



HAL
open science

Pulse rate variability measurement with camera-based photoplethysmography

Peixi Li

► **To cite this version:**

Peixi Li. Pulse rate variability measurement with camera-based photoplethysmography. Image Processing [eess.IV]. Université Bourgogne Franche-Comté, 2021. English. NNT : 2021UBFCK041 . tel-03474187

HAL Id: tel-03474187

<https://theses.hal.science/tel-03474187>

Submitted on 10 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE DE DOCTORAT DE L'ÉTABLISSEMENT UNIVERSITÉ BOURGOGNE
FRANCHE-COMTÉ**

PRÉPARÉE À L'UNIVERSITÉ DE BOURGOGNE

École doctorale n°37
Sciences Pour l'Ingénieur et Microtechniques

Doctorat d'Instrumentation et informatique de l'image

par

LI PEIXI

Pulse Rate Variability Measurement with Camera-based Photoplethysmography

Thèse présentée et soutenue à Dijon, le 30 Mars 2021

Composition du Jury :

LAURENT HÉLÈNE	Maître de Conférences, HDR, INSA Centre Val de Loire	Rapporteuse
VASSEUR PASCAL	Professeur des Universités, Université de Picardie Jules Verne	Président
FARRUGIA NICOLAS	Maître de Conférences, IMT Atlantique Bretagne-Pays de la Loire	Examineur
YANG SONG FAN	Professeure des Universités, Université de Bourgogne	Directrice de thèse
BENEZETH YANNICK	Maître de Conférences, HDR, Université de Bourgogne	Co-directeur de thèse
LI CHAO	Associate Professor, Institute of Acoustics, Chinese Academy of Sciences	Encadrant

Titre : Pulse Rate Variability Measurement with Camera-based Photoplethysmography

Mots-clés : Analyse de contenu vidéo, Variabilité cardiaque, Photoplethysmographie (PPG)

Résumé :

La fréquence cardiaque (FC) et la variabilité de la fréquence cardiaque (VFC) mesurées par l'électrocardiogramme (ECG) peuvent refléter les activités du système nerveux autonome (SNA) et être utilisées pour la surveillance de la santé et la reconnaissance des émotions. Récemment, la photopléthysmographie sans contact (remote PPG - rPPG) a émergé en tant que technique pour mesurer ces paramètres cardiaques. Par rapport à l'ECG, cette technique est sans contact, peu coûteuse et peut être utilisée pour la surveillance à long terme. Elle présente un grand potentiel pour l'évaluation de la santé et la détection des émotions à distance. Cependant, la rPPG est une méthode basée sur la vidéo, donc la mesure n'est pas précise et la performance est fortement affectée par le bruit de l'image, le bruit du capteur, la variation de la lumière, le mouvement de la tête, etc. Cette méthode doit donc être soigneusement étudiée et améliorée. Dans ce manuscrit, nous nous sommes concentrés sur deux questions majeures pour la méthode

rPPG. Premièrement, la sélection de la région d'intérêt (ROI) est une étape critique de la technique pour obtenir des signaux fiables. Elle doit contenir autant de pixels de peau que possible et éliminer les pixels non liés à la peau. Deuxièmement, l'estimation de la variabilité cardiaque à partir du signal rPPG est particulièrement difficile à estimer car il est nécessaire de détecter les pics sur le signal rPPG temporel, qui est généralement bruité et a une résolution temporelle inférieure à celle des signaux obtenus par les équipements en contact. Nous avons apporté plusieurs contributions pour traiter ces questions. La première contribution est l'étude comparative de plusieurs méthodes de segmentation de la ROI et des méthodes de combinaison des canaux de couleur. Deuxièmement, nous proposons une nouvelle méthode de détection du ROI. Troisièmement, nous améliorons la mesure à distance de la variabilité cardiaque grâce à de nouvelles méthodes de détection des pics. Nous montrons expérimentalement que les méthodes proposées sont plus efficaces que l'état de l'art.

