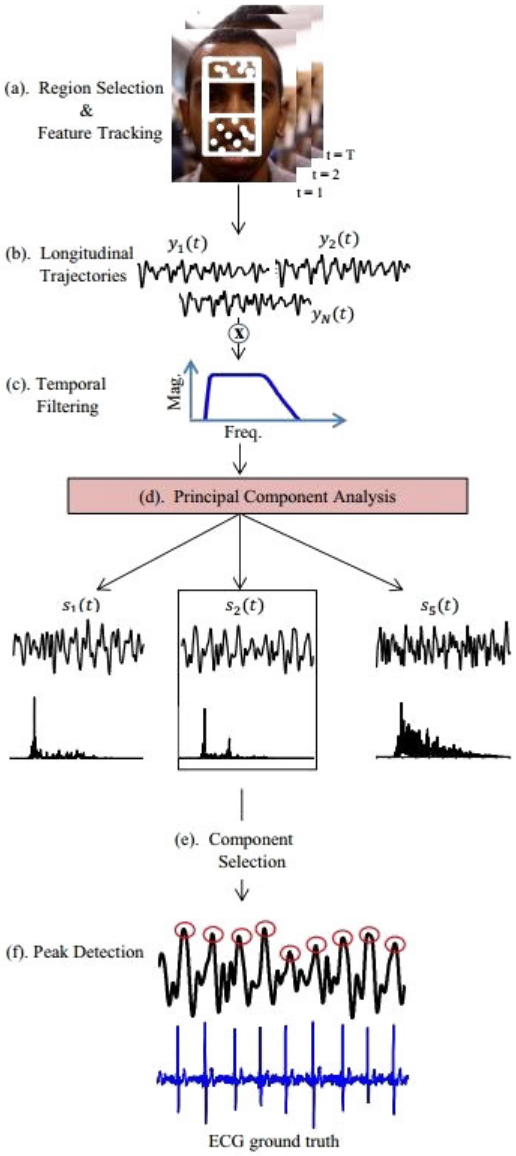


Fig. 2.19 Workflow diagram of the Balakrishnan *et al.* method for the estimation of pulse rate [20].

combine motion-based and color-based methods, it is possible to get a more robust way of measuring the heart rate as compared to using either one independently.

Haque *et al.* [21] addressed the problems of Balakrishnan *et al.* [20] motion-based heart rate measurement method. The subject motion was restricted in Balakrishnan *et al.* [20] study and no internal or external movement was allowed. The internal motion refers to the movement of facial features e.g. eyes, nose, and mouth, whereas external motion means movement of head pose. In realistic scenarios, such motions are expected and therefore more noisy frames may result in the incorrect measurement of heart rate. Haque *et al.* [21] introduced the Face Quality Assessment (FQA) method

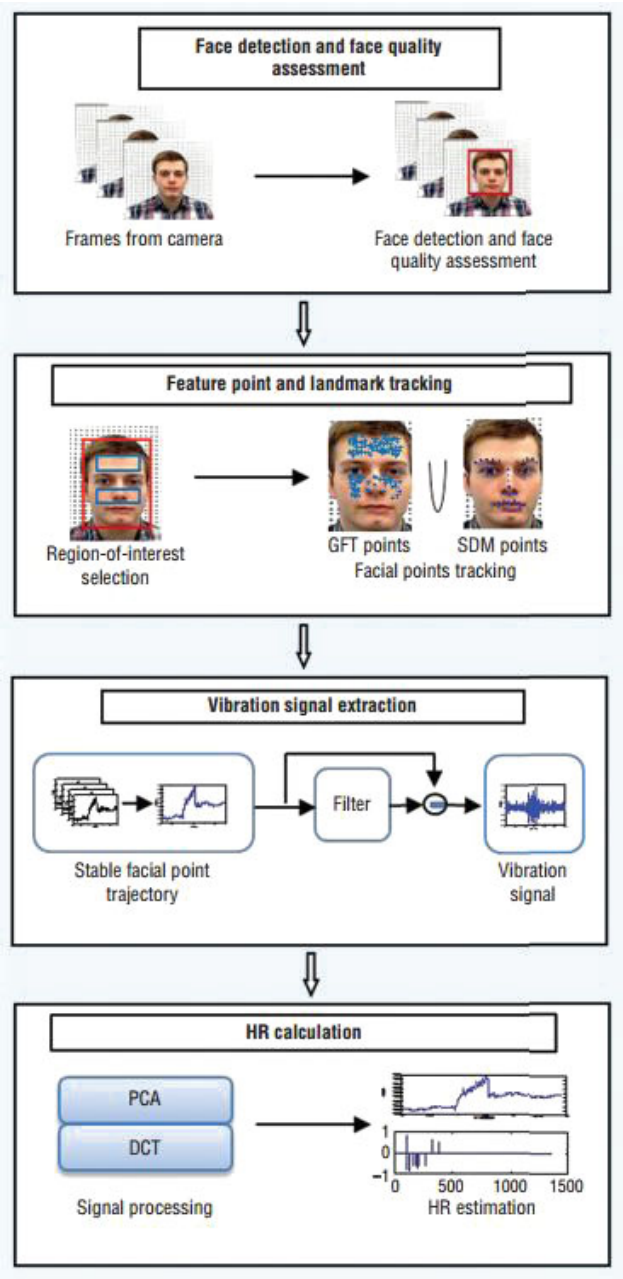


Fig. 2.20 Workflow diagram of the Haque *et al.* method for the estimation of pulse rate [21] from human face videos.

to overcome the aforementioned limitations of the previous study. This method ignores low-quality face frames from the recorded video so that they cannot contribute to the results. The Good Feature to Track (GFT) method [73] was used to track facial feature points from all images of the recorded video. Facial landmarks were also extracted using the supervised descent method (SDM) [74] and then combined with the GFT extracted feature points to track the motion trajectories, which allows for the better estimation of heart rate. Fig. 2.20 shows the overview of their method. The face

detection was performed using OpenCV face detector, which refers to the Viola and Jones's work [54]. The low-quality frames which contained a face were eliminated using an FQA method. The FQA takes into account four parameters which include resolution, brightness, sharpness, and out-of-plan face rotation, in order to decide if the image needs to be discarded. Facial feature points were tracked from the ROI and a trajectory was generated to record the head motions caused due to the heartbeat. Due to the limitations of GFT-based tracking, as it cannot tolerate much subject movement, SDM method to obtain facial landmarks is combined with the GFT tracking to get a better feature point motion trajectory signal. The vibration signal obtained was filtered using temporal filtering technique to eliminate the subject's facial features motion. PCA and DCT (discrete cosine transform) were applied to the vibration signal to extract the pulse source component and calculate heart rate. The FFT of the pulse component was also used to estimates the heart rate. This motion-based heart rate estimation method exhibits better accuracy as compared to other state-of-the-art methods but it is still dependant on little subject motion. For more realistic scenarios, the need for a robust method which can handle motion artifacts and lighting variations is required.

2.5 Existing Challenges in Contact-Free Heart Rate Measurement

In the previous sections 2.3 and 2.4, we have described different approaches for the contact-free heart rate measurement and highlighted their limitations. In this section, we will discuss the existing challenges and the key areas of the non-contact pulse estimation which requires further improvement.

2.5.1 Video Capture

Different types of capturing equipment were used for the acquisition of videos to be processed for the heart rate estimation which includes: thermal camera, CCD camera, smartphone camera, digital video camera, and webcam. They possess distinctive features such as speed, resolution, size, cost etc., which is our concern. The listed capturing devices are all different from each other such as thermal camera obtain thermal/infrared images, CCD camera is a modern-day digital camera which produces digital images, smartphone cameras are digital cameras with reduced size and highly portable, digital video cameras or camcorders produce videos from which we can

Table 2.1 Video capture devices used by previous researchers for the purpose of contact-free heart rate estimation.

Method	Capturing device	Speed(fps)	Resolution(pixels)
Garbey *et al.* [47]	MWIR thermal camera	30	640×480
Poh *et al.* [10]	built-in Macbook webcam	15	640×480
Lewandowska *et al.* [52]	Logitech 9000 webcam	20	640×480
Kwon *et al.* [25]	iphone 4 built-in camera	30	640×480
Yu *et al.* [15]	digital video camera	25	720×576
de Haan & Jeanne [26]	CCD camera	20	1024×752
Our study	Logitech C310 webcam	15	640×480

extract images, and webcam uses CMOS sensor to capture videos. They have different capturing speeds range from 15 fps to 61 fps, but there has been no conclusive evidence found that the camera speed would affect the estimated heart rate. However, the size and cost of the cameras were the main concerns and therefore thermal cameras, CCD cameras, and digital video cameras create the problem of unaffordability of those cameras for this specific application of pulse estimation. The webcam can be the cheapest and most affordable option to choose from all that sensing equipment. In our study, we have used Logitech webcam (C310) for video acquisition which has the following specifications: 1.3 Megapixels (mp) video capture capability, 24 Bit RGB with 3 channels (8 Bit per channel), 15 frames per second (fps) with a pixel resolution of 640×480 pixels. Table 2.1 list different types of cameras and their specifications used by previous researchers.

2.5.2 Experimental Conditions

The previous methods for the non-contact heart rate measurement performed quite well in controlled conditions but the performance may be affected if they are implemented in more realistic capturing scenarios. The lighting was controlled by allowing indirect sunlight as the only source of illumination [10, 24, 25, 52] and some researchers [11, 15, 16] allowed fixed fluorescent light bulbs as an additional light source. The scenarios in which the illumination variations occurs may affect the performance. The

Table 2.2 Experimental conditions of contact-free heart rate estimation methods.

Method	illumination source	Subject's movement
Poh *et al.* [10]	indirect sunlight only	no/slight movement
Lewandowska *et al.* [52]	indirect sunlight only	sit stationary
Kwon *et al.* [25]	indirect sunlight only	sit stationary
Wei *et al.* [13]	sunlight and spotlight	slight movement
de Haan & Jeanne [26]	studio spotlights	sit stationary
Pursche *et al.* [11]	sunlight and bulb light	no/slight movement
Our study	sunlight, bulb & projector's light	45°movement

second element of the controlled capturing conditions was the subject's movement, which was kept to stationary or little in most of the previous work. In a realistic scenario such illumination variations and subject's motion is very much expected, therefore, these two areas must be addressed and need further improvement. In our study, we have made the capturing scenario more challenging and realistic by adding more illumination variations and subject's motion. Table 2.2 presents the experimental setup of some of the previous work for the estimation of heart rate from human face videos.

2.5.3 Robustness of Face Tracker

One of the most important components of the contact-free heart rate measurement from human face videos is the robustness of the face tracker to track the faces from every frame of the recorded video. Even temporary loss of tracking creates a serious problem and can result in spikes in the pulse signal of the tracked ROI, which are falsely indicative of strong physiological responses. Table 2.3 shows the face trackers used by different researchers for the estimation of pulse rate from human imagery. In most of the previous work [10, 13, 16, 25], the OpenCV face detection method [53], which is based on the work of Viola and Jones [54] was employed to track the faces from the recorded videos. The problems of false positives (yielding multiple faces), false negatives (no face detected), and wide variation in how tight fitting the detected

Table 2.3 Face trackers used by different researchers for the purpose of contact-free heart rate measurement.

Method	Face Detector
Poh *et al.* [10]	OpenCV face detector
Kwon *et al.* [25]	OpenCV face detector
Wei *et al.* [13]	OpenCV face detector
Monkaresi *et al.* [16]	OpenCV face detector
Li *et al.* [17]	Combine OpenCV face detector and DRMF
Haque *et al.* [21]	Combine OpenCV face detector and FQA method
Our study	Saragih's face feature tracker [2]

face box is, limits the performance of OpenCV face detection algorithm and therefore result in loss of data frames containing the face. It finds coarse face locations as rectangles, which is not precise enough for pulse rate estimation because the non-skin pixels at corners of rectangles are always included. The situation gets even worse if the subject tilts the face in either direction. Some of the researchers [17, 21] used an additional techniques such as DRMF and FQA (in combination with OpenCV face detector) to make it more robust. In order to create robustness in the non-contact pulse estimation system, robustness in face tracking algorithm is very important. In our study, we introduced the use of Saragih's face feature tracker [2] to minimize the effect of face detection problems while estimating the contact-free heart rate. We have addressed the face detection problems, limitations of Viola & Jones face detector [54], and details of Saragih's face feature tracker [2] later in Chapter 3, section 3.2.1

2.5.4 ROI Selection

The full human face was used as a region of interest (ROI) by most of the previous researchers [10, 24, 25] to estimate the pulse signal from the recorded videos. The face is the most visible part of the human body. Some of the previous heart rate estimation methods [15, 20] reported that certain regions on the face such as eyes do not contain any skin pixels and therefore they are not useful to consider for the estimation of heart rate. Yu *et al.* [15] found that the human face contains regions with less or no blood

Table 2.4 Region of Interest (ROI) selected by previous methods for the purpose of contact-free heart rate measurement.

Method	Region of Interest (ROI)
Poh *et al.* [10]	Full face
Yu *et al.* [15]	Region between eyes and upper lip
Lee *et al.* [12]	Head/neck, inner arm, outer arm, and palm
Pursche *et al.* [11]	Forehead, area around eyes and nose, and mouth
Lewandowska *et al.* [52]	Full face and forehead
Kwon *et al.* [25]	Full face
Ourstudy	40×40 pixls cheeks

vessels such as eyes, hair, and nostrils, which do not contribute to accurate estimation of a pulse signal and result in noise. Balakrishnan *et al.* [20] remove eyes from the detected face so that the blinking noise do not affect the results. Lewandowska *et al.* [52] along with the frontal face selected the forehead as ROI, and they found almost similar results for both regions but the forehead can be covered by the long hair or head covering, which can affect the results. We have chosen 40×40 pixels area of the cheeks (2 centimetres approximately) from the faces as our ROI in our study for the measurement of heart rate. The selection choice of ROI as a cheek is discussed later in chapter 3, section 3.2.1. We also performed an experiment on the full face as ROI and compare it with 40×40 pixels cheek and reported results is section 3.4.2. Table 2.4 list the ROI used by different researchers for the measurement of heart rate.

2.5.5 Databases

In most of the previous work related to contact-free heart rate estimation from human face videos, the local datasets were created on a different number of subjects to test their methods. The only available public dataset is the MAHNOB-HCI database [72], which was used by Li *et al.* [17] and Haque *et al.* [21]. In our study, we have tested our proposed method on two different datasets created locally by us, and the publicly available MAHNOB-HCI database [72]. Table 2.5 presents the number of subjects and video clips used to test the different heart rate estimation methods.

Table 2.5 Datasets used by previous methods for the contact-free heart rate measurement.

Method	No. of Subjects	No. of Videos	Dataset
Poh *et al.* [10]	12	12	local
Yu *et al.* [15]	1	3	local
Lee *et al.* [12]	5	47	local
Haque *et al.* [21]	7	174	MAHNOB
Lewandowska *et al.* [52]	10	10	local
Kwon *et al.* [25]	10	10	local
Our study	16	60	local & MAHNOB

2.6 Human Biometric Recognition Based on Contact-Free Pulse Signals

In the previous sections 2.4 and 2.5, we have discussed in detail the previous work related to the contact-free heart rate measurement from human face videos using a simple webcam. In this section, we will present the human biometric recognition application of the non-contact heart rate estimation system. The study of Haque *et al.* [23], which is the only method presented in this regard will be described. The human biometric recognition part of our thesis (Chapter 4), highlights the effectiveness of the proposed contact-free pulse extraction system.

2.6.1 Human Biometrics

The word Biometrics is derived from two Greek words, "Bio" means *life* and "metric" refers *to measure* [75]. Human biometrics is the identification of a human based on his/her appearance or characteristics. Biometric recognition technology relates to the use of a computer database and algorithms, which automatically identifies a person based on their physical and behavioral characteristics [76]. The biometric identifiers are the distinctive and scalable characteristics of individuals used to label and describe them. These identifiers are classified into two types i.e. physical and behavioral. The

well-known biometrics traits of physical class identifiers include, fingerprint scan [76], retinal image [77], palm prints patterns [78], facial recognition [79], and hand vein scan [80]. The behavioral class identifiers consist of voice recognition [81], typing rhythm [82], hand-written signature [83], and human motion such as gait [80]. All these methods perform effectively and have potential benefits for use in security, crime prevention, and preventing fraud. However, there are some limitations in most of these biometric modalities in certain circumstances. For example, fingerprints can be altered in case of labourer over-use of hands with the passage of time, children's fingerprints also change with age, many people cannot replicate their written signatures in the same manner, hand geometry biometrics are not suitable for people with disabilities, voices can be altered to breach the voice recognition systems, and still pictures can be used to dodge the facial recognition modality of biometrics. The need for a more robust trait in the area of human biometrics is therefore required to overcome these limitations. The heartbeat signal is one possible alternative biometrics traits, which can be very difficult to forge. The pulse signal of every person varies from others due to change in size, position, and anatomy of the heart, chest configuration, and various other factors [84]. This area of human biometrics is investigated by the researchers in the past two decades and the results obtained are found quite satisfactory.

2.6.2 Human Biometrics Recognition Based on Heartbeat Signal

The heartbeat signal can be obtained by Electrocardiogram (ECG) [41] using electrical changes, Phonocardiogram (PCG) [5] using acoustic changes, and Photoplethysmogram (PPG) using changes in light absorption [45]. The human biometric system based on the pulse signal was first introduced by Biel *et al.* in 2001 [85]. They obtained an ECG signal for 20 subjects recorded at rest using the 12-lead ECG equipment (SIEMENS ECG). A set of temporal and amplitude features were extracted from the recorded ECG pulse signals. The number of selected features were reduced after analyzing the correlation matrix. Finally, the chosen features were classified using SIMCA (Soft Independent Modeling of Class Analogy) classifier and the subjects were identified by a multivariate analysis. Venkatesh *et al.* [22] presented the similar concept by adopting a different approach. They extracted 9 features from the ECG pulse signal and classified them using a combination of DTW (Dynamic Time Warping) and FLDA (Fisher's Linear Discriminant Analysis) classifiers in a two-stage classification technique. Fig. 2.21 shows the workflow diagram of their method. Chetana *et al.* [86] applied Radon transformation on the obtained ECG pulse signal. The feature vector was extracted by employing the standard Euclidean distance on the transformed Radon

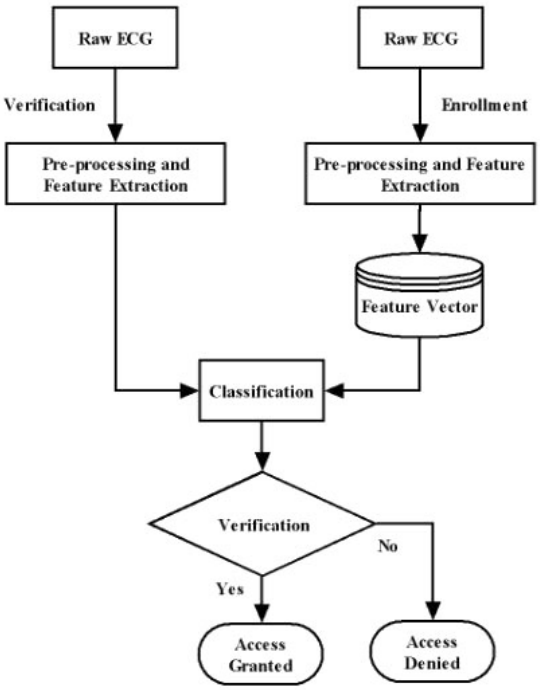


Fig. 2.21 Workflow diagram for the human biometric identification using an ECG based pulse signal [22].

image. The authentication of a person was obtained by computing the correlation coefficient between two such feature vectors. The review of some of the previous important methods presented on the ECG heartbeat based human authentication can be found in [87]. The touch-based ECG and PPG biometrics obtained more than 90 % accuracy. The common drawback of all of the previous human biometric methods based on ECG and PPG heartbeat signals is that they require sensory equipment to obtain a heartbeat signal and, were, therefore dependant on direct human contact. The need for a contact-free method for the heartbeat based biometrics recognition was hence much needed. The development of contact-free heart rate measurement methods from human faces using a webcam opened a way to make the human biometrics based on a heartbeat signal contact-free.

2.6.3 Contact-Free Human Biometric Recognition Based on Pulse Signal

The recent advancements in the contact-free heart rate measurement from human face videos using a webcam created the possibility to extract a pulse signal from the video and use it to identify subjects. In the previous sections, we have described most of the recent heart rate estimation methods. The extraction of a clean pulse signal is,

however, a challenge and requires more robustness. The only work done in the area of contact-free human biometric recognition based on a pulse signal extracted from human face videos that we are aware of was presented by Haque *et al.* [23].

Haque *et al.* [23] presented the contact-free heartbeat signal from facial imagery as a new biometric trait to recognize human subjects. Fig. 2.22 shows the workflow diagram of their proposed method. In the first stage, the heartbeat signal was extracted from the recorded face video of a subject. The video of the subject's face was recorded

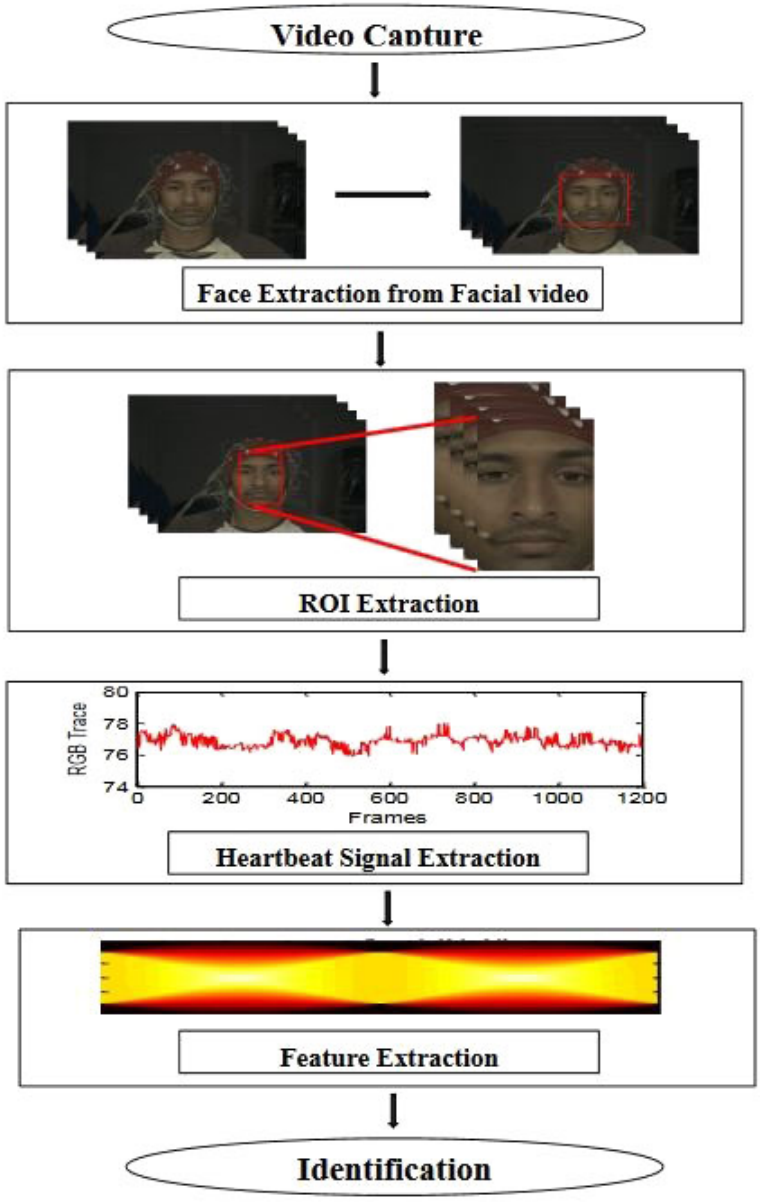


Fig. 2.22 Workflow diagram of the Haque *et al.* method for the human biometric recognition based on contact-free pulse signal [23].

using a simple webcam and the faces were tracked in each frame of the video using the face detector algorithm of [88]. They employed the method of Poh *et al.* [24] to extract the heartbeat signal from the recorded videos. The ROI was extracted from the face region in each frame, which was 60% width of the face area. The average color values of the red, green, and blue channels of the whole ROI were calculated for each frame of the video. The three RGB color traces were normalized and detrended to obtain three RGB signals. The RGB signals contained the pulse signal and the noise, therefore in order to eliminate noise, the JADE algorithm of ICA [56] was used to obtain three independent source components. The component with the highest peak in its FFT power spectrum was selected as the heartbeat component. The heartbeat signal obtained was not a clear representation of the pulse signal when it was compared to the ground truth ECG signal. This was due to the illumination noise and subject's head motion. A denoising filter was applied to the heartbeat signal by detecting the peak in the extracted pulse signal and discarding the outlying RGB traces. The extracted pulse signal was then transferred to the second stage of the biometric recognition module. The second stage was the extraction of features from the pulse signals and their classification to identify a human subject. The features were extracted from the heartbeat signals by first transforming them into the Radon images [89]. The Radon transform is an integral transform computing projections of an image matrix along specified directions and mostly used to reconstruct images from medical CT (computed tomography) scan. The Radon image can be generated by replicating the pulse signal to obtain a waterfall diagram. The number of replications must be equal to the number of frames in the video from which the pulse signal was extracted. The Radon transform [89] was applied on to the generated waterfall diagram to extract features. The features were extracted from the Radon transformed images for authentication by employing a pairwise distance method between every possible pair of pixels in the transformed image. The distance metric of Minkowski was used for this purpose. The feature vector obtained was then used for the authentication of human subjects. The identification process was performed using the decision tree method. They have used videos from the MAHNOB-HCI database [72] for training and testing of their method. The precision rate of 80.86% was achieved during the performance evaluation. The results show the effectiveness of their developed method, but the performance may be affected if it is exposed to a noisy dataset in which illumination changes, subject's motion artifacts, and heart rate variation can be expected. The features extracted from the Radon images of the pulse signals can be tested against other machine learning classification techniques to investigate their behavior and assess performance.

We need to further investigate the biometric recognition application of the contact-free pulse signal extraction from human face videos which can be improved by using cleaner pulse signals obtained from facial imagery and test the system against other machine learning techniques. We will present our findings and experiments about human biometric recognition application of the contact-free pulse extraction in chapter 4 of our thesis.

2.7 Summary

This chapter presented a literature review on different methods used for the measurement of heart rate of human subjects. The pulse rate extraction methods from human face videos using the thermal camera and a CCD camera are discussed. The contact-free heart rate measurement approaches using an ordinary webcam are described in detail and the terminologies used in them are discussed, which includes: data preprocessing and normalization, face tracking algorithms, ROI selection, Independent Component Analysis (ICA), Principle Component Analysis (PCA), Fourier transforms, adaptive and temporal filters, illumination rectification, non-rigid motion elimination and machine learning. The performance of the previous methods were critically analyzed and the limitations were highlighted. We described the existing challenges and the key areas of the non-contact pulse estimation which requires further improvement. The two main problems of most of the previous methods were addressed, which are illumination variations and subject's motion. We attempt to overcome these problems in our work, which will be explained in detail in chapter 3. The human biometrics recognition application of the contact-free heartbeat signal from human face videos is also described. We will present our work about the biometric recognition application based on contact-free pulse signal in chapter 4 of our thesis.

Chapter 3

Robust Contact-Free Heart Rate Measurement From Human Face Videos Using a Webcam

Heart rate measurement plays an important role in human health assessment and there have been a number of methods suggested in order to monitor it remotely with more ease and comfort. Contact-free heart rate measurement is one way to make it user-friendly and can also be used for covert surveillance. The performance of the previously developed methods is found to be efficient and accurate under controlled conditions. In realistic and more challenging scenarios, their performance is degraded, as each method has its limitations. Typically, the performance is dependent on controlled lighting and limited subject movement. More realistic situations require more robust contact-free ways to measure the heart rate. In this chapter, we will present a method that can overcome many of the problems of light reflection and subject's movements, while measuring the heart rate from human faces remotely using an ordinary webcam. We use Saragih face tracker [2] to track the faces from the recorded videos and find the robust mean of the skin pixel's color values of the selected ROI, and calculate the least-squares error optimal filter using our training dataset to estimate the heart rate and obtain a cleaner pulse signal. This technique has potential for advancing human biometric systems employing the pulse signal obtained. Results, presented later in this chapter, show that our proposed contact-free heart rate measurement method has significantly improved on existing methods. We also extracted a cleaner pulse signal which contributes to form a useful human biometric recognition system presented later in chapter 4 of our thesis.

3.1 Introduction

Heart rate refers to the number of times a human heart beats in one minute [32]. The continuous monitoring of a heart rate is very important to assess human health. Several different methods are currently in practice to measure the heart rate such as electrocardiogram (ECG), pulse oximetry and sphygmology. ECG [41] (sensors attached to subject's body detects electrical activity of the heart every time it beats) is considered to be the gold standard for accurately measuring the heart rate, but the need to wear a chest band and associated cables make it quite uncomfortable and restrict movement. Contact-free heart measurement is therefore very much in demand and hence it has been an interesting area for researchers to work on over the past years [9–21, 24–26, 47, 48, 50, 52]

3.1.1 Previous Work

The idea of non-contact heart rate measurement was first introduced by Pavlidis [46] in 2003 and later demonstrated by analyzing thermal videos of the front face [9, 47]. They employed a thermal camera as a sensing element to capture facial images and, based on the thermal signal information emitted from facial vessels, measured the heart rate [9, 47]. Since thermal cameras can be very expensive, in the past few years researchers have reported on the measurement of the heart rate using an inexpensive computer webcam. A number of color based [10, 24, 52, 11, 25, 12–19] and motion based [20, 21] methods were presented to accurately measure the heart rate from human facial imagery. The details of previous work related to contact-free heart rate estimation from human face videos using a webcam is presented in the literature review chapter 2 of our thesis.

3.1.2 Our Work

The previous methods for the measurement of heart rate from human face videos using a webcam performed efficiently, but only with controlled lighting and little subject's motion, and their performance degraded as they were exposed to more realistic and challenging capturing scenarios. The two major problems that affect the accuracy of the existing methods include illumination variations and movement of the subject, which we have addressed in our work and have attempted to overcome them. We use Saragih *et al.* [2] face tracker to track the faces from the recorded videos which provides more reliable extraction of a region of interest (ROI) than the simple face detection. We find the robust mean of the skin pixel's color values of the selected ROI

which can improve the measurement of the skin color in each frame, reducing the impact of facial hair, wrinkles, and lighting. In addition, we calculate the least-squares error optimal filter using our training dataset to estimate the heart rate more accurately from the measured color changes over time and obtain a clean pulse signal. We have trained and tested our system on the subjects in a more challenging environment and compared our results with the ground truth heart rate obtained by means of a pulse oximeter. We have obtained promising results while drawing a comparison of our proposed method with other state-of-the-art methods [10][24–26] using our own datasets and a publicly available dataset.

3.2 Method

In this section, we will describe the different processing steps of the proposed method for the measurement of contact-free heart rate measurement and the extraction of a pulse signal using the facial videos captured by means of a simple webcam. Fig. 3.1 shows the entire workflow of the proposed system.

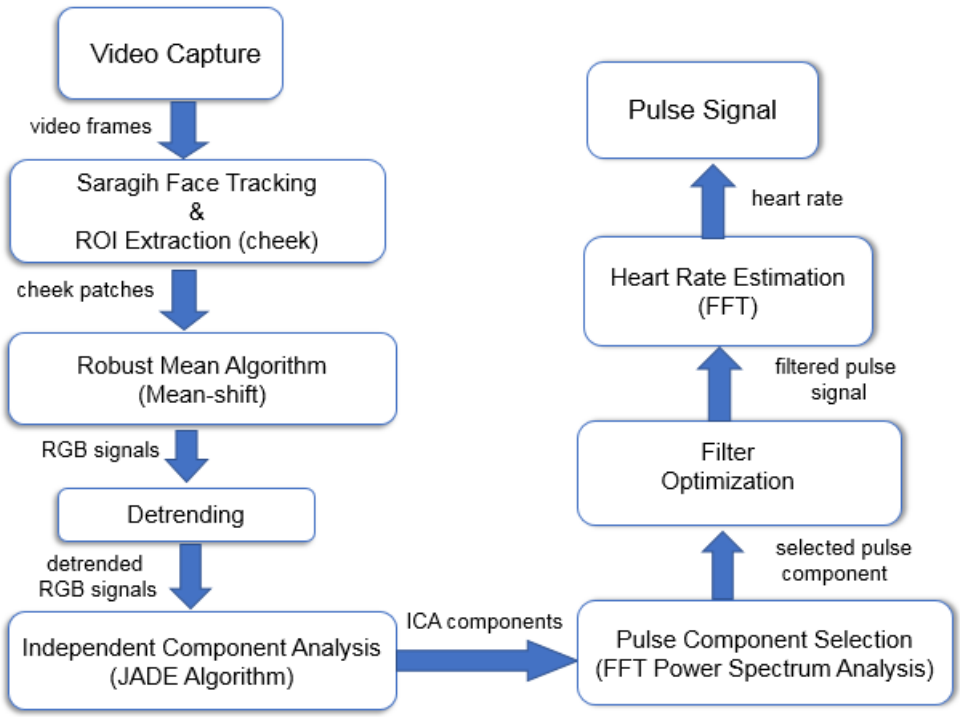


Fig. 3.1 Workflow of the method.

3.2.1 Face Tracking and ROI Extraction

The color of the facial skin changes with respect to each heartbeat, and the RGB color sensor of a webcam can detect subtle changes of blood volume in the facial area. We refer to this small reflected light color change as a plethysmographic or pulse signal that can be used to estimate the heart rate. In order to continuously monitor such variations of skin color pixels, we need to track the skin area of the face from the recorded videos.

Face tracking is an important component in order to get the heart rate measured accurately. In most of the previous work [10, 13, 17, 25], the OpenCV face detection method [53], which is based on the work of Viola and Jones [54] was employed to detect faces from the recorded videos.

The **Viola and Jones** face detection is a machine learning technique, where a cascade function is trained using the training data consisting of a number of positive and negative images. The cascade classifier is then used to detect objects of interest in the new or testing data presented to it. The OpenCV face detection algorithm [53] deals with the faces and a number of positive (which contains at least one face) and negative (images with no face) images are used to train the classifier. The Haar-like features (see Fig.3.2) are extracted from each image, and each feature is represented by a single value which can be computed by subtracting the sum of pixels under white rectangle from the sum of pixels under the black rectangle. In order to reduce the computational complexity, an integral images model was introduced by Viola and Jones [54], which simplifies the calculation using an operation based on the calculation of only four pixel values. During this process, a number of irrelevant features are also predicted by the algorithm. In order to minimize the number of features and to select only the critical visualize features from the large set of features, another classifier was introduced, which is based on the AdaBoost learning algorithm [90]. Finally, the classifiers are grouped into a cascade, which discards the background regions present in the image very quickly, while searching for the face-like regions. The problems

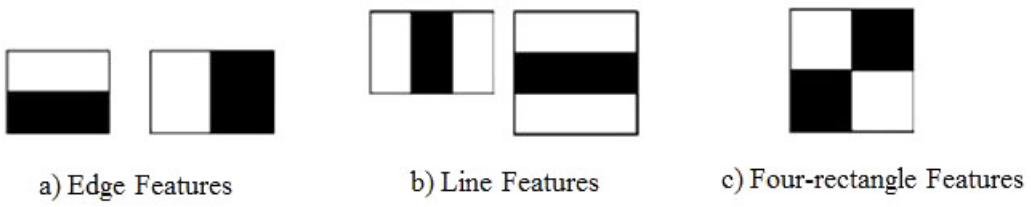


Fig. 3.2 The Haar-like features

of false positives (yielding multiple faces), false negatives (no face detected), and wide variation in how tight fitting the detected face box is, limits the performance of OpenCV face detection method [53] and therefore results in loss of data frames containing the face.

We have introduced the use of **Saragih's face feature tracker** [91] to track the faces from the recorded videos for heart rate estimation because it reduces the loss of frames for the detected face and gives more reliable and consistent ROI. It detects and tracks 68 points on different features of the face (eyes, nose, cheeks, and lips). Saragih *et al.* [2] proposed a principled optimization strategy where a non-parametric representation of the landmark distributions is maximized within a hierarchy of smoothed estimates. The resulting update equations are reminiscent of mean-shift [92] but with a subspace constraint placed on the shape's variability. They focussed on optimizing different deformable model fitting techniques. Deformable models refer to the collection of algorithms that extract distinct features from an image to model them into a certain class. A number of such different techniques were posed under a consistent probabilistic framework in which they were expected to present different parametric approximations which matches the true likelihood maps of the ground truth landmark locations. They then presented a new improved approximation approach which was based on a non-parametric representation that leads to an optimization in the form of subspace constraint mean-shift. The proposed method not only performed better than the previous optimization methods for generic face fitting such as ASM (Active Shape Model), CQF (Coined convex Quadratic Fitting), and GMM (Gaussian Mixture Model), but also resulted in a faster real-time computation. They performed a qualitative analysis of their face feature detection on the labeled faces in the Wild database [93]. The database contains images that have varying lighting, resolution, image noise and partial occlusion. The results show that Saragih's face detection exhibits a higher degree of robustness towards variations encountered in realistic facial image capture [2]. We have found in our experiments that if a participant can move their face up to 90 degrees each way then Saragih's face tracker is able to detect and track the face in the video but, at 90 degrees angle, only half a face can be visible to the camera sensors. Since our region of interest is a single cheek of a subject, a movement of the ROI up to 45 degrees each way can track it, otherwise we have to target both cheeks as a ROI to allow 90 degrees of freedom to move the face. The problem of motion artifacts is resolved to a considerable extent by using Saragih's face tracker algorithm, which allows participants to move their face freely to a certain degree while they were being recorded. Fig. 3.3 shows a set of frames from the tracked video as the subject moves his head and changes facial expressions, which give an impression

Fig. 3.3 A set of sample frames showing the robustness of Saragih's face tracker.

of how well the tracker copes with subject's motion. We have achieved 100% face detection rate from the recorded videos with no loss of frames containing the face, although we acknowledge this is only a relatively small sample of videos.

We have chosen the cheeks from the faces as our **region of interest (ROI)** for the measurement of heart rate because it tends to be a highly sensitive area of the face [94] which is visible to the camera sensors, and is least likely to be affected by a beard, hair, head covering, and glasses. It is also less affected by facial feature movements and we considered it to be the most stable area of the face to extract maximum skin region. We extract a 40×40 pixels ROI from the face and one of the points of the Saragih's face tracker was used to position it. We have set the ROI size to 40×40 pixels because we want to extract the pulse signal from a small patch of the face rather than the full face. The chosen ROI is an acceptable cheek capture measure for all sizes of faces

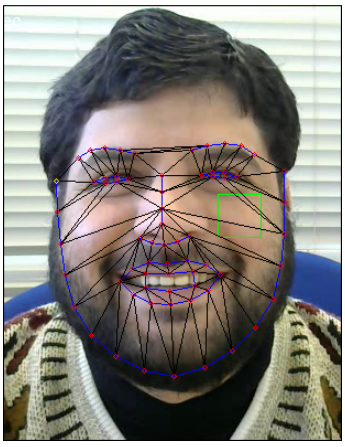


Fig. 3.4 A sample frame showing Saragih's Face Tracker (in blue/black) and ROI extraction (in green).

but could easily be scaled/rotated based on the eye location. Fig. 3.4 shows a sample frame of the video with the tracked face (in blue/black) and detected ROI (in green).

3.2.2 Spatial Averaging Using a Robust Mean Algorithm

We are interested in the RGB color intensities of the skin pixels in the extracted ROI in each frame of the recorded video, and therefore we have calculated the average value of each color channel for all pixels in the ROI of each frame. Since the ROI may be affected by illumination changes, shading, non-skin pixels in the ROI, and exposure adjustment of the webcam, long-term changes can be found in the averaged ROI values over the period of video capture. In order to compensate for these changes, we have introduced a robust mean algorithm that tries to take into account only the skin color pixels and ignores any non-skin pixels that are affected by lighting and other noisy sources and calculates the average of each color channel for only unaffected skin color pixel values in the ROI of each frame.

The robust mean algorithm is based on the Mean Shift algorithm [95]. The intuition behind this algorithm is that we will create a small window within the RGB color space and will move that window to the area of maximum pixel density (or a maximum number of pixels). Let us consider Fig. 3.5, which presents the distribution of pixel colors from the ROI (shown here in 2D for ease of illustration, but calculated in 3D RGB space in the actual algorithm). If we are given a circular region A indicated on the left with center A_0 and centroid A_c, we move the center of the circle to the position of its centroid, which contains maximum pixel distribution. We will keep on moving this circle until the center of the circle matches its centroid and that area will be considered

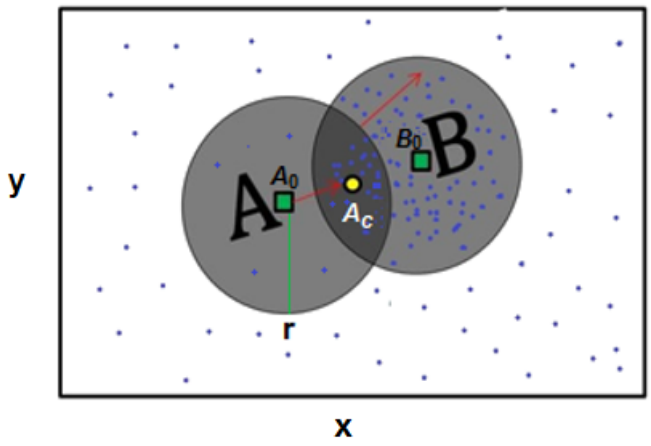


Fig. 3.5 Robust mean algorithm mechanism: original mean lies at A_0 and is moved to B_0 ,which is consider to be the maximum skin pixels area.

as the area of interest with the maximum number of skin pixels (represented by circle B). The algorithm is initialized using the mean of all the points A_0. All points within a radius r of A_0 are then used to re-estimate the mean. The radius r can be any pixel value that fall within the RGB color intensities range (0-255). For example, if we set r = 50 (actually used in our experiments), means we will create a window within the RGB color space with a radius of 50 pixel values, then the diameter of 100 pixel values will be covered by the created window of the algorithm, which is less than the maximum pixel value of 255. This is repeated until the mean stabilizes at location B_0. The aim is that pixels affected by e.g. specular reflections or the appearance of hair or wrinkles are hence eliminated and only skin color pixels are taken in to account for further processing. We have calculated the mean of each color channel values of ROI pixels in each frame using our robust mean algorithm. Let us consider $x_j(i)$ as a mean value of color j, where j represents red, green and blue color channels , for the i^{th} frame. The estimated skin color from each frame of the entire video yields three raw signals $x_r(t), x_g(t)$ and $x_b(t)$, representing red, green and blue color traces, where t denotes time.

3.2.3 Detrending

The three RGB signals that we have obtained may still contain some uncertain variations in their average values despite the effect of lighting having been considerably reduced by using the robust mean algorithm. We, therefore, need to normalize the values of RGB signals in order to achieve further robustness and a cleaner signal. Detrending can eliminate certain variations and normalize the signals for further processing. We have performed this by dividing each sample of these signals by its running average over an interval of 20 samples so that it can contain at least one beat period. The sampling rate of our recorded videos is 15 frames per second, therefore considering a beat every 20 samples is possible.

$$x'_j(t) = \frac{x_j(i)}{\mu}$$

where i represent the i^{th} frame of the signal $j \in \{r, g, b\}$ and μ is the running average over a certain period of time t. The three detrended signals subsequently obtained are $x'_r(t)$, $x'_g(t)$ and $x'_b(t)$. Fig. 3.6 shows the effect of detrending on the three RGB signals.