Title: Pulse Rate Variability Measurement with Camera-based Photoplethysmography

Keywords: Video content analysis, Photoplethysmography (PPG), cardiac variability

Abstract:

Heart Rate (HR) and Heart Rate Variability (HRV) measured by Electrocardiogram (ECG) can reflect the activities of Autonomic Nervous System (ANS) and be used for health monitoring and emotion recognition. Recently, remote photoplethysmography (rPPG) has evolved as a non-contact technique for measuring these cardiac parameters. Compared with ECG, this technique is non-contact, low-cost and possibly utilized in long-term monitoring. It has great potential in remote health assessment and emotion detection. However, the rPPG is a video-based method, thus the measurement is not precise and the performance is heavily affected by the image noise, sensor noise, light variation, head movement, etc. Therefore, this method should be carefully studied and improved. In this manuscript, we have focused on two major issues for the rPPG method. Firstly, the selection of

region of interest (ROI) is a critical step of the technique to obtain reliable pulse signals. It should contain as many skin pixels as possible and discard the non-skin pixels. Secondly, as a possible replacement of HRV in some conditions, the Pulse Rate Variability (PRV) is more complicated to measure than HR because it is then necessary to detect the peaks on the temporal rPPG signal, which is usually noisy and has a lower temporal resolution than the signals obtained by contact equipment. We have made several contributions to address these issues. Firstly, we have conducted the comparative study on several ROI segmentation and color channel combination methods. Secondly, we propose a novel method for ROI detection. Thirdly, we improve the remote measurement of PRV with novel peak detection methods. We show that the proposed methods are more effective than the state of the art.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank Ms. H el ene Laurent and Mr. Pascal Vasseur for agreeing to read the manuscript and attending the PhD defence. I really appreciate the effort and time they put into reviewing and assessing my thesis.

I am equally thankful to Mr. Nicolas Farrugia for joining the jury and giving me the chance to present and defend my thesis.

I would like to express my gratitude to my supervisor Dr. Yannick Benezeth. He has helped me with tremendous patience combined with a pleasing modesty during the PhD project.

I am also grateful to my director Professor Fan Yang. She never hesitates to offer aid to overcome the difficulties I have faced in the study and administrative works.

I would like to thank my another supervisor Dr. Chao Li for giving me advice on my research.

I have University of Burgundy and ImViA lab to thank for providing me with the platform and resources. I am really grateful to the professors and students working in the lab.

TABLE OF CONTENTS

I	Context	1
1	Introduction	3
1.1	The measurement of cardiac parameters	4
1.1.1	Electrocardiography	4
1.1.2	Photoplethysmography	6
1.1.3	Remote photoplethysmography	7
1.2	Heart Rate Variability and Pulse Rate Variability	10
1.3	Objectives	12
1.4	Organization	13
2	State of the art	15
2.1	Remote photoplethysmography framework	16
2.1.1	Overview of rPPG technology pipeline	16
2.1.2	Region of interest selection	17
2.1.3	Color channel signal analysis	20
2.1.4	Physiological parameter estimation	24
2.2	RPPG with novel devices	25
2.3	Pulse rate variability measurement	27
2.3.1	PRV measurement and applications	28
2.3.2	Peak detection of Blood Volume Pulse signal	29
2.4	Conclusion	31
II	Contribution	33
3	Comparative study of existing rPPG methods	35
3.1	Description of the algorithms implemented in the comparative study	36
3.1.1	ROI detection methods	36
3.1.2	RPPG signal selection and combination methods	37

3.2	Experiments	39
3.2.1	Dataset	39
3.2.2	Software	41
3.2.3	Evaluation metrics	42
3.2.4	Results and discussion	44
3.3	Conclusion	47
4	Model-based Region of Interest segmentation for rPPG	49
4.1	Introduction	50
4.2	Model-based ROI segmentation	51
4.3	Experiments and results	53
4.4	Conclusion	56
5	Remote PRV measurement	59
5.1	Introduction	60
5.2	Slope Sum Function	60
5.2.1	Algorithm	60
5.2.2	Experiments	61
5.2.3	Results	62
5.3	Adaptive Slope Sum Function	67
5.3.1	Algorithm	67
5.3.2	Experiments and results	68
5.4	Adaptive Two-window Peak Detection	71
5.4.1	Algorithm	71
5.4.2	Experiments and results	74
5.5	Conclusion	77
III	Conclusion	81
6	Conclusion and future work	83
	List of acronyms	105