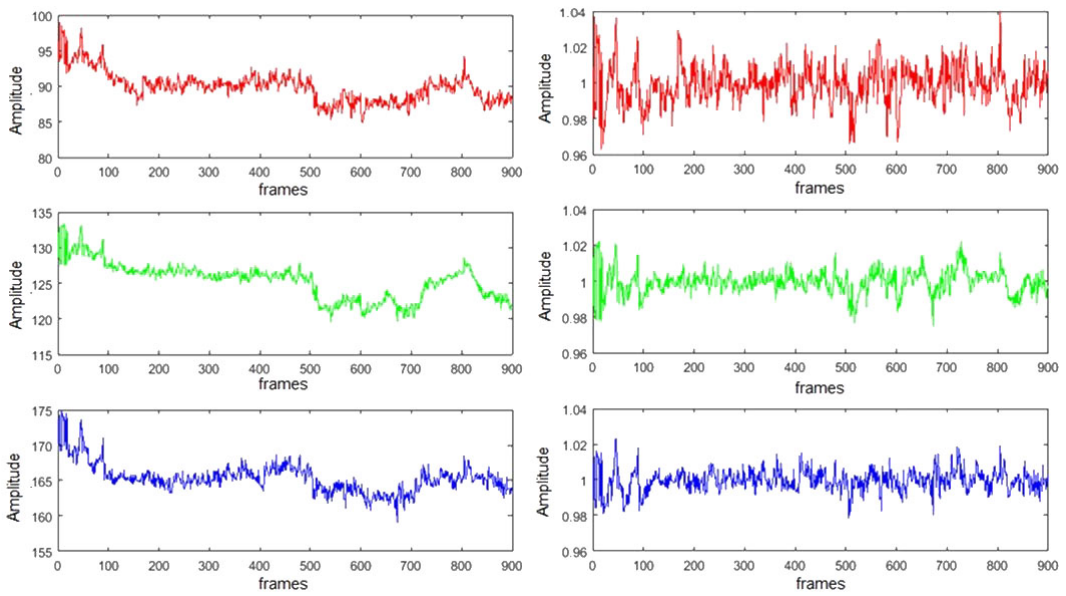


Fig. 3.6 RGB signals before (left) and after (right) detrending.

3.2.4 Independent Component Analysis (ICA)

The normalized RGB signals that we have obtained after detrending are a mixture of the desired pulse signal and other noise sources. In order to extract the pulse from these signals, previous authors [10, 25] have employed the ICA technique of blind source separation (BSS), which is capable of separating different sources mixed together. The concept of BSS is presented in Fig. 3.7, which shows that two persons named A and B are speaking at the same time. Two microphones labeled mic1 and mic2 are recording their voices as a mixture of voices A and B, and it is difficult to distinguish between the voice of person A and B. The voice data recorded by the two microphones is meaningless at this point. The blind source separation is introduced, which is capable of unmixing the voices of both persons and separate them as two different independent

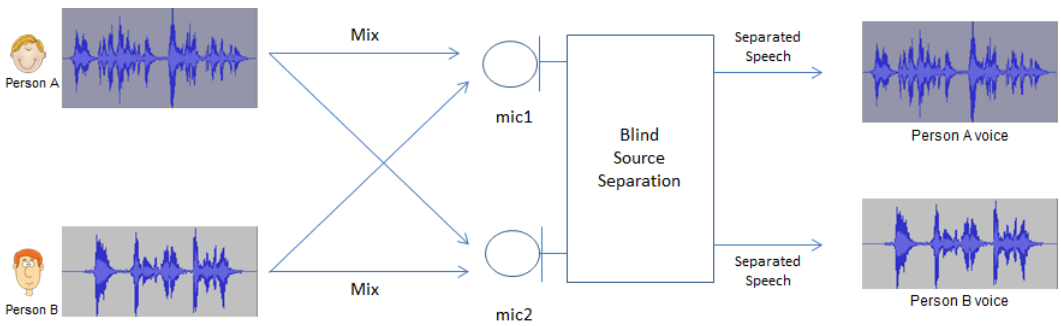


Fig. 3.7 Illustration of Blind Source Seperation (BSS) concept.

sources. The aim of BSS in this context is to try and demix the speech signals into a set of signals each containing the speech signal of a single person. The output obtained is the original voices of person A and person B. Independent component analysis (ICA) [96] is a special case of the broader concept of blind source separation (BSS). The ICA model assumes the observed signals as a linear mixture of different sources, which is represented by the following equation:

$$x_i(t) = \sum_{j=1}^{3} a_{ij} s_j(t) \tag{3.1}$$

where x_i for i = 1,2,3 represents three observed signals, j= 1,2,3 are the three source signals (one source will be pulse and other two will represent noise), a_{ij} are the linear weights, s_j are the statistically independent signals, and t is time. We can rewrite Equation 3.1 compactly using the following mixing equation:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{3.2}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), x_3(t)]^T$ is a column vector containing the observed signals and $\mathbf{s}(t) = [s_1(t), s_2(t), s_3(t)]^T$ is a column vector that contains the source signals, and $\mathbf{A}$ is a 3×3 linear mixing matrix with mixture coefficients a_{ij}. The ICA aims to approximate the inverse of $\mathbf{A}$ as a demixing matrix $\mathbf{W}$ for which

$$\tilde{s}(t) = \mathbf{W}\mathbf{x}(t) \tag{3.3}$$

where $\tilde{s}(t) = [s_1(t), s_2(t), s_3(t)]^T$ is an estimate of the source signal vector **s**(t). The central limit theorem [97] states that the sum of independent random variables is more Gaussian than the original variables. It is, therefore, to demix the observed signals, we must maximize the non-Gaussianity of each recovered signal. Different ICA algorithms use different means of measuring the non-Gaussianity such as kurtosis, negentropy or mutual information.

There are different algorithms of ICA such as JADE, FastICA, RADICAL, and RobustICA [98] [99]. It is quite interesting to know that all these algorithms perform the same for the estimation of heart rate from facial videos. This has been demonstrated by Mannapperuma *et al.* [98] and Christinaki *et al.* [99], and we have also observed this behaviour in our experiments.

Although, there are other methods which could be used to extract the pulse signal cleanly such as PCA, SVD (Singular Value Decomposition), etc., but we have prioritized ICA over other methods. ICA and PCA certainly process the data differently

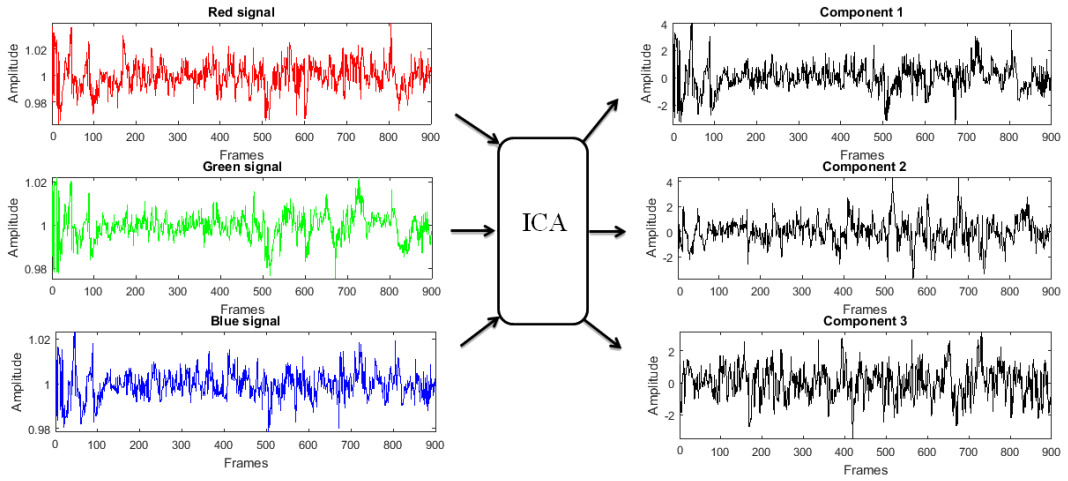


Fig. 3.8 ICA JADE algorithm decomposes three RGB signals (left) in to three independent source components (right).

and will give different results. In PCA all the directions are at right angles (in the high dimensional space), the first along the axis of most variation, the second at right angles to it in the direction of next most variation etc. In ICA the data is first whitened (i.e. scaled so that all directions have the same variance) and axes are found along which the (projected) data are maximally statistically independent. In ICA the axes are not (necessarily) at right angles to each other. The review paper of Anton [100] also recommends ICA over PCA as the best suitable option for pre-processing of an extracted pulse signal. Relating to PCA vs SVD, SVD can be more stable in some situations but it is probably not so important if we are using ICA. In our study, we have used the joint approximate diagonalization of eigenmatrices (JADE) algorithm for ICA developed by Cardoso [101], and used by [10, 25] for the extraction of pulse signal. This approach by tensorial methods uses fourth-order cumulant tensors and involves the joint diagonalization of eignmatrices; the solution of this approximates statistical independence of the sources (to the fourth order). We have used the R version [102] of Cardoso's JADE ICA algorithm ported from Matlab. Conventionally in ICA, the number of outputs is set equal to the number of inputs assigned to it. We have supplied the three normalized RGB signals as the input, and we obtained three source components as output, in which one of them will be considered as the pulse signal and the other two are noise sources. We named these three output sources as component 1, component 2, and component 3 represented in Fig. 3.8.

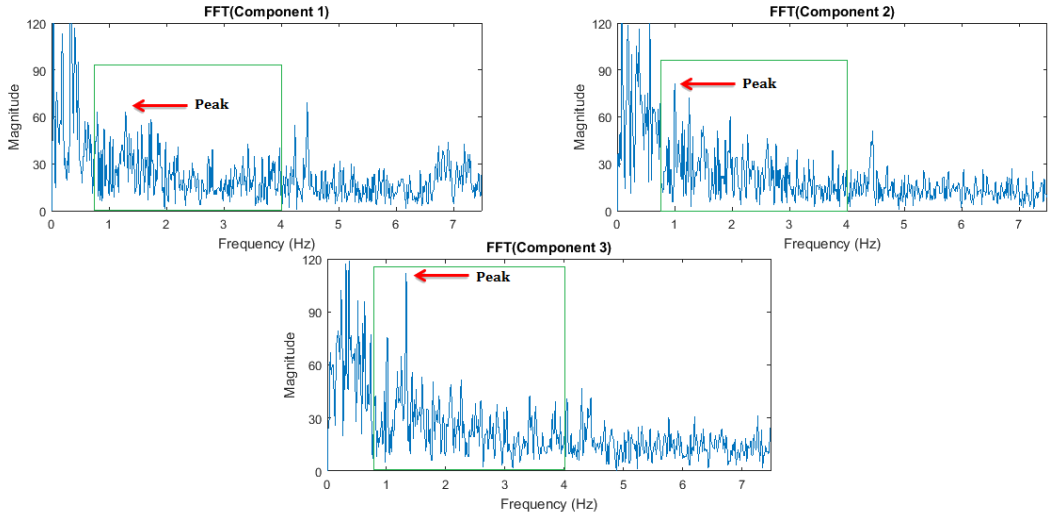


Fig. 3.9 Single-sided FFT power spectrum of component 1 (top left), component 2 (top right), and component 3 (bottom, with the highest peak within set frequency range). Green rectangle represents the set frequency range. The red arrow indicates heart rate peak.

3.2.5 Pulse Signal Identification

ICA does not reveal any information regarding the identification of the expected pulse signal among the three extracted source components. We have employed Poh *et al.*'s [24] criteria, who have used a fast Fourier transform (FFT) to obtain the power spectrum of all the extracted components, and the component whose power spectrum contains the highest peak within the operational frequency was selected as the pulse signal. The highest peak means the peak that has the highest magnitude in the spectrum. We set the operational frequency range to [0.8-4.0] Hz, which corresponds to [48-240] bpm, to cover a wide range of acceptable heart rate measurements. Fig. 3.9 shows the FFT power spectrum of Components 1, 2, and 3 (green rectangle indicates set frequency range). The heart rate is then calculated using the following equation:

$$\textit{Heart Rate Value} = \textit{Pulse component highest peak frequency} \times 60$$

3.2.6 Optimized Filter-based Pulse Signal Extraction

Detection of the highest peak within a fixed window is not very reliable as the estimated pulse component may have been affected by noise, resulting in the selection of the wrong peak. In order to make it possible for a non-contact heart rate measurement system to work accurately in realistic and more challenging conditions, where lighting

and uncertain movements may have been expected, we need a more robust way to extract the heart rate from an estimated pulse component accurately. We have proposed a new method for pulse estimation based on the MOSSE tracking algorithm [103], in which we have created a filter using the ground truth pulse signal and the color signal from training videos to improve the extraction of the pulse signal. Although the idea has been picked up from MOSSE tracking algorithm [103] but the difference is that MOSSE tracking worked on 2D images and instead of our ground truth they constructed a simple Gaussian function centered on the object/feature of interest to create a 2D filter. We are using the same algorithms on 1D signals. Let us consider the following equation:

Pulse Signal (oximeter) = Filtered Pulse Component (video)

$$s = f * c \tag{3.4}$$

where s is the ground truth pulse signal, c is the pulse component obtained from the video, f is the optimal filter function and * is the convolution operator. For a single pair s and c, it is possible to find a function f that satisfies the above equation by converting into the Fourier domain where convolution becomes multiplication i.e.

Equation 3.4 can be rewritten as:

$$S = F^*.C \tag{3.5}$$

Where F, S and C are the Fourier transforms of f, s and c respectively. Since we have the pulse signals and pulse components, we can create a filter using the following equation:

$$F^* = \frac{S}{C} \tag{3.6}$$

The above filter is only suitable for the specific input signals used to construct it. In the more general case, we can design a filter that can be applied to any of the training signals in order to approximate the desired output signals as closely as possible. We wish to find the filter F which minimizes the error function:

$$\chi^2 = \sum_{i=0}^{N} |F^*.C_i - S_i|^2 \tag{3.7}$$

The least squares solution is given by:

$$F = \frac{\sum_i S_i.C_i^*}{\sum_i C_i.C_i^*} \tag{3.8}$$

where S_i is the i^{th} training pulse signal and C_i is the i^{th} training pulse component signal, and * denotes complex conjugate.

We have trained our system using the expected pulse components obtained from the videos and the ground truth pulse signals obtained from the oximeter as an input. Using equation (3.8), we have created a filter and by applying it to each expected pulse component, we obtained an estimated pulse signal. We have employed a leave-one-video out cross-validation (LOOCV) strategy so that the result of components not used in training can be tested against the filter. The average correlation between the filter calculated using all the training data and the filter in each trial is 98 %, which shows that it is not significantly affected by leaving one video out. We have extended the leave one video out validation strategy to the k-fold cross-validation (k=10) and reported the results. The datasets used are explained in the dataset section 3.3.1 in detail. The Fourier transform of the estimated pulse signal K_j from input signal C_j is then estimated as follows:

$$K_j = F.C_j \tag{3.9}$$

The frequency corresponding to the highest peak of the Fourier transform of the

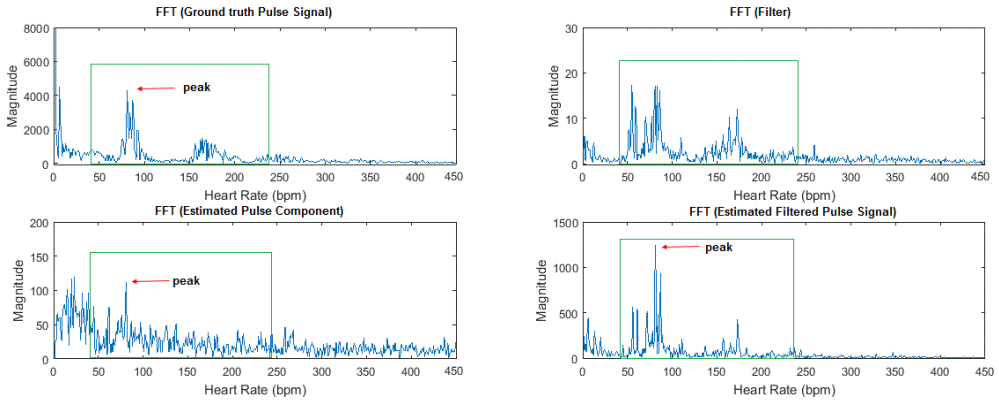


Fig. 3.10 Single-sided FFT power spectrum of the ground truth pulse signal (top left), FFT of the proposed filter (top right), FFT of the estimated pulse component (bottom left) and FFT of the estimated filtered pulse signal (bottom right), red arrow indicates heart rate peak. Green rectangle represents the set frequency range (window size) for prediction of heart rate peak.

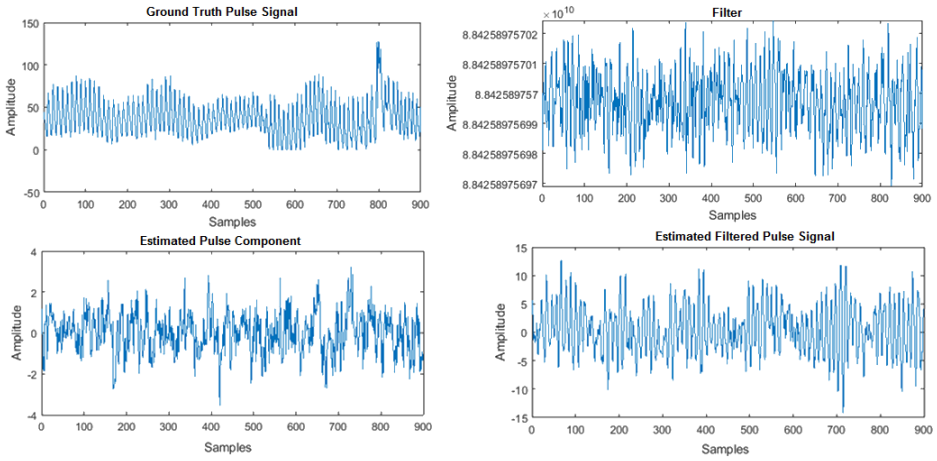


Fig. 3.11 Filter (top right) applied to the estimated pulse component (bottom left), we obtained the estimated filtered pulse signal (bottom right) which looks cleaner and more periodic in line with the ground truth pulse signal (top left).

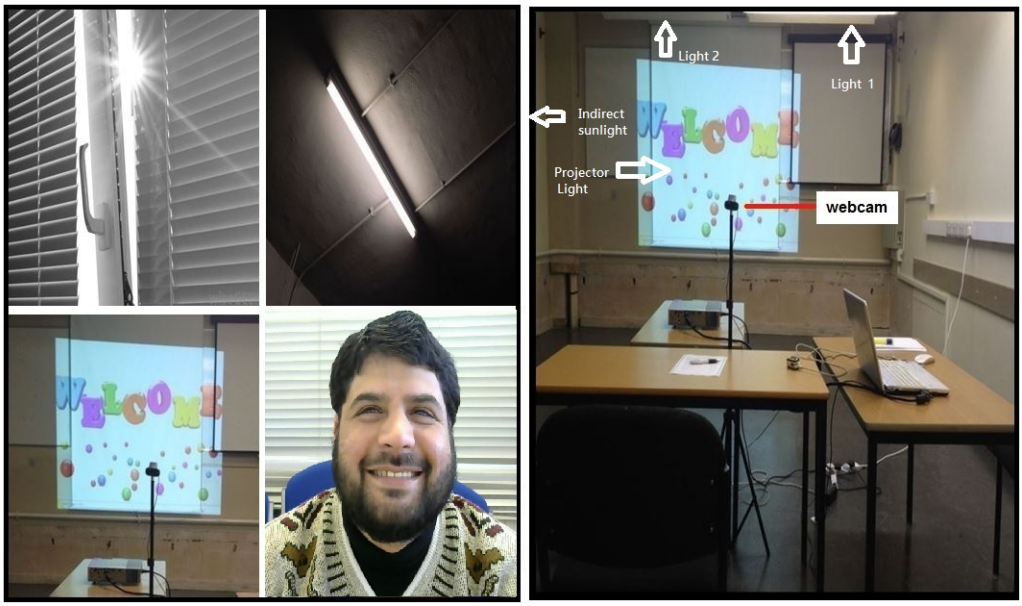
estimated pulse signal estimates the heart rate (see Fig. 3.10) and the inverse Fourier transform of the same signal gives us the estimated pulse signal in the time domain (see Fig. 3.11).

3.3 Experiments

We have implemented the proposed method for the contact-free measurement of pulse rate from human face videos using an ordinary webcam, on a set of three different datasets to generate results. The purpose of testing the proposed method on three different databases is to validate its robustness.

3.3.1 Datasets

The first dataset that was used to evaluate the performance of our proposed method was a local dataset, comprised of human subjects. It consists of 40 videos (1 minute each) from 5 participants (age 25-40) of both genders from different ethnic origins (Asians, Africans, and Caucasians) and skin colors. Informed consent was obtained from all the participants before the experiment and the study was approved by the Ethics Committee of our Institute. All experiments were conducted indoors with office fluorescent lights (one light was flickering randomly throughout the experiments) and indirect sunlight as the main sources of illumination, and a projector which reflected light making the lighting conditions more challenging. The lighting conditions can be more challenging if there exists more variations in the intensity of light and the



(a) (b)

Fig. 3.12 (a) Challenging capture parameters includes indirect sunlight, light flickering, projector's light reflection, and participant's facial features movement. (b) Experimental setup.

reflection of light. The heart rate is measured from the skin color changes on the human face and the more illumination distractions can make the scenario more noisy and challenging, as the change of color on the facial skin caused by the affect of a heartbeat pulse is very tiny. More robust pulse estimation method is required to measure that change and count the number of instances it is caused to be changed. An ordinary Logitech webcam (C310) was used in our experiments which has the following specifications: 0.3 Megapixels (Mp) video capture capability, 24 Bit RGB with 3 channels (8 Bit per channel), 15 frames per second (fps) with a pixel resolution of 640×480 pixels. We asked the participant to sit and relax on a chair in front of a webcam at a distance of approximately 0.5 meters. We played a video clip on the projector screen that contained different emotional cues (happy, sad, and scary videos) and asked the participant to view it while we recorded a color video of the participant's face. The happy clip was a video from a famous comedian Mr. Bean, the sad video clip chosen was the comparison between rich and poor people specifically focussed on poor and needy children, the scary video was an animated movie clip (San Andreas 2015) showing the destructions by a tsunami (links to the shown video clips are provided in Appendix B). The purpose of showing them the video clip was to try to vary their heart rates, add projector light to our lighting conditions, and to encourage the participants to move their facial features (eyes, nose, mouth, head etc.) due to their

Fig. 3.13 Thumbnails from video recordings of 5 subjects

emotional expression response, to a certain degree. It should be noted that we are not doing emotion research, but only trying to make the conditions more challenging. The experimental scenario was hence made more realistic and challenging as we included extra lighting and movement artifacts. Fig. 3.12 (a) presents the capturing conditions and Fig. 3.12 (b) shows the lighting locations and experimental setup. We have recorded a color video of each participant synchronized with the projector display (refresh rate of 15 fps). All recorded videos were saved in .AVI format on a laptop under unique identifiers. Fig. 3.13 shows a thumbnail from each video of all subjects. We used a pulse oximeter Contec CMS50E (clipped to the participant's finger) in order to measure the ground truth heart rate at 60 Hz simultaneously for training and testing purposes. We asked the participants to avoid the finger movement so that the risk of movement artifacts can be minimized. The effect of subject's finger movement noise can be viewed in Fig. 2.5. in chapter 2.