I

CONTEXT

INTRODUCTION

Studies that investigate and improve the cardiac physiological parameters measurement have drawn huge amount of attention of the biomedical researchers as these parameters are widely used in biomedical and psychological applications such as assessing health conditions and detecting human emotion. Usually, these parameters are measured by expensive and complex medical equipment. While the conventional measurement is precise and stable for critical medical analysis, the cost and complexity potentially reduce the availability and the wider adoption. Recently, computer vision researchers join this field because the physiological parameters are possibly extracted by image and signal processing algorithms, and this leads to the potential intelligent solutions for health and emotion monitoring. Unlike the conventional physiological parameter measurement, the computer vision methods are usually implemented in non-contact scenarios with cameras. This improves the convenience and reduces the cost, however, it is usually challenging to get precise results since the measurements are influenced by the noise of the images and the sensors. Therefore, computer vision researchers have to put much effort to improve the methods and test them in the framework of physiological measurement and related applications.

The autonomic nervous system (**ANS**) is the portion of the nervous system that controls the body's visceral functions, including action of the heart, movement of the gastrointestinal tract and secretion by different glands, among many other vital activities. The ANS consists of two parts, the sympathetic and parasympathetic nervous system (Fig. 1.1). The sympathetic nervous system activates the fight or flight response during a threat or perceived danger, and the parasympathetic nervous system controls the state of calm. Therefore, the mental and emotional states directly affect the ANS [Levenson, 2014]. Cardiac parameters such as Heart Rate (**HR**) and Heart Rate Variability have been widely studied as a reflection of many physiological factors modulating the normal rhythm of the heart as they provide a powerful means of observing the interplay between the sympathetic and parasympathetic nervous systems. The HR is a nonstationary signal. This signal and its variation contain indicators of current disease, or warnings about impending cardiac diseases. Besides the ANS, these parameters have been proven to be related to blood pressure, myocardial infarction, cardiac arrhythmia, diabetes, renal failure, and so forth. The diagnostic indicators represented by HR and Heart Rate Variability may be present at all times or may occur at random—during certain intervals of the day. It is time consuming to study the data collected over several hours. Recently, it has been shown that the Pulse Rate Variability measured by low-cost photoplethysmography can replace Heart Rate Variability

lity in some conditions [Gil et al., 2010]. Hence, HR and Pulse Rate Variability analysis with computer technologies has become a popular noninvasive tool for applications such as medical diagnostics and emotion state detection. These cardiac parameters will be discussed in details in section 1.2.