In order to validate the robustness of the proposed method, we have used our second dataset in which we have recorded an additional 10 videos (1 minute each) of a single subject under 10 different lighting conditions. Fig. 3.14 shows a thumbnail from each video. We started capturing in the presence of indirect sunlight as the only source of illumination, then we added artificial light in the form of one fluorescent office light and further recorded in adding more light bulbs. We also performed some captures outdoors in the presence of natural light.

Fig. 3.14 Thumbnails from video recordings of single subjects in 10 different lighting.

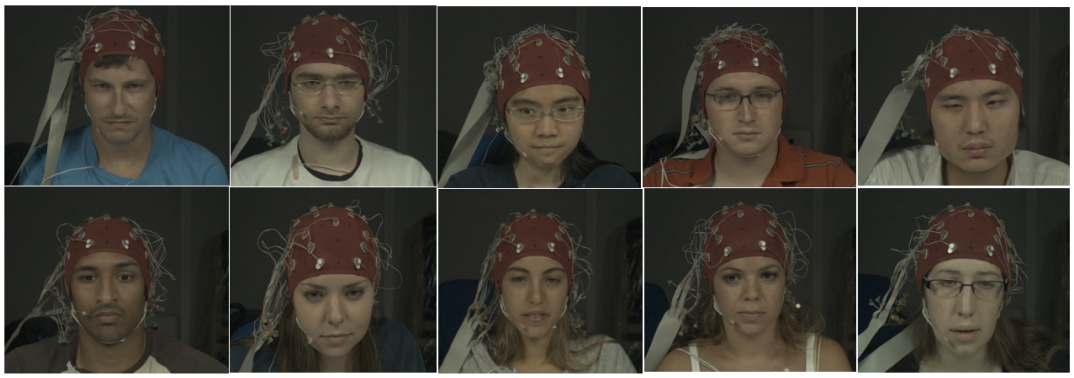


Fig. 3.15 Thumbnails from video recordings of 10 subjects from MAHNOB database.

The third dataset used was from a publicly available MAHNOB-HCI database [72]. The database consists of 30 subjects from different cultural and educational backgrounds. They were shown fragments of movies and pictures, while monitoring them with 6 video cameras, a head-worn microphone, an eye gaze tracker, as well as physiological sensors measuring ECG, EEG (32 channels), respiration amplitude, and skin temperature. We have used 10 videos (recorded at 61 fps) of one minute each for 10 different subjects in our experiment. Fig. 3.15 shows a thumbnail from each video. The ground truth was measured by the physiological sensor at the sampling rate of 256 Hz. We detrended the RGB signals by using a moving average at the rate of 80 frames due to the frame per second rate of 61 so that it can contain one beat period. We found that the videos in the database are not synchronised very well with the ground truth pulse signals. The recorded videos also contain the problem of dropping out the frames over the duration of videos. The low quality video recording and non-synchronisation of pulse signals with the recorded video made this dataset hard for the estimation of heart rate.

3.4 Results

We have evaluated the performance of our proposed method (described in Section 3.2) for the measurement of heart rate in the presence of challenging illumination interference and motion artifacts using the following statistics:

Mean error ($\bar{e}$) - the mean value of the differences between the estimated heart rate from the videos and the ground truth heart rate measured by means of a pulse oximeter.

Mean absolute error ($\overline{|e|}$) - the mean value of the absolute differences between the estimated heart rate and the reference heart rate.

Standard deviation of error (σ) - the sample standard deviation of the differences

between the estimated and the reference heart rate.

Root mean square error (RMSE or ε) - the root mean square error of the differences between the estimated and the ground truth heart rate.

Correlation coefficient (r) - Pearson's correlation coefficient which is a measure of the linear dependence between two variables. We will find the correlation between the estimated heart rate and the ground truth heart rate. The greater r value will indicate the best result and the lower r the opposite. A p-value is calculated to test the statistical significance of correlation between the estimated heart rate and the ground truth heart rate. If $p < 0.05$ then the correlation will be significant and if $p > 0.05$ then the correlation will be not statisticaly significant.

We compared the estimated heart rate obtained from recorded videos to the ground truth heart rate measured by a pulse oximeter. Further, to illustrate the effectiveness and accuracy of the proposed method, we have also compared our results with the state-of-the-art methods [10] [24–26] for the measurement of heart rate from human face videos. In order to validate the robustness of the proposed method, we present the results of the heart rate measured in different experimental setups with varying lighting conditions. Finally, we analyzed the results obtained using the publicly available dataset of MAHNOB-HCI [72].

3.4.1 Experiment 1

As mentioned in Section 3.3.1, our first experiment is based on the video data from our own dataset. The dataset consists of 40 videos (one minute each) of 5 different subjects recorded under challenging experimental conditions in the presence of varying amount of illumination and motion artifacts. The Leave-one-video-out cross-validation was used for training and testing the dataset of 40 videos. The filter was created (see section 3.2.6) by leaving one video out in each iteration and use the remaining 39 videos for training purpose. The left video is then tested against the created filter by applying the filter to that video. We will explain in detail the performance of every processing step of our proposed method in the following section.

The robustness of the Saragih's face tracker [91] used in our experiments can be seen in Fig. 3.3, which shows a sequence of frames tracking the face from the captured color video of the subject. The tracker can cope with left, right, up, and down movement of the face very well. The freedom of 90-degree rotation each way can easily track the face. The ROI (single-sided cheek) can be tracked if the subject moves the face up to 45° each way, and if we target both-sided cheeks as ROI then the tracker allows 90° freedom to move the face each way. We obtained 100% success rate for

the detection of frames containing the face. This overcomes the problem of loss of frames containing the face by the trackers reported in previous studies [10] [24, 25]. Fig. 3.4 presents a sample frame for the tracking of a face from the recorded video and extraction of the ROI (single-sided cheek) is shown as a green square. The ROI yields three RGB color traces using the robust mean algorithm (which tries to ignore the pixels affected by specular illumination and taking into account only facial skin color pixels) and calculated the mean of RGB color intensities for all pixels in each frame of the recorded video. Table 3.1 shows the effectiveness of the robust mean algorithm, which we have used instead of using the average mean for the measurement of heart rate from facial videos. The algorithm reduced the absolute mean error from 11.22 bpm to 8.72 bpm, the standard deviation of error was reduced from 18.12 bpm to 11.49 bpm, the RMSE reduced from 18.02 bpm to 13.29 bpm, and the correlation coefficient was improved from 0.14 (p = 0.35) to 0.53 (p = 0.0003), and therefore has a positive impact on improving the overall results. We can, therefore, assume that the robust mean algorithm ignores pixels affected by illumination and is calculating the mean of those pixels which are most likely to be representing skin.

The impact of detrending can be viewed in Fig. 3.6, which normalizes the RGB signals for post-processing. We can see that the three RGB color signals on the left-hand side contain uncertain variations in their mean values over the course of the video,

Table 3.1 Analysing the effect of the Robust mean algorithm in the heart rate measurement using 40 videos of the local database with challenging conditions.

	Average	Robust mean algorithm
$\bar{e}$ (bpm)	-2.17	6.92
$\overline{\lvert e \rvert}$ (bpm)	11.22	8.72
σ (bpm)	18.12	11.49
ε (RMSE)	18.02	13.29
$p-value$	0.35	0.0003
r	0.14	0.53

The table shows mean error $\bar{e}$, absolute mean error $\overline{|e|}$, standard deviation of error σ, root mean square error ε , $p-value$, and Pearson correlation r.

caused by illumination and motion artifacts, which are normalized after the application of detrending. The right-hand side of Fig. 3.6 shows the effect of detrending, which tends to compensate the uncertain variations in the average values of the RGB signals.

The detrended RGB signals were used as an input for ICA, resulting in three independent source components: i.e. component 1, component 2, and component 3 as shown in the Fig. 3.8. The ICA returns the output source components in random order, it is therefore not known which component needs to be selected for the estimation of the heart rate. The three source components were decomposed from their time domain into their frequency domain using an FFT. Fig. 3.9 clearly shows that component 3 has the highest peak in its power spectrum among the three components within the set frequency range, therefore this was chosen as the desired pulse component. The frequency that corresponds to the highest peak of component 3 in its power spectrum within the set frequency range was selected as the heart rate frequency. This was one subject's example for which component 3 was selected as an estimated pulse component, other subjects may have different components which may qualify for the highest peak pulse component.

As shown in Fig. 3.9, we have obtained a prominent peak in the third component but the signal (in the Fourier domain) is not clean as yet because some noise is still present. Our proposed optimal filter method (explained in section 3.2.6) has not only improved the accuracy of the heart rate measured but also resulted in a cleaner pulse signal, which is closer to the measured ground truth pulse signal. Fig. 3.10 shows the FFT power spectrum of the ground truth pulse signal and the designed filter. The filter, when applied to the FFT power spectrums of the expected heart rate component (component 3) revealed the estimated pulse signal. Fig. 3.10 compares the FFT power spectrum of pulse signals obtained before and after the application of the optimal filter. The power spectrum of both of these signals shows that the heart rate peak (indicated by a red arrow) became more prominent after we applied the filter. The signal noise is also reduced, which shows that the effect of illumination interference and motion distractions is reduced and we get a cleaner pulse signal. Fig. 3.11 shows the ground truth pulse signal, designed optimal filter, expected heart rate component and the estimated pulse signal in the spatial domain, which are obtained by the inverse FFT of their frequency domain counterparts. We find that the estimated filtered pulse signal is more periodic in nature as compared to the estimated pulse component. The noise is hence filtered out and we get a cleaner periodic signal, which is a better representation of the pulse signal.

Fig. 3.16 presents the comparison of different processing steps (explained in section 3.2) and their impact on the improvement of the measured heart rate. We can

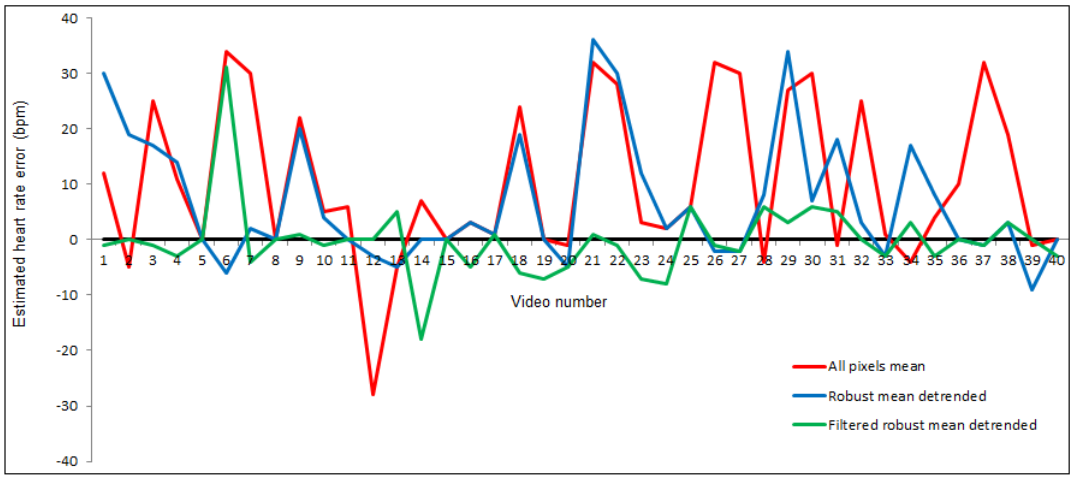


Fig. 3.16 Plot of comparison of different processing steps (section 3.2) involved in the measurement of heart rate against the ground truth for 40 videos of the local database captured in challenging conditions. The black line is a ground truth heart rate error, the red line indicates the use of all pixels mean, the blue line shows the impact of the robust mean algorithm. The green line represents the proposed adaptive filter method estimated heart rate error and shows that it remains closer to the ground truth.

see that the green line, which represents the final step yielding the estimated heart rate from facial videos using the filtered robust mean method, comes closer to the ground truth heart rate, with the introduction of additional processing steps i.e. robust mean algorithm, detrending, and filter optimization.

Table 3.2 presents the robustness of the proposed method compared to the other state-of-the-art methods [10, 24–26] for the measurement of heart rate using 40 videos of the local dataset. The absolute mean error is reduced to 3.77 bpm as compared to de Haan *et al.* (10.30 bpm), Kwon *et al.* (13.67 bpm), and Poh *et al.* (12.75 bpm & 11.22 bpm). The proposed method has a lowest standard deviation error of 6.75 bpm as compared to the second best of Kwon *et al.* 13.19 bpm, and lowest RMSE of 6.67 bpm as compared to the second best of Poh *et al.* 17.70 bpm . The higher value of Pearson correlation coefficient r = 0.82 ($p = 6.85e-11$) indicates the strong relationship of the proposed method estimated heart rate and the ground truth heart rate as compared to other methods. Fig. 3.17 illustrates the correlation coefficient plots of all the implemented methods. Fig. 3.18 shows the performance of each method, and we can notice that the green line of the proposed method estimated heart rate error remains closer to the black line of the ground truth error. We can conclude that the proposed method has the minimum absolute mean error, standard deviation of error, RMSE, and higher correlation value as compared to other methods, therefore, it shows

Table 3.2 Performance comparison between the proposed method and state-of-the-art methods for heart rate measurement using 40 videos of the local database with challenging conditions

	Poh *et al.* [10]	*Poh*^{2} *et al.* [24]	Kwon *et al.* [25]	de Haan and Jeanne [26]	Proposed method
$\bar{e}$ (bpm)	10.30	-2.17	13.17	-5.15	-0.22
$\overline{\lvert e \rvert}$ (bpm)	12.75	11.22	13.67	10.30	3.77
σ (bpm)	14.57	18.12	13.19	21.65	6.75
ε (RMSE)	17.70	18.02	18.53	21.99	6.67
$p-value$	0.03	0.35	0.01	0.73	$6.85e-11$
r	0.33	0.14	0.40	-0.05	0.82

The table shows mean error $\bar{e}$, absolute mean error $\overline{\lvert e \rvert}$, standard deviation of error σ, root mean square error ε , $p-value$, and Pearson correlation r.

that the proposed method performed the best in comparison to other state-of-the-art methods under more realistic and challenging conditions.

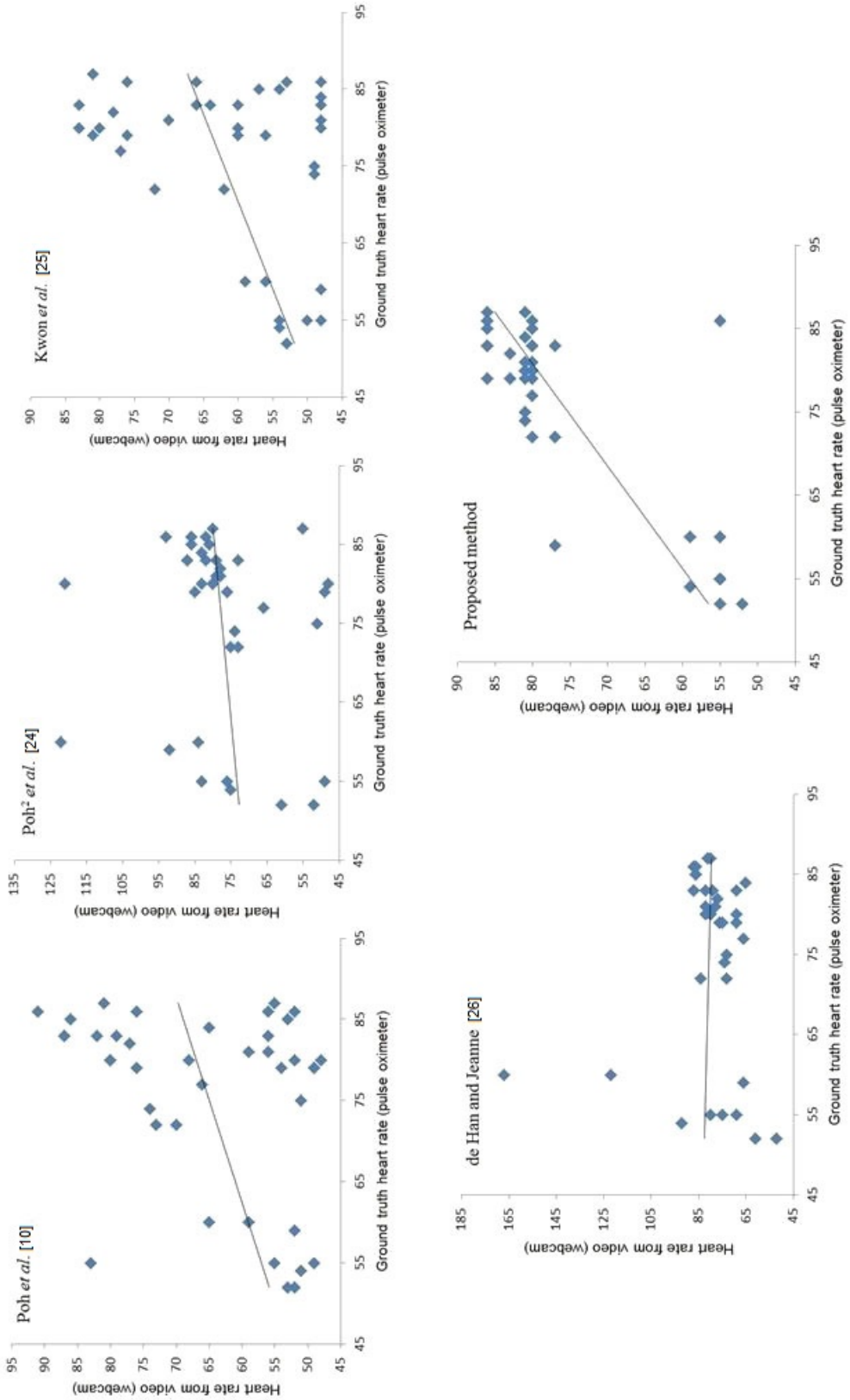


Fig. 3.17 Correlation performance comparison between the proposed method and the state-of-the-art methods [10] [24–26] of heart rate measurement using 40 videos of the local database with challenging conditions. The proposed method plot shows that the estimated heart rate from videos is highly correlated to the ground truth heart rate as compared to other methods.

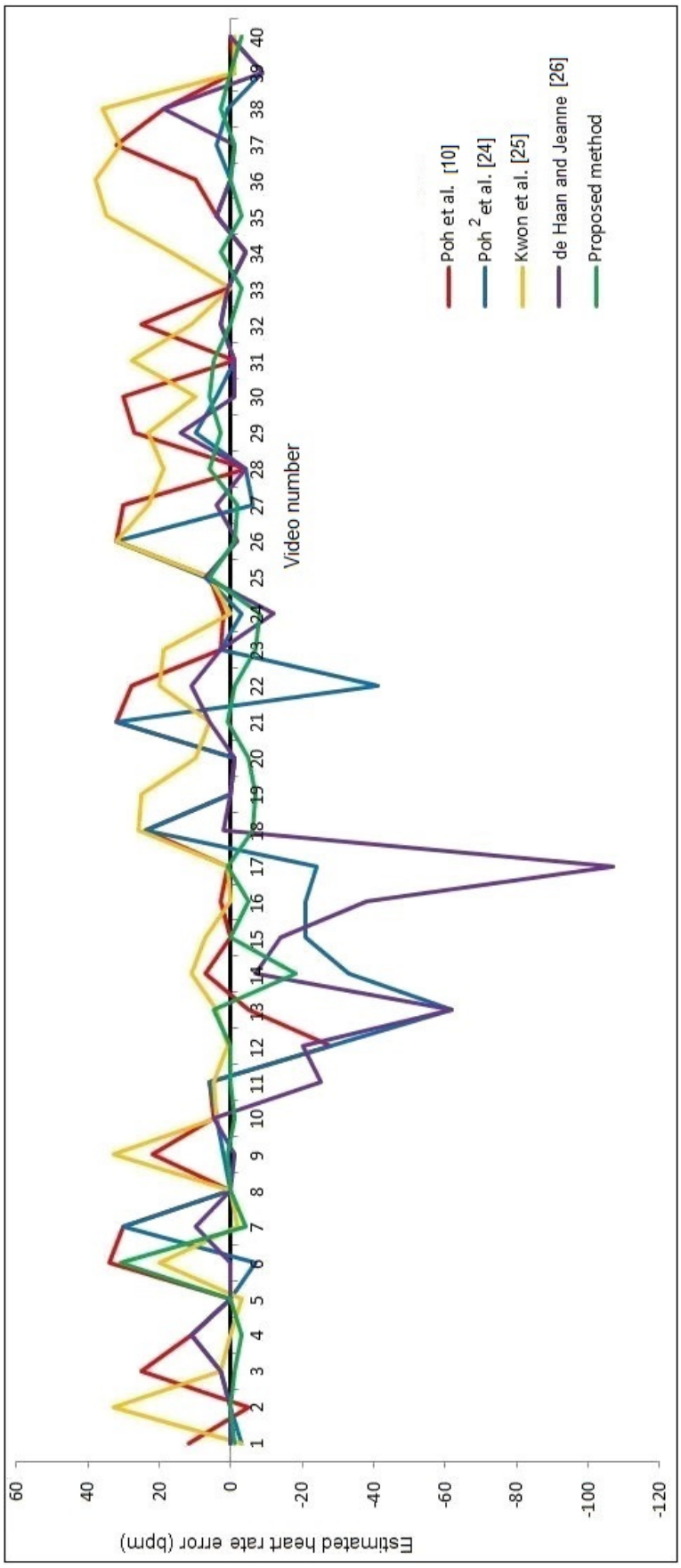


Fig. 3.18 Performance comparison between the proposed method and the state-of-the-art methods of the heart rate measurement using 40 videos of the local database with challenging conditions. Black line is a ground truth heart rate error, red line indicates Poh *et al.* 2010 [10], blue line represents Poh *et al.* 2011 [24], yellow line shows Kwon *et al.* [25], purple line indicates de Haan and Jeanne [26], Green line represents the proposed method estimated heart rate error (which shows that it remains closer to ground truth as compared to other methods).

Full face ROI vs Cheek as an ROI

In order to validate that the cheeks is the most suitable region to cleanly extract the pulse signal from, we extend our experiment 1 to estimate the heart rate from the full face rather than choosing 40×40 pixels cheek as a ROI. We used our own created dataset explained in the dataset section 3.3.1 which consists of 40 videos (1 minute each) recorded from 5 different subjects under challenging experimental conditions. We applied all the processing steps of our proposed method for the estimation of heart rate explained in section 3.2 to obtain the results. The following Table 3.3 shows the results of the experiment.

Table 3.3 Comparison of heart rate estimation using Full face and 40×40 Cheek as ROI using 40 videos of the local database with challenging conditions.

	Full face ROI	40×40 pixels Cheek ROI
$\bar{e}$ (bpm)	-3.70	-0.22
$\overline{\lvert e \rvert}$ (bpm)	7.60	3.77
σ (bpm)	10.83	6.75
ε (RMSE)	11.32	6.67
$p-value$	0.00037	$6.85e-11$
r	0.44	0.82

The table shows mean error $\bar{e}$, absolute mean error $\overline{\lvert e \rvert}$, standard deviation of error σ, root mean square error ε , $p-value$, and Pearson correlation r.

We can find from the Table 3.3 that the absolute mean error for the estimated heart rate in case of a full face as ROI is 7.60 bpm, which is reduced to 3.77 bpm if we choose only 40×40 pixels cheek as ROI. The standard deviation of error is also reduced from 10.83 bpm to 6.75 bpm, and the greater value of the Pearson correlation coefficient r = 0.82 (p=$6.85e-11$) in case of cheek ROI as compared to full face r= 0.44 (p=0.00037) validates that we can obtain better results if we use the cheek as an ROI because it tends to be a highly sensitive area of the face which is visible to the camera sensors, and is least likely to be affected by a beard, hair, head covering, and glasses. It is also less affected by facial feature movements and we considered it to be the most stable area of the face to extract maximum skin pixels region.

k-Fold Cross-Validation (k=10) vs LOOCV:

We have used k-fold cross-validation (k-fold CV) method instead of LOOCV to validate the small dependency on the size of the training set, and reported the results of the estimated heart rate from human face videos. The dataset of 40 videos was split into 10 folds with 4 videos in each fold. Table 3.4 compares the results obtained using leave-one-out cross-validation and k-fold cross-validation. We can find that the k-fold cross-validation produce almost similar results as compared to LOOCV. The absolute mean error is calculated as 3.90 bpm, and the the accuracy of 80 % is obtained. The drop in performance with k-fold CV is only very marginal, indicating only a small dependency on the size of the training set.

Table 3.4 Comparison of the proposed heart rate estimation method using Leave-one-out cross-validation (LOOCV) and k-fold cross-validation (k-fold CV) with k=10 using 40 videos of the local database with challenging conditions.

	LOOCV	k-fold CV (k=10)
$\bar{e}$ (bpm)	-0.22	-0.50
$\overline{\lvert e \rvert}$ (bpm)	3.77	3.90
σ (bpm)	6.75	6.97
ε (RMSE)	6.67	6.90
$p-value$	$6.85e-11$	$3.21e-10$
r	0.82	0.80

The table shows mean error $\bar{e}$, absolute mean error $\overline{|e|}$, standard deviation of error σ, root mean square error ε , $p-value$, and Pearson correlation r.

3.4.2 Experiment 2

We can see from the results of experiment 1 that the adaptive filtering contributes significantly to reduce the error in the heart rate measured from human face videos. In order to further evaluate the robustness of the proposed method, we have used our second dataset mentioned in section 3.3.1, in which we have recorded an additional 10 videos (1 minute each) of a single subject under 10 different lighting conditions. The Leave-one-video-out cross-validation was used for training and testing the dataset

of 10 videos. The filter was created (see section 3.2.6) by leaving one video out in each iteration and use the remaining 9 videos for training purpose. The left video is then tested against the created filter by applying the filter to that video.Table 3.5 shows the results. The absolute mean error of the estimated heart rate using our proposed method is reduced to 4.20 bpm as compared to de Haan *et al.* (8.10 bpm), Kwon *et al.* (16.70 bpm), and Poh *et al.* (19.80 bpm & 22.70 bpm). The proposed method has a lowest standard deviation of error i.e. 4.40 bpm as compared to the second best Poh *et al.* standard deviation of 8.32 bpm. The RMSE was also found the lowest i.e. 4.91 bpm as compared to second best of de Haan *et al.* of 11.19 bpm. The higher value of Pearson correlation coefficient r = 0.68 (p = 0.02) indicates the strong relationship of the proposed method estimated heart rate and the ground truth heart rate as compared to other methods. Fig. 3.19 illustrates the comparison of the correlation coefficient plots of all the implemented methods. Fig. 3.20 shows the overall performance of each method, and we can notice that the green line of the proposed method estimated heart rate error remains closer to the ground truth . We can conclude that the proposed method has the minimum mean error, absolute mean error, standard deviation of error, RMSE, and higher correlation value as compared to other methods, therefore, it shows

Table 3.5 Performance comparison between the proposed and the state-of-the-art methods of heart rate measurement for a single subject at 10 different lighting conditions setup.

	Poh *et al.* [10]	*Poh*^{2} *et al.* [24]	Kwon *et al.* [25]	de Haan and Jeanne [26]	Proposed method
$\bar{e}$ (bpm)	22.70	18.00	14.50	3.50	-2.60
$\overline{\lvert e \rvert}$ (bpm)	22.70	19.80	16.70	8.10	4.20
σ (bpm)	8.32	13.77	15.32	11.20	4.40
ε (RMSE)	24.03	22.24	20.53	11.19	4.91
$p-value$	0.78	0.58	0.43	0.55	0.02
r	0.10	-0.19	-0.27	0.21	0.68

The table shows mean error $\bar{e}$, absolute mean error $\overline{\lvert e \rvert}$, standard deviation of error σ, root mean square error ε, $p-value$, and Pearson correlation r.

that the proposed method performed the best in comparison to other state-of-the-art methods under different lighting conditions.

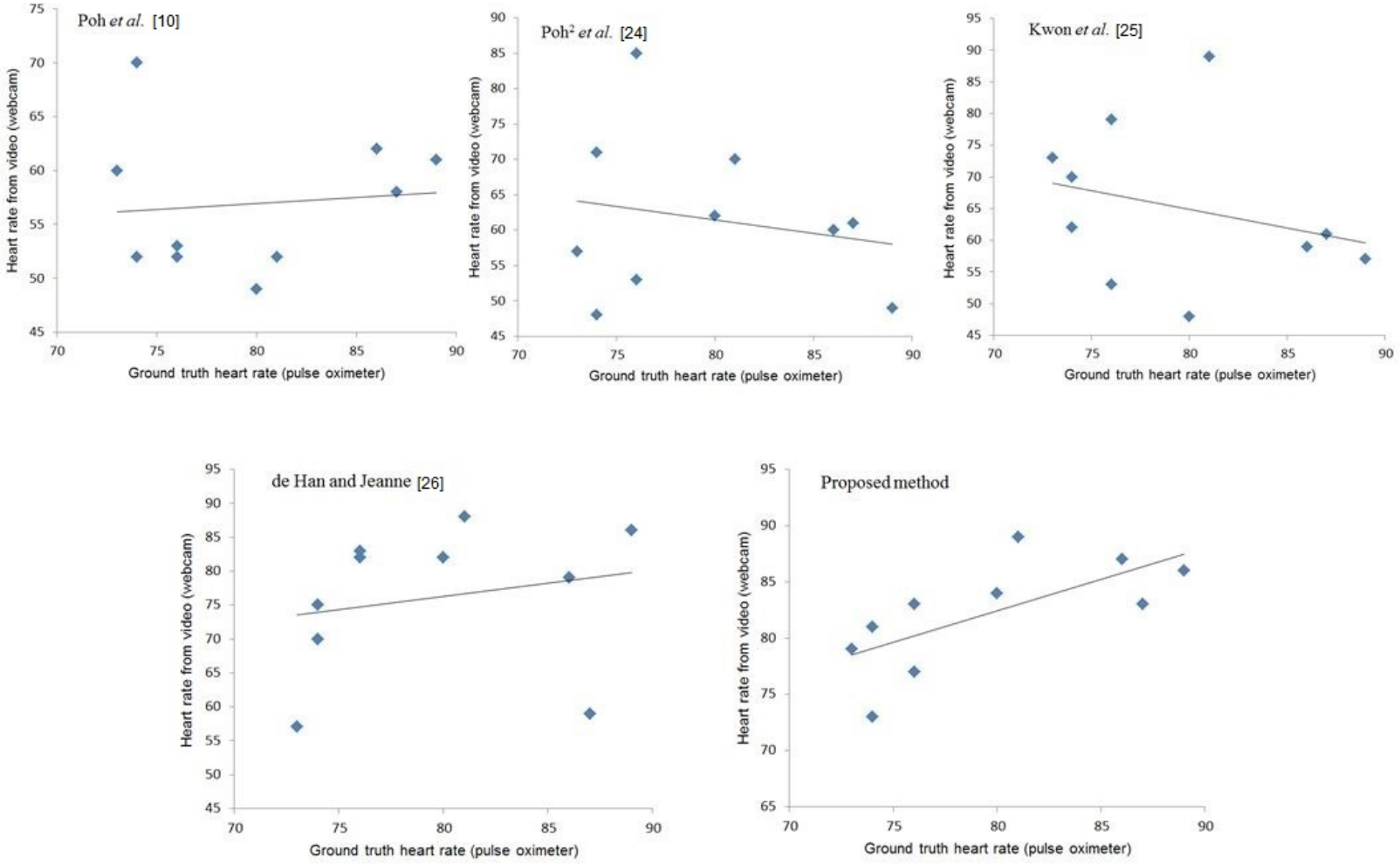


Fig. 3.19 Correlation performance comparison between the proposed method and the state-of-the-art methods [10] [24–26] of the heart rate estimation using 10 videos of a single subject under different lighting conditions. The proposed method plot shows that the estimated heart rate from videos is highly correlated to the ground truth heart rate as compared to other methods.

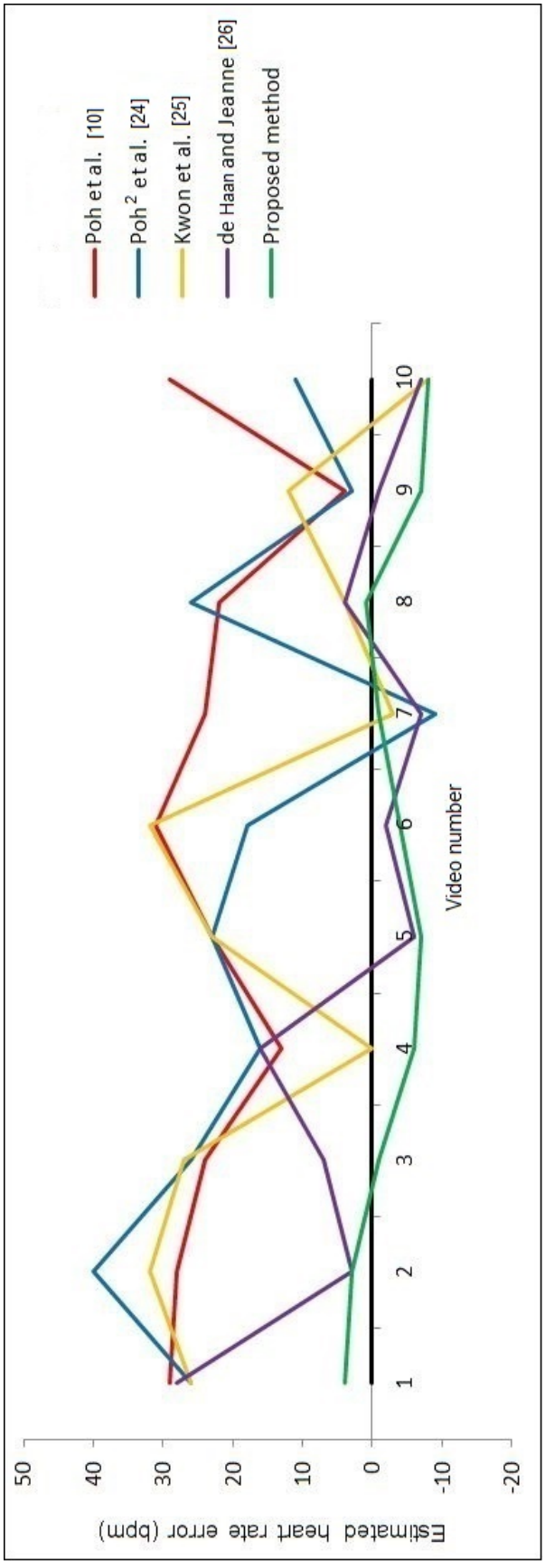


Fig. 3.20 Performance comparison between the proposed method and the state-of-the-art methods of heart rate measurement for a single subject under 10 different lighting conditions scenarios. Black line is a ground truth heart rate error, red line indicates Poh *et al.* 2010 [10], blue line represents Poh *et al.* 2011 [24], yellow line shows Kwon *et al.* [25], purple line indicates de Haan and Jeanne [26], Green line represents the proposed method estimated heart rate error(which shows that it remains closer to ground truth as compared to other methods).

3.4.3 Experiment 3

In order to validate the robustness of our proposed heart rate estimation method (explained in section 3.2) from facial videos, we have tested our method on the publicly available database of MAHNOB-HCI [72]. We have used one minute videos of 10 different subjects (i.e. dataset of 10 videos) who were recorded under controlled conditions (controlled lighting and little face movement). The Leave-one-video-out cross-validation was used for training and testing the dataset of 10 videos. The filter was created (see section 3.2.6) by leaving one video out in each iteration and use the remaining 9 videos for training purpose. The left video is then tested against the created filter by applying the filter to that video.Table 3.6 presents the results. The absolute mean error of the estimated heart rate using our proposed method is reduced to 5.40 bpm as compared to de Haan *et al.* (7.10 bpm), Kwon *et al.* (6.80 bpm), and Poh *et al.* (9.60 bpm & 7.50 bpm). The proposed method has a lowest standard deviation of error i.e. 6.82 bpm as compared to the second best Kwon *et al.* standard deviation of 7.78 bpm. The RMSE was also found the lowest i.e. 6.52 bpm as compared to second best of de Kwon *et al.* of 9.03 bpm. Fig. 3.21 illustrates the comparison of the correlation coefficient plots of all the implemented methods. Fig.

Table 3.6 Performance comparison between the proposed and the state-of-the-art methods [10] [24–26] of heart rate measurement using 10 videos of 10 subjects from the MAHNOB-HCI database.

	Poh *et al.* [10]	*Poh*^{2} *et al.* [24]	Kwon *et al.* [25]	de Haan and Jeanne [26]	Proposed method
$\bar{e}$ (bpm)	2.20	0.70	5.20	-0.90	-0.80
$\overline{\lvert e \rvert}$ (bpm)	9.60	7.50	6.80	7.10	5.40
σ (bpm)	12.20	10.66	7.78	9.70	6.82
ε (RMSE)	11.78	10.14	9.03	9.24	6.52
$p-value$	0.90	0.98	0.62	0.71	0.43
r	-0.04	-0.008	0.17	-0.13	-0.27

The table shows mean error $\bar{e}$, absolute mean error $\overline{\lvert e \rvert}$, standard deviation of error σ, root mean square error ε , $p-value$, and Pearson correlation r.

3.22 shows the overall performance of each method, and we can notice that the green line of the proposed method estimated heart rate error remains closer to the ground truth. We can conclude that the proposed method has the minimum absolute mean error, standard deviation of error, RMSE, and higher correlation value as compared to other methods, therefore, it shows that the proposed method performed the best in comparison to other state-of-the-art methods.

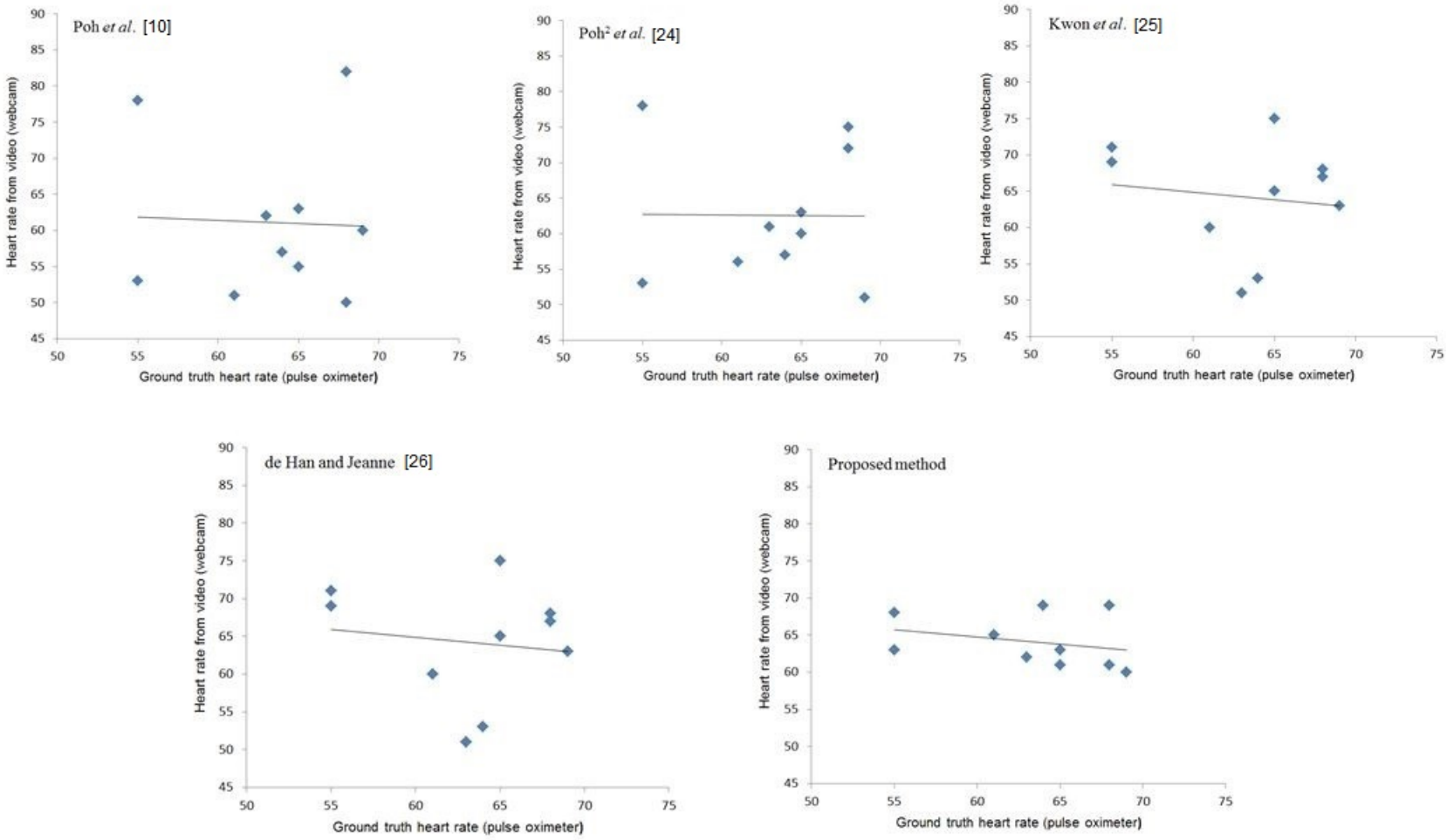


Fig. 3.21 Correlation performance comparison between the proposed method and the state-of-the-art methods [10] [24–26] of the heart rate estimation using 10 videos of 10 subjects from the MAHNOB-HCI database. The proposed method plot shows that the estimated heart rate from videos is highly correlated to the ground truth heart rate as compared to other methods.

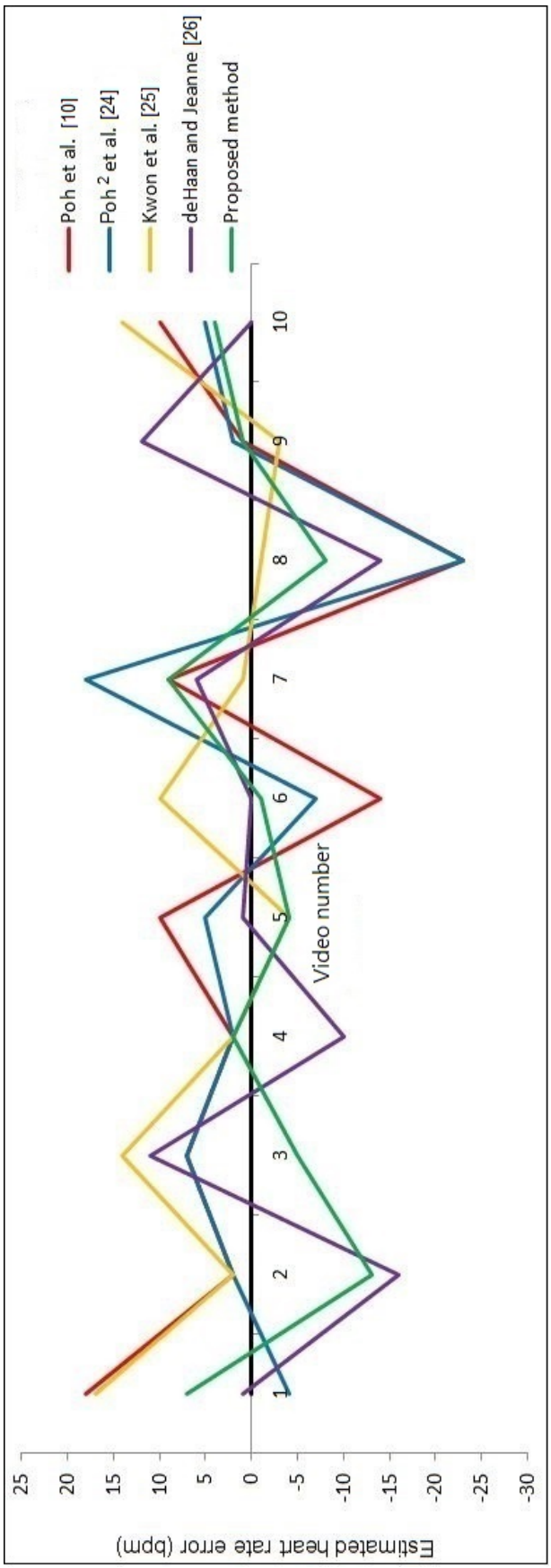


Fig. 3.22 Performance comparison between the proposed method and the state-of-the-art methods of heart rate measurement using 10 videos of 10 subjects from the MAHNOB-HCI database. Black line is a ground truth heart rate error, red line indicates Poh *et al.* 2010 [10], blue line represents Poh *et al.* 2011 [24], yellow line shows Kwon *et al.* [25], purple line indicates de Haan and Jeanne [26], Green line represents the proposed method estimated heart rate error(which shows that it remains closer to ground truth as compared to other methods)

3.5 Discussion

Measurement of the heart rate from human face videos captured by means of a webcam under realistic scenarios is a challenging task as we have found that most of the previous methods are dependent on highly controlled conditions. The controlled conditions in the heart rate estimation means no/less motion of the subject and the difficult lighting settings. Most of the previous work was based on such controlled experimental setups in which they restricted subject's movement while they have been captured, and the no artificial lighting or fixed spot lighting bulbs are used for lighting the capturing scene. In realistic scenarios both subject motion and lighting variations are expected. We have allowed the subjects to move normally and they were allowed to tilt their face upto 90 degrees angle. The subjects were watching the emotional content video clips which triggered their facial features to move considerably as a response to the content of the video, which made the experimental setup more challenging. In order to employ this technology in real world applications, it needs to be robust and easily adaptable to a wide range of capturing scenarios. We have addressed these problems in our work and therefore made our experimental environment challenging with difficult lighting and some natural facial motion. As far as the usage of ground truth pulse signal in the creation of optimal filter was concerned, we have tested the filter against the video which was not a part of the creation of that filter. So we generalize the filter which can be applied to any video to extract the pulse signal. In other words, the tested instances were not a part of the training instances and are independent of being a part of the optimal filter. Results obtained demonstrate the improved performance of the proposed method, but there are some aspects that need to be taken into account which could lead to further improvement.