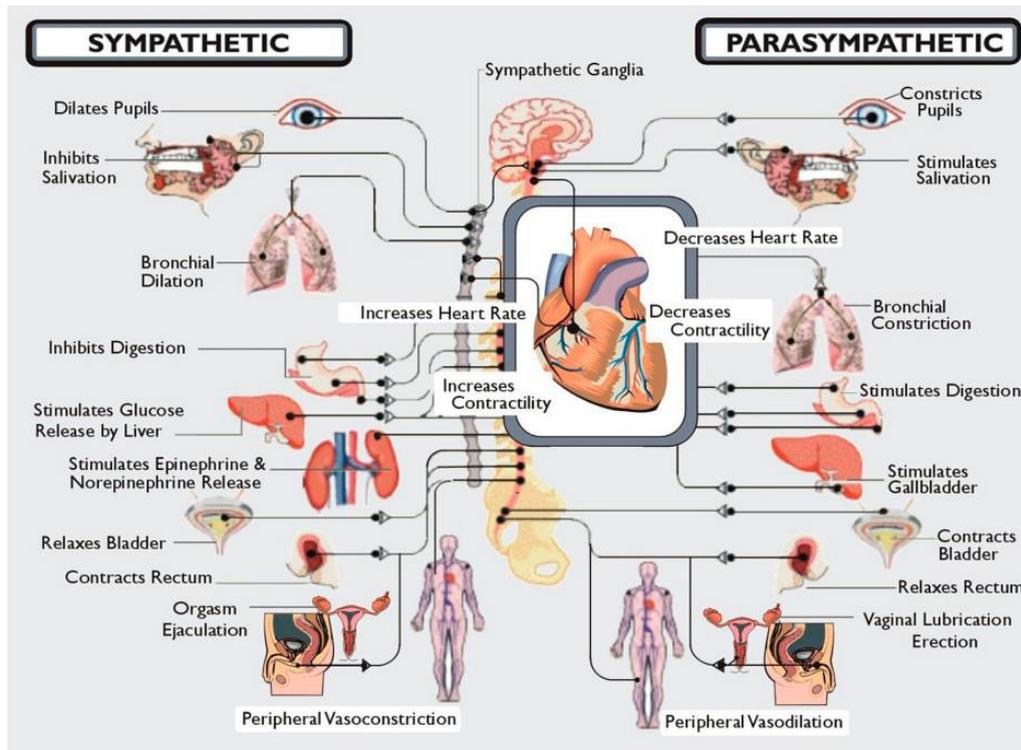


FIGURE 1.1 – The two branches of autonomic nervous system (source : <http://www.dysautonomiainternational.org/page.php?ID=122>).

The aim of the PhD project is to precisely measure the Pulse Rate Variability with remote photoplethysmography technology. In this chapter, we introduce the scientific background of cardiac parameters measurement which is discussed in three parts, namely the electrocardiogram, photoplethysmography and remote photoplethysmography. Then we describe the background of the key physiological parameters, namely heart rate variability and pulse rate variability. The objectives of the project and the organization of the thesis are introduced in this chapter as well.

1.1/ THE MEASUREMENT OF CARDIAC PARAMETERS

1.1.1/ ELECTROCARDIOGRAPHY

Electrocardiography (ECG) is a graph of voltage of the electrical activity of the heart versus time measured by electrodes attached on human skin. The heartbeat is driven by the complex interaction of myocardial cells that give the heart its ability to pump blood. The basic principle of ECG [Yanowitz, 2012] is based on the small electrical change on the

skin surface caused by the depolarization of the myocardial cells during each heartbeat. This small change can be captured by the electrocardiogram recording equipment and then magnified. When the myocardial cells are in a resting state, there is a potential difference which is formed by the difference of the concentration of positive and negative ions on both sides of the myocardial cell membrane. Therefore, depolarization is the process in which the potential difference of the myocardial cells rapidly decreases to zero and causes the myocardial cells to shrink. In a cardiac cycle of a healthy heart, the depolarization waves generated by the sinoatrial node cells propagate in the heart in an orderly manner, firstly to the entire atrium, and then to the ventricles through the "internal conduction pathway". If two electrodes are placed on any two sides of the heart, then the voltage changes between the two electrodes can be recorded during this process. The electrocardiogram can reflect the rhythm of the entire heart beat, as well as the weak part of the heart muscle. Fig. 1.2 shows one pulse of the ECG.

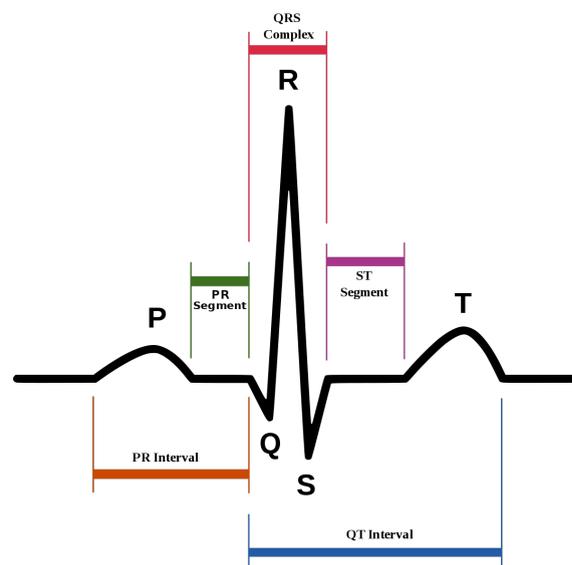


FIGURE 1.2 – The QRS complex (image taken from [Isin et al., 2017]).