The used ROI was situated on the cheeks in our system but we could also investigate other parts of the body that may give a better estimation of variations in skin color pixels with respect to each heartbeat. We have normalized the RGB signals before ICA, but additional pre-processing of the raw signals may perform well at the post ICA stage, where we are identifying the expected pulse signal. We have chosen the independent source component with the highest peak in the power spectrum within the set frequency range as the expected pulse component, but we have found that in some components a smaller peak was more prominent (high compared to its neighbours) than the highest peak components, and the heart rate obtained was closer to the ground truth heart rate. The criterion for selecting the correct source component for estimating heart rate may need to be improved by setting some threshold or providing additional processing. We found that our method performed less efficiently in the case of very

dark skin color subject. Video no. 6 and 14 in Fig. 3.18 shows the heart rate of a dark skin subject, which can be seen in Fig. 3.17 as well as the out-liners. The possible solution to this problem can be the use of near-infrared cameras, so that if the heart rate cannot be measured by the skin pixels color variation of a webcam then thermal energy radiated from the skin using near-infrared cameras can cover it. If we eliminate these dark skin subject's videos from the dataset, we achieved an accuracy of 92%. It should be noted in the Fig. 3.17 of the experiment 1, which consists of videos captured using our own created database, that the heart rates are highly concentrated around the 85bpm, which shows the impact of emotional video shown to the participants. The emotional content in the shown video raises the heart rate for all subjects. This concludes that the heart rate is highly related to the human emotions, and the heart rate from videos method can help identify the human emotions. We will try to detect human emotions using our proposed heart rate estimation method in the future. We noticed that in the case of using the MAHNOB-HCI dataset, the proposed method did not perform significantly better as compared to other methods (in comparison to our results obtained using local datasets) due to the fact that the videos in the database are not synchronised very well with the ground truth pulse signals. The recorded videos also contain the problem of dropping out the frames over the duration of videos. The poor quality of video recording and non-accurate synchronisation problems made this dataset hard to estimate the accurate heart rate and extract a cleaner pulse signal. In Fig. 3.11, we can find that the estimated filtered pulse signal is more periodic in nature as compared to the estimated pulse component. The noise is hence filtered out and we get a cleaner periodic signal (highest peak), which is a better representation of the pulse signal. Fig. 3.10 present FFT representation of the filtered signal. It compares the FFT power spectrum of pulse signals obtained before and after the application of the optimal filter. The power spectrum of both of these signals shows that the heart rate peak (indicated by a red arrow) became more prominent (with the peak at highest altitude) after we applied the filter. The signal noise is also reduced, which shows that the effect of illumination interference and motion distractions is reduced and we get a cleaner pulse signal. We can also notice that the filtered signal look more like the ground truth pulse signal in Fig. 3.11

3.6 Conclusion

In this chapter, we have presented an improved method for the measurement of heart rate from human face videos under realistic and more challenging conditions. We

introduced the use of Saragih face tracker [2] for the face tracking and extraction of ROI in the images, which performed better and reduced the error occurring due to the subject's movement. The proposed robust mean algorithm minimized the effect of illumination variations. The optimized filtering approach not only improved on the existing methods for the measurement of the heart rate but also resulted in a cleaner pulse signal, which could be integrated into many other real-life applications such as human biometrics, emotion recognition, driving monitoring systems, and medical care equipment.

3.7 Summary

In this chapter, we have addressed the two main problems of contact-free heart rate measurement from human face videos using a webcam. They are illumination noise and subject motion artifacts. We have proposed a solution to minimize the effect of these problems and presented the results using local and public databases. The results show that our proposed method significantly improved the contact-free heart rate measurement from human face videos using a webcam.

Chapter 4

Human Biometrics Based on Pulse Signals Extracted From Facial Videos

The improved performance of our proposed contact-free pulse signal extraction from human face videos using a webcam, presented in chapter 3, opens the ways to apply it directly to many other useful applications. In this chapter, we will present the human biometric recognition application of our presented pulse signal extraction method. We will compare our results with the Haque *et al.* method [23], which is the only method presented in the literature, for the human biometric based on non-contact pulse signal extraction from facial imagery.

4.1 Introduction

Human biometrics is the identification of a human based on his/her appearance or characteristics. The heartbeat signal is one possible biometrics traits, which can be very difficult to forge. The pulse signal of every person varies from others due to change in size, position, and anatomy of the heart, chest configuration, and various other factors [84]. The recent advancements in the contact-free heart rate measurement from human faces using a webcam created the possibility to extract a pulse signal from the videos and use it to recognize subjects. The extraction of a clean pulse signal is, however, a challenge and requires more robustness. The only work done in the area of contact-free human biometrics based on pulse signal that we are aware of was presented by Haque *at al.* [23]. They presented the heartbeat signal from facial videos as a new biometric trait to recognize human subjects. The results show the effectiveness of their developed method but the performance may be affected if it is

exposed to a noisy dataset in which illumination changes, subject's motion artifacts, and heart rate variations are expected.

4.1.1 Our Work

Haque *et al.* [23] method for the human biometrics recognition based on pulse signals obtained from facial videos performed well using the MAHNOB dataset [72]. The videos in the MAHNOB dataset were recorded under controlled conditions, i.e. little subject movement and controlled illumination conditions. The performance of their method can be affected if we implement it on a more challenging dataset, where the subject's motion is permitted and lighting condition varies while capturing the facial videos. We found in our experiments for the contact-free heart rate measurement from human face videos that the MAHNOB dataset contain videos of poor quality with the problem of dropping the frames over the duration of entire video. The pulse signals recorded were also not synchronised properly with the recorded videos. In our work, we have used a more challenging dataset and our proposed pulse signal extraction method presented in chapter 3. We have used Radon transform to extract features from the pulse signal and two machine learning techniques i.e. decision tree and artificial neural networks (ANNs) were employed to classify those features and identify the correct subjects.

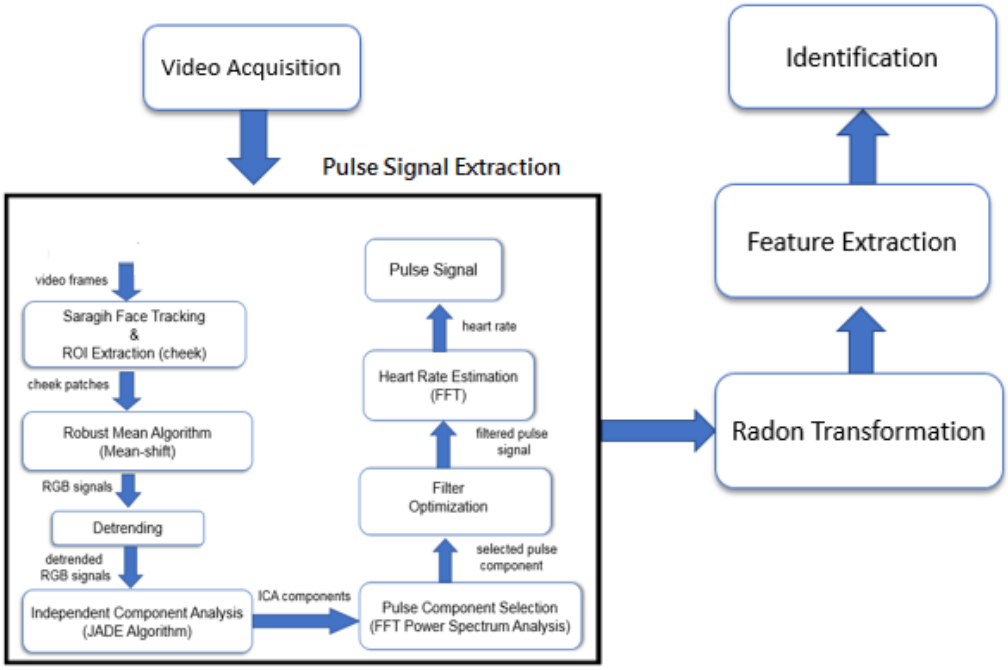


Fig. 4.1 Workflow of the method.

4.2 Method

In this section, we will describe the different processing steps involved in the proposed human biometric system based on the contact-free pulse signal extraction from human face videos. Fig. 4.1 shows the entire workflow of the method. The details of all these processing steps will be presented in the following sub-sections.

4.2.1 Video Acquisition and Face Tracking

We have recorded the color videos of the subjects using an ordinary webcam. The webcam used for the purpose of video acquisition was Logitech (C310), which has the following specifications: 0.3 Megapixels (Mp) video capture capability, 24 Bit RGB with 3 channels (8 Bit per channel), 15 frames per second (fps) with a pixel resolution of 640×480 pixels. The face area was focused while capturing the videos because the color of facial skin changes with respect to each heartbeat, and the RGB color sensor of a webcam can detect such subtle changes of blood volume in the facial area. This small reflected light color change derives the pulse signal, which we will use to detect human biometric. In order to continuously monitor such variations of skin color pixels, we need to track the skin area of the face from the recorded videos. In most of the previous work [10, 25], the OpenCV face detection method [53], which is based on the work of Viola and Jones [54] was employed to track the faces from the recorded videos. The problems of false positives (yielding multiple faces), false negatives (no face detected), and wide variations in how tight fitting the detected face box is, limits the performance of this method and therefore result in loss of data frames containing the face. We have used Saragih's face feature tracker [2] to track the faces from the recorded videos because it ensures no loss of frames for the detected face. We have found that this face tracker exhibits a 90 degrees freedom for the subject to move the face while they have been recorded. The cheeks of the subject were chosen as the region of interest (ROI) for the extraction of pulse signals because it tends to be a highly sensitive area of the face [94] and is least likely to be affected by a beard, hair, head covering and glasses. We extract a 40×40 pixels ROI from the face and one of the points of the Saragih's tracker was used to position it.

4.2.2 Pulse Signal Extraction

We have used our proposed method (explained in Chapter 3) for the extraction of pulse signals from the recorded videos. We briefly review the extraction of the pulse signal here to aid the reader, the full details can be found in chapter 3. We recorded

color videos of the subjects and track the faces from each frame of the recorded videos using Saragih's face tracker [2]. We were interested in the RGB color intensities of the skin pixels in the extracted ROI, therefore, calculated the mean of each color channel values of all ROI pixels in each frame using our robust mean algorithm. The three RGB signals that we have obtained were detrended to remove the uncertain variations in their mean values. The resulted normalized RGB signals were the mixture of the desired pulse signal and other noise sources. JADE algorithm of ICA [101] was employed to separate different signals mixed together and three independent source components were obtained, among which one of them will be the desired pulse signal. The fast Fourier transform (FFT) was used to obtain the power spectrum of all the extracted components, and the component with the highest peak within the set operational frequency qualifies for the next stage. We have optimized a filter using the expected pulse components and the ground truth pulse signals to obtain a clean pulse signal. Fig. 4.2 shows an example of the pulse signal extracted from the recorded video for one subject using our proposed pulse signal extraction from facial imagery method.

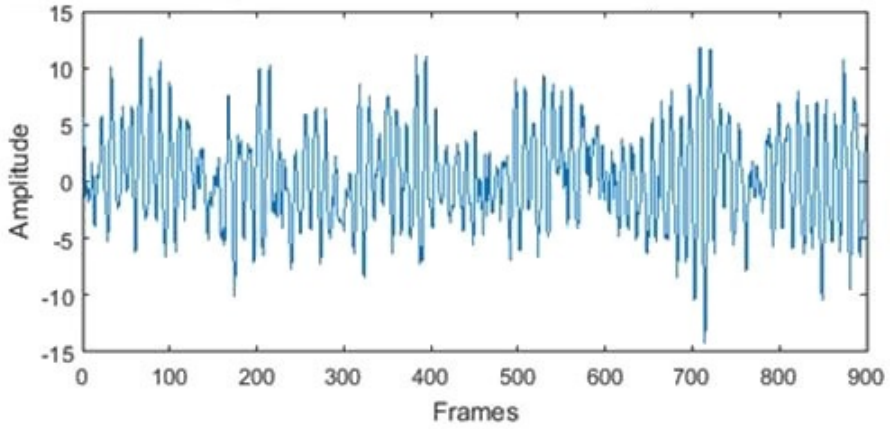


Fig. 4.2 An example of extracted pulse signal obtained from the facial video of a single subject using our proposed pulse signal extraction from facial imagery method.

4.2.3 Feature Extraction

We have extracted the features from the pulse signal by first transforming it into a Radon image [24]. Radon images were used in the Hedge *et al.* [86] ECG based biometric recognition method, and produced satisfactory results with the False Acceptance ratio of 2.19% and False Rejection ratio of 0.128%. The Radon images can be generated by the waterfall diagram of the pulse signals. The waterfall diagram can be

constructed by replicating the pulse signal. The number of replications depends on the number of frames in the video from which that pulse signal is extracted. For example, if we have to extract the pulse signal from the video of 60 seconds duration recorded at 15 frames per second, then the number of replications must be equal to $60 \times 15 = 900$. Fig. 4.3 shows different steps to obtain the Radon image of the pulse signal. The pulse signal is replicated and concatenated in the second dimension of the signal in order to create the waterfall diagram for 60 seconds long video captured at the speed of 15 frames per second. The Radon transform [89] was applied on to the generated waterfall diagram to obtain a two dimensional Radon image. The Radon transform is an integral transform which computes projections of an image matrix along specified directions. A projection of a two-dimensional image constitutes a set of line integrals. Let us consider *f (x,y)* as a two-dimensional image representing image intensity in the *(x,y)* coordinates. We can define the Radon transform *(R)* of the image, $R(\theta)[f(x,y)]$ as follows [23]:

$$R(\theta)[f(x,y)] = \int_{-\infty}^{\infty} f(x'cos\theta - y'sin\theta, x'sin\theta - y'cos\theta)dy' \tag{4.1}$$

where θ (θ = [0,180]) is the angle created by the distance vector of a line from the line integral with the relevant axis in the Radon space, and

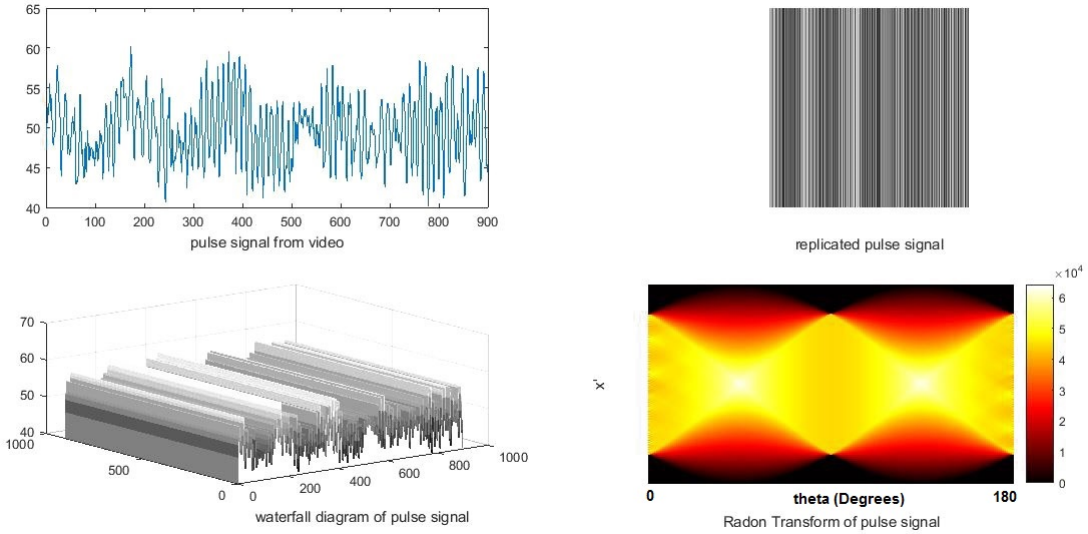


Fig. 4.3 Radon transform steps: Original pulse signal (top left) is replicated 900 times (top right), we obtain the waterfall diagram (bottom left), and finally generate the Radon transform image of the waterfall pulse signal (bottom right).

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{4.2}$$

We applied the Radon transform on the waterfall diagram of the pulse signal and obtained a two-dimensional Radon image shown in the bottom right of Fig. 4.3.

The features were extracted from the Radon transformed images for authentication by employing a pairwise distance method between every possible pair of pixels in the transformed image. The distance metric of Minkowski was used for this purpose. The following equation measure the pairwise distance of $m \times n$ pixels of the Radon image R :

$$d_{st} = \sqrt[p]{\sum_{i=1}^{n} |R_{si} - R_{ti}|^p} \tag{4.3}$$

where R_s and R_t are the row vectors representing each of the $m(1 \times n)$ row vectors of R, and p is a scalar parameter with $p = 2$. The feature vector obtained was then used for the identification of human subjects.

4.2.4 Identification

We have employed the Haque *et al.* [23] criteria for the identification of subjects using the feature vectors obtained from the Radon images of the pulse signals. The decision tree based method of [104] was utilized, which is a flowchart-like structure, consists of three types of components. The internal node represents a test on a feature, the branch component represents the outcome of the test, and the leaf node is a class label obtained as a decision after comparing all the features in the internal nodes. The tree was trained by utilizing the feature vectors data to split nodes and setup the decision tree. During the testing phase, the feature vector of a testing data passes through the tests in the nodes and finally gets a group label, where a group stands for a subject to be identified.

We extend the identification of subjects to the artificial neural networks (ANNs) technique of the machine learning. ANNs are biologically inspired computer programs designed to simulate the way in which the human brain processes information. ANNs gather their knowledge by detecting the patterns and relationships in data and learn (or are trained) through experience, not from programming. An ANN is formed from

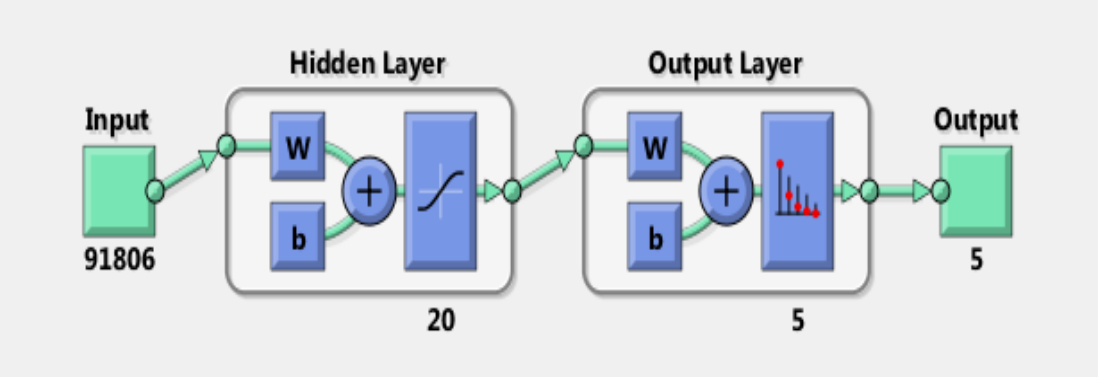


Fig. 4.4 Network structure.

hundreds of single units, artificial neurons or processing elements (PE), connected with coefficients (weights), which constitute the neural structure and are organized in layers [105]. Neural networks consist of input and output layers, as well as a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize. We have used pattern recognition application of the neural network toolbox developed by MATLAB [106]. In pattern recognition problems, we want a neural network to classify inputs into a set of target categories. The inputs to the network are the feature vectors obtained from the Radon images of the pulse signals. We create a two-layer (Hidden layer and Output layer) feed-forward network, with sigmoid hidden and softmax output neurons (patternnet) to classify feature vectors, given 20 neurons in its hidden layer. The network will be trained with scaled conjugate gradient backpropagation (trainscg). Fig. 4.4 shows the structure of the neural network created using 91806 features of 150 samples (30 samples per class) of 5 different classes of subjects. We used 70% of the total feature data for training and 30% data for testing in our experiments. The results obtained are presented in the following Section 4.4.

4.3 Experimental Setup

We have used our own dataset for the human biometric recognition experiments. The dataset consists of 5 different subjects of both genders (3 males, 2 females). Informed consent was obtained from all the subjects before the experiment and the study was approved by the Ethics Committee of our institute. All experiments were conducted indoors with the office fluorescent lights (one light was flickering randomly throughout the experiments) and indirect sunlight as the main sources of illumination and a

projector which reflected light making the lighting conditions more challenging. An ordinary Logitech webcam (C310) was used in our experiments. We played a video clip on the projector screen that contained different emotional cues (happy, sad, and scary videos) and asked the participant to view it while we recorded a color video of the subject's face. The happy clip was a video from a famous comedian Mr. Bean, the sad video clip chosen was the comparison between rich and poor people specifically focussed on poor and needy children, the scary video was an animated movie clip (San Andreas 2015) showing the destructions of a tsunami (links to the shown video clips is provided in Appendix B). The purpose of showing them the video clip was to try to vary their heart rates, add projector light to our lighting conditions, and to encourage the participants to move their facial features (eyes, nose, mouth, head etc.) due to their emotional expression response, to a certain degree. It should be noted that we are not doing emotion research, but only trying to make the conditions more challenging. The experimental scenario was hence made more realistic and challenging as we included extra lighting and movement artifacts. We have recorded a color video of each subject synchronized with the projector display (refresh rate of 15 fps). All recorded videos were saved in the .AVI format on a laptop under unique identifiers. We used a pulse oximeter Contec CMS50E (clipped to the participant's finger) in order to measure the ground truth heart rate at 60 Hz simultaneously for training and testing purposes. We have used 50 videos of 60 seconds each for 5 subjects (10 videos per subject) in our experiments. After the extraction of pulse signals from these 50 videos, we chopped them into 3 parts to obtain 150 samples (30 samples each subject). The 80 % of the total data was used for training and 20 % for testing purpose in a single-fold experiment for the evaluation of our biometric recognition system using decision tree method. We also test the performance of our system using k-fold cross-validation (k=5). In case of ANNs method of identification, we used 70% of the total feature data for training and 30% data for testing in our experiments.

4.4 Results

We have evaluated the performance of the proposed method (described in Section 4.2) for the human biometric recognition based on the contact-free pulse signal in the presence of challenging illumination interference and motion artifacts using the following statistics defined by Jain *et al.* [107]. These statistics are also used by Haque *et al.* [23] and we will compare our work with Haque *et al.* [23], therefore, utilizing the same statistics. These include False Positive Identification Rate (FPIR), False Negative

Identification Rate (FNIR), True Positive Identification Rate (TPIR). We can define these statistics as follows:

FPIR:

The FPIR can be defined as the probability of a test sample falsely identified as a subject. If TP = True Positive, TN = True Negative, FP = False Positive, and FN = False Negative identifications among N number of trials in an experiment, then we can define FPIR by the following equation.

$$FPIR = \frac{1}{N}\sum_{n=1}^{N}\frac{FP}{TP + TN + FP + FN} \tag{4.4}$$

FNIR:

The FNIR refers to the probability of a test sample falsely identified as a different subject. It can be defined by the following equation.

$$FNIR = \frac{1}{N}\sum_{n=1}^{N}\frac{FN}{TP + TN + FP + FN} \tag{4.5}$$

TPIR:

Using the above equation 4.5, we can also calculate the TPIR, which represents the overall identification performance of the proposed biometric method as follows.

$$TPIR = 1 - FNIR \tag{4.6}$$

We have also evaluated the proposed biometric system performance using four other metrics defined by [108], which includes precision, recall, specificity, and accuracy. These metrics can be defined as follows:

Precision:

Precision represents the total number of positive samples divided by the predicted positives. It can be calculated as follows.

$$Precision = \frac{1}{N}\sum_{n=1}^{N}\frac{TP}{TP + FP} \tag{4.7}$$

Recall:

Recall refers to the total number of positive samples divided by the actual positives in the experiment. It can be defined by the following equations.

$$Recall = \frac{1}{N}\sum_{n=1}^{N}\frac{TP}{TP + FN} \tag{4.8}$$

Specificity:
Specificity refers to the ratio of correctly identified negative samples with total number of negative samples and can be calculated using the following formula.

$$Specificity = \frac{1}{N}\sum_{n=1}^{N}\frac{TN}{TN + FP} \tag{4.9}$$

Accuracy:
Accuracy presents the ratio of correctly identified positive and negative samples with total number of positive and negative samples. This metric can be calculated by the following equation.

$$Accuracy = \frac{1}{N}\sum_{n=1}^{N}\frac{TP+TN}{TP+TN+FP+FN} \tag{4.10}$$

4.4.1 Decision Tree Results

We used 80% of the total feature data for training and 20% data for testing in a single-fold experiment for the evaluation of our biometric recognition system using decision tree method. The decision tree was developed from the training data and the authentication results were obtained using the testing data. Each test feature for every subject was compared in the decision tree to find the best match from the training set. Table 4.1 shows the confusion matrix for the identification of 150 samples for 5 subjects. The prefix 'S' represents the subject followed by the subject number. The values in the first diagonal (shown in bold) of the confusion matrix represents the true positive (*TP*) identifications. The false positive (*FP*) values are shown in each column (except the first diagonal value), and the values in each row represent the false negative (*FN*) identifications (excluding the first diagonal value). We can observe from the confusion matrix that a good number of true positive identifications are achieved for most of the subjects.

Table 4.2 presents the identification results obtained using the Haque *et al.* [23] method. We can observe that the confusion matrix shows almost the similar behavior as our proposed method results and a good number of true positive identifications are achieved for most of the subjects.

Table 4.1 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the decision tree based proposed human biometric method.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**23**	1	1	5	0
S2	0	**25**	0	5	0
S3	0	0	**24**	6	0
S4	0	0	0	**30**	0
S5	0	0	0	6	**24**

Table 4.2 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the decision tree based the Haque *et al.* [23] human biometric method.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**24**	0	0	6	0
S2	0	**27**	0	3	0
S3	0	0	**24**	6	0
S4	0	0	1	**29**	0
S5	1	0	1	5	**23**

Table 4.3 shows the confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals as an input to the proposed system.

Table 4.3 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the the decision tree based ground truth pulse signals.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**28**	0	1	1	0
S2	1	**28**	1	0	0
S3	1	0	**29**	0	0
S4	0	0	1	**29**	0
S5	1	0	3	0	**26**

Table 4.4 Comparison of the overall performance of the ground truth, Haque *et al.* [23] method and the proposed system for human biometrics recognition from facial imagery using decision trees.

Statistical Metrics	**Ground truth**	**Haque *et al.* [23]**	**Proposed method**
FPIR	1.33%	3.06%	3.20%
FNIR	1.33%	3.06%	3.20%
TPIR	98.70%	97.00%	96.80%
Precision rate	93.60%	89.40%	89.97%
Recall rate	93.33%	84.60%	84.00%
Specificity	98.33%	96.16%	96.00%
Accuracy	97.33%	93.86%	93.60%

Table 4.4 compares the overall performance of the ground truth, Haque *et al.* [23] method, and the proposed system evaluated using the statistical metrics defined above. We can see that the precision rate of both methods is touching 90% and the overall accuracy of above 90% is achieved. The results show that both methods are identifying the subjects effectively with a high accuracy.

k-Fold Cross-Validation (k=5):

We extend our experimental results by testing the proposed biometric recognition method based on contact-free pulse signal using the k-fold Cross-validation technique based on decision trees. The database explained in section 4.3 was used which consists of 50 videos of 60 seconds each for 5 subjects (10 videos per subject) in our experiment. After the extraction of pulse signals from these 50 videos, we chopped them into 3 parts to obtain 150 samples (30 samples each subject). These 150 samples are split into 5 folds and each fold contains 30 samples. Table 4.5 shows the confusion matrix for the identification of 150 samples for 5 subjects. using the proposed biometric method. Table 4.6 presents the identification results obtained using the Haque *et al.* [23] method. Table 4.7 shows the confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals. Table 4.8 compares the overall performance of the ground truth, Haque *et al.* [23] method, and the proposed system based on k-fold cross validation decision trees evaluated using the statistical metrics defined above.

Table 4.5 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the proposed human biometric method and k-fold Cross-validation decision trees technique (k=5).

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**30**	0	0	0	0
S2	3	**27**	0	0	0
S3	7	0	**23**	0	0
S4	8	0	0	**22**	0
S5	5	0	0	0	**25**

Table 4.6 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the Haque *et al.* [23] human biometric method and k-fold Cross-validation decision trees technique (k=5).

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**30**	0	0	0	0
S2	2	**28**	0	0	0
S3	7	0	**23**	0	0
S4	8	0	0	**22**	0
S5	4	1	0	0	**25**

Table 4.7 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals using k-fold Cross-validation decision trees technique (k=5).

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**27**	1	0	1	1
S2	0	**29**	0	0	1
S3	0	0	**30**	0	0
S4	1	0	1	**28**	0
S5	0	0	0	0	**30**

We can find from the Table 4.8 that the k-fold cross-validation technique has almost the same overall performance for both methods i.e. Haque *et al.* and our proposed method. This validates the efficiency of our biometric identification system.

Table 4.8 Comparison of the overall performance of the ground truth, Haque *et al.* [23] method and the proposed system for human biometrics recognition from facial imagery using k-fold Cross-validation (k=5) based on decision trees.

Statistical Metrics	**Ground truth**	**Haque *et al.* [23]**	**Proposed method**
FPIR	0.80%	2.93%	3.00%
FNIR	0.80%	2.93%	3.00%
TPIR	99.20%	97.10%	97.00%
Precision rate	95.98%	91.06%	91.32%
Recall rate	96.00%	85.33%	84.66%
Specificity	99.00%	96.33%	96.16%
Accuracy	98.40%	94.13%	93.86%

4.4.2 ANN Results

In the following section, we will present results obtained for the human biometric recognition based on the ANNs classification method. The ANNs classification technique will be applied on both Haque *et al.* [23] method and our proposed method. The results obtained will be compared to see the behaviour of the pulse signals obtained from facial videos and how well they can perform in the human biometric recognition system. Table 4.9 presents the confusion matrix for the identification of 150 samples for 5 subjects using the proposed method based on the ANNs technique of the neural networks. The prefix 'S' represents the subject followed by the subject number. The values in the first diagonal (shown in bold) of the confusion matrix represents the true positive (*TP*) identifications. The false positive (*FP*) values are shown in each column (except the first diagonal value), and the values in each row represent the false negative (*FN*) identifications (excluding the first diagonal value). Table 4.10 shows the results obtained by Haque *et al.* [23] method using the ANNs identification technique instead of a decision tree. The extraction of pulse signals from facial video and the feature extraction methods remain the same as obtained by Haque *et al.* [23]. Table

4.11 shows the confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals.

Table 4.9 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the proposed human recognition method based on ANNs.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**25**	1	4	0	0
S2	2	**25**	3	0	0
S3	5	1	**21**	3	0
S4	3	0	4	**23**	0
S5	1	0	0	0	**29**

Table 4.10 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the Haque *et al.* [23] human recognition method based on ANNs.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**20**	3	4	0	3
S2	4	**20**	3	3	0
S3	5	2	**20**	0	3
S4	1	3	2	**23**	1
S5	3	4	5	0	**18**

Table 4.8 compares the overall performance of the ground truth, Haque *et al.* [23] method, and the proposed system based on the ANNs, evaluated using the statistical metrics defined above. We can observe that our proposed method performed better than

Table 4.11 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals based on ANNs method.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**22**	0	1	6	1
S2	0	**30**	0	0	0
S3	0	0	**29**	1	0
S4	0	0	0	**30**	0
S5	0	1	0	0	**29**

Table 4.12 Comparison of the overall performance of the Haque *et al.* [23] method and proposed system using ANNs.

Statistical Metrics	**Ground truth**	**Haque *et al.* [23]**	**Proposed method**
FPIR	1.33%	6.50%	3.60%
FNIR	1.33%	6.50%	3.60%
TPIR	98.7%	93.50%	96.40%
Precision rate	94.18%	68.40%	83.18%
Recall rate	93.33%	67.33%	82.00%
Specificity	98.33%	91.83%	95.50%
Accuracy	97.33%	86.93%	92.80%

the Haque *et al.* [23] method. The overall system accuracy of 92.80% of the proposed method is higher than the 86.93% of the Haque *et al.* [23] method. We can, therefore,

draw a conclusion that the pulse signals obtained from the facial video using our proposed method performs better than those obtained by the Haque *et al.* [23] method. This validates the robustness of our proposed contact-free pulse signal extraction method presented in chapter 3.

4.4.3 Original Pulse Signal Based Identification Results

We perform another experiment to test the behaviour of the original pulse signals as an input to the biometric identification system instead of the features extracted from the Radon images of the pulse signals. We applied the k-fold (k=5) cross-validation decision trees technique to test the system. The same dataset of 150 samples of 5 different subjects (explained in section 4.3) was used and all the processing steps of our proposed method (described in section 4.2), except the Radon transformation step (section 4.2.3), are performed. Table 4.13 shows the confusion matrix for the identification of 150 samples for 5 subjects. using the proposed biometric method. Table 4.14 presents the identification results obtained using the Haque *et al.* [23] method. Table 4.15 shows the confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals. Table 4.16 compares the overall performance of the ground truth, Haque *et al.* [23] method, and the proposed system based on k-fold cross validation decision trees evaluated using the statistical metrics defined above.

Table 4.13 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the proposed human recognition method based on k-fold cross-validation decision trees.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**30**	0	0	0	0
S2	3	**27**	0	0	0
S3	7	0	**23**	0	0
S4	7	0	0	**23**	0
S5	5	0	0	0	**25**

Table 4.14 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the Haque *et al.* [23] human recognition method based on k-fold cross-validation decision trees.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**30**	0	0	0	0
S2	2	**28**	0	0	0
S3	7	0	**23**	0	0
S4	8	0	0	**22**	0
S5	5	0	0	0	**25**

Table 4.15 Confusion matrix for identification of 150 samples of 5 subjects (30 samples each) using the ground truth pulse signals based on k-fold cross-validation decision trees.

Subjects	**S1**	**S2**	**S3**	**S4**	**S5**
S1	**24**	2	0	4	0
S2	1	**29**	0	0	0
S3	0	1	**24**	3	2
S4	1	0	4	**24**	1
S5	1	3	0	0	**26**

We can find from the results shown in Table 4.16 that the original pulse signal based biometric identification performed equally better as compared to the Radon images feature based identification method. We can conclude that the transformation of pulse signals extracted from facial images into Radon images and extract features

Table 4.16 Comparison of the overall performance of the Haque *et al.* [23] method and proposed system using k-fold cross-validation decision trees.

Statistical Metrics	**Ground truth**	**Haque *et al.* [23]**	**Proposed method**
FPIR	3.06%	2.93%	2.93%
FNIR	3.06%	2.93%	2.93%
TPIR	96.93%	97.07%	97.07%
Precision rate	84.86%	91.52%	91.52%
Recall rate	84.66%	85.33%	85.33%
Specificity	96.16%	96.33%	96.33%
Accuracy	93.86%	94.13%	94.13%

from those images has no significant impact on the improvement of human biometric system.

4.5 Discussion

The use of pulse signals obtained from facial videos for a biometric identification purpose is a challenging task, specifically if obtained from the capturing scenario where motion and illumination artifacts can be expected. The extraction of a clean pulse signal from facial imagery is still a problem in realistic conditions. Haque *et al.* [23] attempted to use contact-free pulse signals from videos as a biometric trait and they presented satisfactory results. It should be noted that the touch-based pulse signal extraction such as from ECG [87] or PPG [109] produced an accuracy of over 90% on some local databases. The identification results of our system are also above 90% accuracy line, which is quite competitive to the contact-based heartbeat signal biometric recognition methods. However, the contact-free way of human biometrics system from pulse signal make our method more comfortable and user-friendly. The accuracy could possibly be further improved by employing different

machine learning techniques to extract features from the heartbeat signal and to provide a better classification of those features. We have performed two experiments for the human biometric recognition based on the pulse signals obtained from the facial videos. In the first experiment, we use our local dataset which was captured in challenging conditions where illumination variations and motion noise was present. We use our contact-free pulse signal extraction from facial videos method explained in chapter 3. The pulse signals obtained were then passed on to the biometric identification system as presented by Haque *et al.* [23]. The features obtained from the Radon images of the pulse signals were classified using decision tree and the overall system performance was evaluated. We compare our results with the results obtained by the Haque *et al.* [23] method and found that we achieved almost the same accuracy. The overall system performance was above 90%, which is quite satisfactory. In the second experiment, we tested the behaviour of the same data using a different machine learning classification technique. We classified the features using ANNs and found that our proposed method performed better than the Haque *et al.* [23] method. We achieved an accuracy of 92.80% which is better than the 86.93% accuracy of Haque *et al.* [23]. We can conclude that the pulse signals obtained using our pulse signal extraction method performed better than the Haque *et al.* [23] method in case of ANNs based human biometric identification. In order to further test the robustness of the pulse signal extraction method, we need to test these signals using other machine learning techniques for human biometrics recognition. The feature extraction process from pulse signals could also be improved to achieve better identification results. The human biometric recognition system based on contact-free pulse signals from facial videos can be improved by improving the contact-free pulse signal methods. We also extend our experiment and test the biometric system using the original pulse signals without transforming them into Radaon images and found the satisfactory results. We can conclude from that experiment that Radon images transformation of original pulse signals obtained from facial videos did not have significant impact on the overall results. We, therefore, need more robust feature extraction mechanism and more powerful machine learning algorithm to make the identification method more efficient.

4.6 Conclusion

We have presented a method for the biometric recognition of a person based on the contact-free pulse signal extracted from facial videos. A dataset was created

in the presence of motion and illumination artifacts to make the capturing scenario more challenging and realistic. The pulse signal was extracted using our proposed contact-free heart rate measurement from human face videos using a webcam, which is capable of extracting a cleaner pulse signal in realistic conditions. The heartbeat signal extraction method utilizes the Saragih's face tracker, the robust mean color values algorithm, and filter optimization technique to cleanly obtain the pulse signal which is important to use heartbeat as a human biometric trait. Radon transformation was applied to the waterfall diagram of the pulse signals. The pairwise distance metric of Minkowski was used to extract features from the radon images of the heartbeat signals. The decision tree and ANNs machine learning techniques were employed for identification of subjects and we found the behaviour of the pulse signals extracted better in our proposed method using the ANNs identification method. The accuracy of almost 90% was obtained and considering the contact-free nature of the method such accuracy is quite acceptable. This way of recognizing the human biometrics can be integrated into many other daily life activities with more ease and comfort.

4.7 Summary

In this chapter, we presented a method for the human biometrics using the non-contact pulse signal extracted from facial imagery. We introduced human biometrics followed by the importance of it and the current limitations in this evolving technology. Contact-free heartbeat signal used as a biometric trait was introduced and the importance of implementing this contact-free way of recognizing humans in a realistic situation was highlighted. We presented a biometric recognition method using our local dataset captured in challenging conditions. The robustness of pulse signal extraction from human face videos was already evaluated in chapter 3, and was used to extract a cleaner pulse signal for the identification of human subjects. The Radon transform images of the pulse signals were used for the extraction of features, and the decision tree and ANNs machine learning techniques were employed to classify those features for identification purpose. The results show satisfactory performance in the human biometric recognition system using a contact-free heartbeat signal. Results imply the cleaner pulse signal provided by our approach can improve performance significantly for some machine learning approaches. In order to get further improvement, we need to improve the feature extraction from pulse signals and try more sophisticated machine learning techniques for identification of subjects.

Chapter 5

Conclusions

In this chapter, we summarise the content of the thesis, list the main contributions of our work and propose future directions.

5.1 Summary of the Thesis

The heart is the most important muscular organ of a human body and its strength can be assessed by the rate at which it beats. The continuous monitoring of heart rate is very important for appropriate health treatment. The heart rate has been measured from over two thousand years ago when it was assessed by the traditional way by placing the two fingers over the subject's radial arteries. The technological era of the digital world measured the heart rate using different sensory equipment such as ECG machine, pulse oximeter, Doppler probe, etc. The performance of these types of machinery is quite satisfactory but the ease and comfort is sacrificed. The contact-free way to measure heart rate can make it more comfortable and applicable in many other daily life applications. Several attempts have been made over the past decade to estimate the heart rate from videos captured by means of a thermal camera, CCD camera, and a simple computer webcam. The use of a webcam to measure pulse rate has become popular due to its easy and affordable usage. In this thesis, we have introduced different methods for the measurement of the heart rate from human face imagery using an ordinary webcam. The techniques employed in these methods have been described, the performance was evaluated, and the limitations have been highlighted. The two main problems of almost all the previous work are the subject's movement artifacts and illumination variations causing noise. We attempt to minimize the effect of these problems in our work and to make the contact-free heart rate measurement more robust so that it can be integrated more reliably into other applications. We

re-implemented the existing heart rate measurement methods and by refining certain aspects to try to get them improved. We presented our own method to compensate the aforementioned problems of heart rate estimation systems and extract a cleaner pulse signal. The extracted pulse signal from human face videos was then used to recognize human biometrics and to validate the feasibility and comfort of contact-free heart rate measurement as example application.

In the literature review (Chapter 2), we briefly described the heart rate measurement methods in practice and their limitations have been highlighted. The detailed literature review of the contact-free pulse rate estimation methods has been conducted and the limitations of those methods were described. Existing challenges in this area, including motion artifacts and illumination noise, have been addressed. The biometric recognition application of the non-contact pulse signal extraction from human face videos was also described in detail.

After the identification of the existing challenges in the contact-free heart rate estimation, we attempted to propose a solution to overcome them. In chapter 3, we introduced the use of Saragih's face tracker [91] for the first time in heart rate estimation to minimize the effect of motion artifacts. We also have proposed the robust mean algorithm to calculate the average value of the skin color pixels on the face, which has attempted to ignore the non-skin pixels color values and minimize the effect of illumination variations. We have optimized a filter to increase the accuracy of the heart rate measurement system and to extract a cleaner pulse signal, which can be integrated into other applications. The results have been compared to the ground truth and other state-of-the-art methods to evaluate the performance of the proposed system. The dataset used in our experiments have been made challenging to ensure the useful applications of the proposed system in the real world. The results have shown significant improvement in the contact-free heart rate estimation methods.

The proposed heart rate extraction method from human face videos can have many applications. We have presented a human biometric recognition application of our work in chapter 4. The extracted pulse signals have been transformed into the radon images and the features have been extracted from those images to classify them. The decision tree and ANNs machine learning techniques have been used to identify the correct subjects. The accuracy of 90% has been achieved in challenging experimental conditions, which is quite satisfactory. We have concluded that the heartbeat signal if extracted cleanly from the human face videos can have many useful applications in our daily life activities.

5.2 Contributions

The contributions of our work presented in this thesis are:

- Different modalities of heart rate measurement were critically discussed in terms of their efficiency, applications, affordability, advantages, and disadvantages. This includes sphygmology, phonocardiogram (PCG), pulse oximetry, and electrocardiogram (ECG) (Chapter 2).

- An extensive literature review of existing contact-free heart rate measurement methods [46, 9, 47, 48, 50, 26] and the human biometric application of the extracted pulse signal was conducted. The most famous and promising methods [10–17, 20, 21][24, 25][52] for the estimation of pulse rate using an ordinary webcam were described. From a thorough review, we were able to find out the common problems of most of these methods which include, face tracking failure, subject's movement, illumination variations, and other noisy artifacts of the realistic conditions (Chapter 2).