There are many different ECG equipment in medical applications. In a typical twelve-lead ECG system (Fig. 1.3), ten electrodes are placed on the patient's chest, arms or legs [Yanowitz, 2012]. The magnitude of the heart's electrical potential is thus measured from twelve different leads and the signals are recorded in a certain period of time. Then the magnitude and direction of this electrical depolarization is detected at each moment in the cardiac cycle. There are three main components in an ECG (Fig. 1.2): the QRS complex, which represents the depolarization of the ventricles ; the P wave, which represents the depolarization of the atria; and the T wave, which represents the repolarization of the ventricles. To the trained clinicians, an ECG represents a large amount of information about the structure of the heart and the function of its electrical conduction system. In conventional medical applications, the ECG can be used to measure the presence of damage to the heart's muscle cells, the function of implanted pacemakers, the effects of heart drugs, the rate and rhythm of heartbeats, the size and position of the heart chambers, etc.

In the heart related studies, the ECG could be the precise reference, but it is not convenient

because of the leads and contact.



FIGURE 1.3 – The 12-lead ECG (source: <https://www.vyaire.com/products/vyntus-ecg-12-lead-pc-ecg>).

1.1.2/ PHOTOPLETHYSMOGRAPHY

Photoplethysmography (PPG) is an optical measurement technique for estimating changes in blood volumes in tissues. This technology is now well known and very often used for cardiac parameter measurements.

PPG was first introduced in 1937 by Alrick B. Hertzman. As the Greek etymology of the word photoplethysmography suggests, this technology measures the volumetric variations of the cutaneous vessels optically. In the first prototype of a photoelectric PPG [Hertzman, 1937], variations in light absorption through the skin were measured by a photoelectric cell placed over a region of the skin (of the hand) illuminated by a light source located above it (Fig. 1.4). Following these pioneering works, much research has been carried out to study this pulse waveform whose modulation is not only based on the pulsations of the blood arteries as originally assumed. This hypothesis with which the pulse oximeter was developed in the 1970s has subsequently proved to be misleading; the shape of the PPG wave is in fact more complex and would correspond to the sum of the interactions between arterial blood and venous with the cardiac, respiratory and autonomic systems [Alian et al., 2014]. These secondary sources of modulation make the PPG signal complex. Today, this signal measured by pulse oximeters are mainly used to estimate heart rate as well as pulsed oxygen saturation derived from measurements made with different wavelengths.

In the original conception of Hertzman [Hertzman, 1937] [Hertzman, 1938], the photo-receptor measures the small variations in light intensity reflected by the skin associated with arterial pulsation. This principle can be used with reflection or transmission where

the light source and the detector are separated by a tissue, typically a finger, toe or earlobe. The light that passes through the tissues is modulated by the blood pulse and then collected by the sensor. The residual light reaching the detector has two components. First of all a stationary component, called a DC component, given by the skin, non-arterial blood and other tissues that are between the source and the detector. It contains information on breathing, venous flow, sympathetic nervous system and thermoregulation [Allen, 2007]. Second, a pulsatile component, also known as the AC component, which is mainly created by variations in arterial blood volume. In addition to heart rate, pulse oximeters are conventionally used to measure pulsed oxygen saturation (SpO_2). This measurement is based on the differences in absorption of light by oxygenated blood and deoxygenated blood. Oxygenated blood absorbs little red light but much more near-infrared light, while deoxygenated blood absorbs little near-infrared light but more red light. Therefore most pulse oximeters use two light sources: a red source at 650 nm and an infrared source at 950 nm.