- We have improved the performance of the state-of-the-art methods [10] [24–26] for estimation of heart rate using a webcam by re-implementing them using additional parameters and challenging capturing scenarios. We have introduced the use of Saragih face tracking technique [2] for accurately track the faces from the captured face videos and minimize the effect of subject's motion (it should be noted that Saragih face feature tracker has been used in our work for the first time as per our knowledge for the estimation of heart rate from human face videos), robust mean algorithm based on the Mean Shift algorithm [95] was applied to calculate the mean color values of the skin color pixels which attempt to minimize the effect of illumination, detrending was applied to the color channels to reduce the uncertain variations in their mean values in each frame of the recorded video caused due to lighting reflection. Chapter 3 shows the improvement in the results obtained after the addition of the above-mentioned useful parameters.

- We have developed a method for the contact-free heart rate measurement from human face videos using a webcam under more challenging and realistic capturing scenarios. The proposed method has the potential to overcome many of the problems of subject's movement and light reflection. We find the robust mean of the skin pixel's color values and calculate the least-squares error optimal filter using our training dataset to estimate the heart rate. Chapter 3 presents the

performance evaluation of the proposed method as compared to the reference ground truth and other state-of-the-art methods [10] [24–26] for the estimation of heart rate from human faces using an ordinary webcam. This method not only improved the efficiency of heart rate measurement systems but also resulted in extraction of a cleaner pulse signal, which can be integrated into many other real-world applications such as human biometric recognition.

- We have developed a method for the human biometric recognition based on the pulse signals extracted from the facial videos. The pulse signal was extracted using our proposed contact-free heart rate measurement from human face videos using a webcam (Chapter 3), which is capable of extracting a cleaner pulse signal in realistic conditions. The heartbeat signal extraction method utilizes the Saragih's face tracker, the robust mean color values algorithm, and filter optimization technique to cleanly obtain the pulse signal, which is important to use heartbeat as a human biometric trait. Radon transformation was applied to the waterfall diagram of the pulse signals. The pairwise distance metric of Minkowski was used to extract features from the radon images of the heartbeat signals. The decision tree and ANNs machine learning techniques were employed to classify them and identify the correct subject. The accuracy of almost 90% was obtained, and considering the contact-free nature of the method, such accuracy is quite acceptable (Chapter 4). After comparing our proposed method with the Haque *et al.* [23] method, we achieved almost the same results in case of decision tree classification technique but get improved results as we apply ANNs classification technique and therefore conclude that the behaviour of pulse signals obtained by our proposed contact-free pulse signal extraction method performed better than Haque *et al.* [23] method for the purpose of human biometric recognition.

5.3 Future Work

Based on the work presented in this thesis, a list of future directions are proposed as follows:

- A comparison between our contact-free heart rate measurement method from human face videos using a webcam and other popular pulse rate estimation methods reviewed in Section 2.4 could be conducted.

- We can further improve our heart rate estimation method by employing additional parameters into it such as the selection criteria for the expected pulse component can be further investigated using machine learning techniques to correctly identify the desired pulse component in the post-ICA stage, the near-infrared cameras can be used along with the webcam to cleanly extract a pulse signal of a dark skin color subject, and further pre-processing of the raw RGB signals in the pre-ICA stage could be useful.
- We will implement our pulse rate estimation method on a large dataset under more challenging and realistic conditions.
- The human biometric application of our extracted pulse signal from facial videos can be improved by employing different techniques of machine learning. We can extend our investigation to deep learning to test the behavior of the system.
- We could try to find the human emotions and extend our work to lie detection based on the contact-free pulse signal extracted from the facial imagery.
- Furthermore, we will try to implement our work in the real-time scenarios such as in hospitals, in vehicles, and other HCI environments, and compare the results with the ground truth to make it applicable in many other daily life activities facilitating ease and comfort.

Appendix A

Glossary of Terms

Acronyms

bpm	beats per minute
fps	frames per second
kNN	k-nearest neighbor
ANNs	artificial neural networks
ARMA	Autoregressive Moving Average
BSS	blind source separation
BVP	blood volume pulse
CCD	Charge Coupled Device
CT	computed tomography
DFT	discrete-time Fourier transform
DRMF	Discriminative Response Map Fitting
DTW	Dynamic Time Warping
ECG	Electrocardiogram
FFT	fast Fourier transform
FLDA	Fisher's Linear Discriminant Analysis
FN	False Negative
FNIR	False Negative Identification Rate

FP False Positive
FPIR False Positive Identification Rate
FQA Face Quality Assessment

GFT Good Feature to Track

HR heart rate
Hz hertz

IBIs interbeat intervals
ICA Independent Component Analysis

JADE Joint Approximate Diagonalization of Eigenmatrices

KLT Kanade-Lucas-Tomasi

LE Laplacian Eigenmap

MRI Magnetic Resonance Imaging
MT measurement tracker

NLMS Normalised least mean square

OpenCV Open Computer Vision

PCA Principle Component Analysis
PCG Phonocardiogram
PE processing elements
PPG Photoplethysmography
PSD Power Spectral Density

RGB red, green, blue
RMSE root mean square error
ROI region of interest

SDM supervised descent method

SIMCA	Soft Independent Modeling of Class Analogy
SNR	signal-to-noise ratio
STFT	short-time Fourier transform
SVD	singular value decomposition
TPIR	True Positive Identification Rate
TP	True Positive
TT	tandem tracker
TN	True Negative

Appendix B

Links to the Emotional Content Video Clips

Happy Clip:

https://www.youtube.com/watch?v=OoI57NeMwCc

Sad Clip:

https://www.youtube.com/watch?v=-9pNeT55U_c

Scary Clip:

https://www.youtube.com/watch?v=r_E7yrCAq58

References

[1] Best Life. 10 Surefire Signs Your Heart Is Super Strong. https://bestlifeonline.com/signs-healthy-heart/. [accessed September 10, 2018].

[2] Jason M. Saragih, Simon Lucey, and Jeffrey F. Cohn. Face alignment through subspace constrained mean-shifts. *2009 IEEE 12th International Conference on Computer Vision*, pages 1034–1041, 2009.

[3] Wikimedia Commons. Diagram of the human heart, created by wapcaplet in sodipodi. cropped by yaddah to remove white space. https://commons.wikimedia.org/wiki/File:Diagram_of_the_human_heart_(cropped).svg, 2006. File:Diagram of the human heart (cropped).svg.

[4] Wellcome Collection. An ayurvedic medical practitioner taking the pulse. watercolour, ca. 1825. https://wellcomecollection.org/works/nzrgnx69.

[5] Tamal Chakrabarti, Sourav Saha, Sathi Roy, and Ishita Chel. Phonocardiogram signal analysis - practices, trends and challenges: A critical review. *2015 International Conference and Workshop on Computing and Communication (IEMCON)*, 2015.

[6] Wikimedia Commons. Picture of a wrist mounted pulse oxymeter. https://commons.wikimedia.org/wiki/File:Wrist-oximeter.jpg, 2011. File:Wrist-oximeter.jpg.

[7] Wikimedia Commons. Schematic diagram of normal sinus rhythm for a human heart as seen on ecg (with english labels). https://commons.wikimedia.org/wiki/File:SinusRhythmLabels.svg, 2007. File:SinusRhythmLabels.svg.

[8] Wikimedia Commons. The picture shows the standard setup for an ekg. https://commons.wikimedia.org/wiki/File:Ekg_NIH.jpg, 2013. File:Ekg NIH.jpg.

[9] Ioannis T Pavlidis, Jonathan Baldwin Dowdall, Nanfei Sun, Colin Puri, Jin Fei, and Marc Garbey. Interacting with human physiology. *Computer Vision and Image Understanding*, 108:150–170, 2007.

[10] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010.

[11] Thomas Pursche, Jarek Krajewski, and Reinhard Moeller. Video-based heart rate measurement from human faces. *In IEEE International Conference on Consumer Electronics (ICCE)*, pages 544–545, 2012.

[12] Kual-Zheng Lee, Pang-Chan Hung, and Luo-Wei Tsai. Contact-free heart rate measurement using a camera. *2012 Ninth Conference on Computer and Robot Vision (CRV)*, pages 147–152, 2012.

[13] Lan Wei, Yonghong Tain, Yaowei Wang, Touradj Ebrahimi, and Tiejun Huang. Automatic webcam-based human heart rate measurements using laplacian eigenmap. *Asian Conference on Computer Vision*, pages 281–292, 2012.

[14] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4), 2012.

[15] Yong-Poh Yu, Ban-Hoe Kwan, Chern-Loon Lim, Siaw-Lang Wong, and Raveendran Paramesran. Video-based heart rate measurement using short-time fourier transform. *International Symposium on Intelligent Signal Processing and Communication Systems*, pages 704–707, 2013.

[16] Hamed Monkaresi, Rafael A. Calvo, and Hong Yan. A machine learning approach to improve contactless heart rate monitoring using a webcam. *IEEE Journal of Biomedical and Health Informatics*, 18(4):1153–1160, 2014.

[17] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikäinen. Remote heart rate measurement from face videos under realistic situations. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4264–4271, 2014.

[18] Frederic Bousefsaf1, Choubeila Maaoui1, and Alain Pruski1. Automatic selection of webcam photoplethysmographic pixels based on lightness criteria. *Journal of Medical and Biological Engineering*, 37(3):374–385, 2017.

[19] Lai-Man Po, Feng Litong, Li Yuming, Xu Xuyuan, Cheung Terence Chun-Ho, and Cheung Kwok-Wai. Block-based adaptive roi for remote photoplethysmography. *Multimedia Tools and Applications*, 77(6):6503–6529, 2018.

[20] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3430–3437, 2013.

[21] Mohammad A. Haque, Ramin Irani, Kamal Nasrollahi, and Thomas B. Moeslund. Heartbeat rate measurement from facial video. *IEEE Intelligent Systems*, 31(3):40–48, 2016.

[22] Venkatesh N and Srinivasan Jayaraman. Human electrocardiogram for biometrics using dtw and flda. *2010 International Conference on Pattern Recognition (ICPR)*, pages 3838–3841, 2010.

[23] Mohammad Ahsanul Haque, Kamal Nasrollahi, and Thomas B. Moeslund. Heartbeat signal from facial video for biometric recognition. *In Image Analysis: 19th Scandinavian Conference, SCIA 2015*, pages 165–174, 2015.

[24] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011.

[25] Sungjun Kwon, Hyunseok Kim, and Kwang Suk Park. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),*, pages 2174–2177, 2012.

[26] Gerard de Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[27] NHS. Electrocardiogram (ECG). https://www.nhs.uk/conditions/electrocardiogram/. [accessed March 30, 2017].

[28] Wikipedia. Pulse Oximetry. https://en.wikipedia.org/wiki/Pulse_oximetry. [accessed April 19, 2017].

[29] The Free Dictionary. Doppler probe. https://medical-dictionary.thefreedictionary.com/Doppler+probe. [accessed March 20, 2017].

[30] David L. Schriger. *Cecil Medicine, 24th ed.* Elsevier, 2007.

[31] New Health Advisor. Diagram of Human Heart and Blood Circulation in It. http://www.newhealthadvisor.com/Heart-Diagram-Labeled.html. [accessed March 10, 2017].

[32] British Heart Foundation. Your heart rate. https://www.bhf.org.uk/publications/heart-conditions/medical-information-sheets/your-heart-rate. [accessed May 20, 2017].

[33] Valerie C. Scanlon and Tina Sanders. *Essentials of Anatomy and Physiology*. F.A. Davis Company, 2007.

[34] Hirofumi Tanaka, Kevin D. Monahan, and Douglas R. Seals. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156, 2001.

[35] Decoded past. Sphygmology: Measuring the Pulses in Traditional Chinese Medicine And Beyond. http://decodedpast.com/taking-the-pulse/2144. [accessed March 20, 2017].

[36] R.M.Marchant-Forde, D.J.Marlin, and J.N.Marchant-Forde. Validation of a cardiac monitor for measuring heart rate variability in adult female pigs: accuracy, artefacts and editing. *Physiology Behavior*, 80(4):449–458, 2004.

[37] COPD. Clinical use of Pulse Oximetry. http://www.copd-alert.com/OximetryPG.pdf. [accessed April 20, 2017].

[38] Geaorge Mardirossian and Ronald E. Schneider. Limitations of pulse oximetry. *Anesthesia Progress*, 39(6):194–196, 1992.

[39] CONTEC. CMS50E Pulse Oximeter. http://www.contecmed.com/index.php?page=shop.product_details&flypage=flypage.tpl&product_id=186&category_id=7&option=com_virtuemart&Itemid=592. [accessed March 15, 2017].

[40] Wearable. FITNESS TRACKER FEATURE . https://www.wareable.com/fitness-trackers/how-your-fitness-tracker-works-1449. [accessed Decemeber 10, 2018].

[41] Wikimedia Common. Schematic diagram of normal sinus rhythm for a human heart as seen on ECG. https://commons.wikimedia.org/wiki/File%3ASinusRhythmLabels.svg. [accessed April 30, 2017].

[42] HCE. ECG machine. https://www.hce-uk.com/epages/HCE-UK.sf/en_GB/?ObjectID=10993&ViewAction=FacetedSearchProducts&SearchString=ECG+machine. [accessed May 10, 2017].

[43] A. B. Hertzman and C. R. Spealman. Observations on the finger volume pulse recorded photo-electrically. *American Journal of Physiology*, 119:334–335, 1937.

[44] Wim Verkruysse, Lars O. Svaasand, and J. Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.

[45] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):1–39, 2007.

[46] Ioannis Pavlidis. Continuous physiological monitoring. *Engineering in Medicine and Biology Society, 2003. Proceedings of the 25th Annual International Conference of the IEEE*, pages 1084–1087, 2003.

[47] Marc Garbey, Nanfei Sun, Arcangelo Merla, and Ioannis Pavlidis. Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. *IEEE Transactions on Biomedical Engineering*, 54(8):1418–1426, 2007.

[48] Rodolfo Gabriel Gatto. *Estimation of instantaneous heart rate using video infrared thermography and ARMA models*. ProQuest Dissertations Publishing, 2009.

[49] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung. *Time Series Analysis. Forecasting and control, Fifth Edition*. John Wiley and sons, 2016.

[50] Chihiro Takano and Yuji Ohta. Heart rate measurement based on a time-lapse image. *Medical Engineering and Physics*, 29(8):853–857, 2007.

[51] Spectral Instruments Inc. What is a CCD? http://www.specinst.com/What_Is_A_CCD.html. [accessed June 10, 2017].

[52] Magdalena Lewandowska, Jacek Rumiński, and Tomasz Kocejko. Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity. *Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011.

[53] OpenCV. Face Detection using Haar Cascades. https://docs.opencv.org/3.3.0/d7/d8b/tutorial_py_face_detection.html. [accessed June 10, 2017].

[54] Paul Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[55] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[56] Jean-François Cardoso. High-order contrasts for independent component analysis (ica). *Neural Computation*, 11(1):157–192, 1999.

[57] Perttu O.Ranta-aho Mika P. Tarvainen and Pasi A. Karjalainen. An advanced detrending method with application to hrv analysis. *IEEE Transaction on Biomedical Engineering*, 49(2):172–175, 2002.

[58] Jonathon Shlens. A tutorial on principal component analysis. *International Journal of Remote Sensing*, 51(2), 2014.

[59] Weijia Cui and Lee E. Ostrander. In vivo reflectance of blood and tissue as a function of light wavelength. *IEEE Transactions on Biomedical Engineering*, 37(6):632–639, 1990.

[60] Kishor K. Bhoyar and Omprakash G. Kakde. Skin color detection model using neural networks and its performance evaluation. *Journal of Computer Science*, 6(9):963–968, 2010.

[61] Linda G. Shapiro and George C. Stockman. *Computer Vision*. Pearson, 2001.

[62] Dorin Comaniciu, Visvanathan Ramesh, and Peter Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[63] Boualem Boashash. *Time-Frequency Signal Analysis and Processing- A Comprehensive Reference*. Elsevier Science, 2003.

[64] Edwin H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–129, 1997.

[65] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[66] Jont B. Allen. Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(3):235–238, 1977.

[67] David W. Aha, Dennis Kibler, and Marc K. Albert. Instance-based learning algorithms. *Machine Learning*, 6(1):37–66, 1991.

[68] Simon O. Haykin. *Adaptive Filter Theory*. Prentice Hall, 2002.

[69] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic. Robust discriminative response map fitting with constrained local models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3444–3451, 2013.

[70] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. *Technical Report CMU-CS-91-132*, 1991.

[71] Peter D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967.

[72] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012.

[73] Kamal Nasrollahi and Thomas B. Moeslund. Extracting a good quality frontal face image from a low-resolution video sequence. *IEEE Transactions on Circuits and Systems for Video Technology*, 21(10):1353–1362, 2011.

[74] Xuehan Xiong and Fernando De la Torre. Supervised descent method and its applications to face alignment. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 532–539, 2013.

[75] TechTarget. biometrics. http://searchsecurity.techtarget.com/definition/biometrics. [accessed September 15, 2016].

[76] Anil K. Jain, Patrick Flynn, and Arun A. Ross. *Handbook of Biometrics*. Springer, 2008.

[77] Wageeh W. Boles. A security system based on human iris identification using wavelet transform. *First International Conference Knowledge-Based Intelligent Electron Systems*, 2:533–541, 1997.

[78] Wageeh W. Boles and S. Y. T. Chu. Personal identification using images of the human palm. *TENCON '97. IEEE Region 10 Annual Conference. Speech and Image Technologies for Computing and Telecommunications*, 2:295–298, 1997.

[79] Ashok Samal and Prasana A.Iyengar. Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition*, 25(1):65–77, 1992.

[80] Richard Jiang, Somaya Al-maadeed, Ahmed Bouridane, Prof. Danny Crookes, and Azeddine Beghdadi. *Biometric Security and Privacy: Opportunities Challenges in The Big Data Era*. Springer International Publishing Switzerland 2017, 2017.

[81] Kresimir Delac and Mislav Grgic. A survey of biometric recognition methods. *Elmar-2004. 46th International Symposium on Electronics in Marine*, pages 184–193, 2004.

[82] Abhinav Gupta, Agrim Khanna, Anmol Jagetia, Devansh Sharma, Sanchit Alekh, and Vaibhav Choudhary. Combining keystroke dynamics and face recognition for user verification. *2015 IEEE 18th International Conference on Computational Science and Engineering*, pages 294–299, 2015.

[83] Chetana Hegde, S Manu, and P Deepa Shenoy. Secure authentication using image processing and visual cryptography for banking applications. *16th International Conference on Advanced Computing and Communications*, pages 65–72, 2008.

[84] Biju P.Simon and C.Eswaran. An ecg classifier designed using modified decision based neural networks. *Computers and Biomedical Research*, 30(4):257–272, 1997.

[85] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide. Ecg analysis: A new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*, 50(3):808–812, 2001.

[86] Chetana Hegde, Rahul H Prabhu, Sagar Davasam, and Lalit M Patnaik. Heartbeat biometrics for human authentication. *Signal Image and Video Processing*, 5(4):485–493, 2011.

[87] Meenakshi Nawal and G.N Purohit. Ecg based human authentication: A review. *International Journal of Emerging Engineering Research and Technology*, 2(3):178–185, 2014.

[88] Mohammad A. Haque, Kamal Nasrollahi, and Thomas B. Moeslund. Real-time acquisition of high quality face sequences from an active pan-tilt-zoom camera. *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 443–448, 2013.

[89] Gabor T. Herman. *Basic Concepts of Reconstruction Algorithms, in Fundamentals of Computerised Tomography*. Springer, 2009.

[90] YoavFreund and Robert ESchapire. A decision-theoretic generalization of online learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[91] Jason Saragih. Face tracker. https://github.com/kylemcdonald/FaceTracker. [accessed April 20, 2014].

[92] Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[93] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Mi ller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Technical Report, University of Massachusetts, Amherst*, pages 7–49, 2007.

[94] Klaus Peter Wilhelm, Hongbo Zhai, and Howard I. Maibach. *Dermatotoxicology, Eighth edition*. CRC Press, 2013.

[95] Zhi-Qiang Wen and Zi-Xing Cai. Mean shift algorithm and its application in tracking of objects. *2006 International Conference on Machine Learning and Cybernetics*, pages 4024–4028, 2006.

[96] PierreComon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[97] H. F. Trotter. An elementary proof of the central limit theorem. *Archiv der Mathematik*, 10(1):226–234, 1959.

[98] Kavan Mannapperuma, Benjamin D Holton, Peter J Lesniewski, and John C Thomas. Performance limits of ica-based heart rate identification techniques in imaging photoplethysmography. *Physiological Measurement*, 36(1):67–83, 2014.

[99] Eirini Christinaki, Giorgos Giannakakis, Franco Chiarugi, Matthew Pediaditis, Galateia Iatraki, Dimitris Manousos, Kostas Marias, and Manolis Tsiknakis. Comparison of blind source separation algorithms for optical heart rate monitoring. *4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, 36(1):339–342, 2015.

[100] Anton M Unakafov. Pulse rate estimation using imaging photoplethysmography: generic framework and comparison of methods on a publicly available dataset. *Biomedical Physics Engineering Express*, 4(4), 2018.

[101] Jean-Francois Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[102] Klaus Nordhausen, Jean-Francois Cardoso, Jari Miettinen, Hannu Oja, Esa Ollila, and Sara Taskinen. Package 'JADE' . https://cran.r-project.org/web/packages/JADE/JADE.pdf. [accessed August 30, 2017].

[103] David S. Bolme, J. Ross Beveridge, Bruce A. Draper, and Yui Man Lui. Visual object tracking using adaptive correlation filters. *IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition (CVPR), pages 2544–2550, 2010.

[104] Richard O. Duda, Richard O. Duda, and Richard O. Duda. *Pattern Classification, Second Edition*. Wiley-Interscience, 2000.

[105] S Agatonovic-Kustrin and R Beresford. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, 22(5):717–727, 2000.

[106] MathWorks. Neural Network Toolbox. https://uk.mathworks.com/products/neural-network.html. [accessed March 30, 2018].

[107] P. A. K. Jain, D. A. A. Ross, and D. K. Nandakumar. *"Introduction," in Introduction to Biometrics*. Springer, 2011.

[108] David L. Olson and Dursun Delen. *"Performance Evaluation for Predictive Modeling," in Advanced Data Mining Techniques*. Springer, 2008.

[109] Koksoon Phua, Jianfeng Chen, Tran Huy Dat, and Louis Shue. Heart sound as a biometric. *Pattern Recognition*, 41(3):906–919, 2008.