The PPG signals measured at these 2 wavelengths are then used in the famous formula of the ratio-of-ratios, i.e. $R = (AC_R / DC_R) / (AC_{IR} / DC_{IR})$, to obtain a quantity proportional to the oxygen saturation. AC_R and DC_R represent the AC and DC components of PPG signal measured at the red wavelength. Similarly, AC_{IR} and DC_{IR} represent AC and DC components measured at the near infrared wavelength. Then, the conversion of R to oxygen saturation was initially based on the model of Beer-Lambert. Unfortunately, this model cannot be used in practice mainly because it does not consider the processes of diffusion in the skin. The oxygen saturation is therefore in practice calculated using an empirical equation (for example $SpO_2 = K1 + K2.R$ where $K1$ and $K2$ are constants estimated by calibration from measured data [Tremper et al., 1989] [De Kock et al., 1993] [Sinex, 1999]). PPG technology is therefore commonly used for heart rate monitoring, the oxygen saturation and also to a certain extent of the pressure arterial, respiration, vascular evaluation (e.g. to aid in the diagnosis of arterial diseases or vascular stiffness) and autonomic function (with the Pulse Rate Variability which will be discussed in details in section 1.2). A detailed presentation of the clinical applications of PPG was first proposed by J. Allen in 2007 [Allen, 2007].

Fig. 1.5 shows the basic principle of contact PPG sensor. It can be seen that the amount of the light detected by the sensor can reflect the different concentration of the hemoglobin. Fig. 1.6 shows an actual pulse oximeter working as a PPG sensor.

Although PPG measurements are non-invasive and relatively easy to perform compared to ECG, they still have some drawbacks. For example, the sensitivity for movement, especially when the PPG signals are measured at the fingers or toes. In addition, as with ECG technology, the contact measure can sometimes be problematic in scenarios involving infants or patients suffering from physical trauma such as burns. Recently, the technique of non-contact PPG helped overcome some of the problems of contact PPG.

1.1.3/ REMOTE PHOTOPLETHYSMOGRAPHY

Imaging techniques play a major role in some medical applications because medical imagery has never ceased to evolve and improve. Unfortunately, these techniques

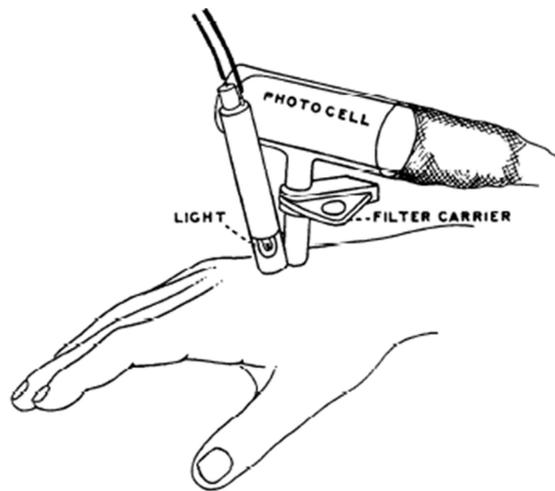


FIGURE 1.4 – The original PPG presented by Hertzman [Hertzman, 1937].

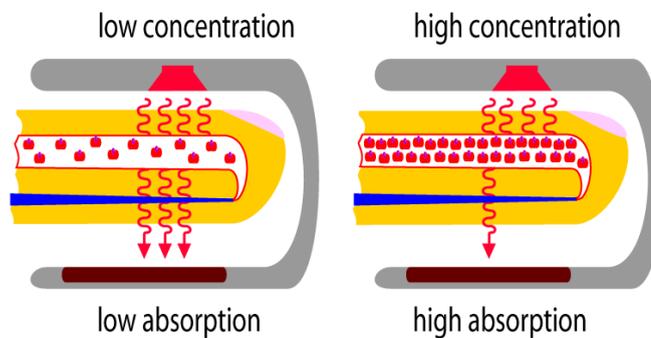


FIGURE 1.5 – Principle of contact PPG (image taken from [Chaithanya et al., 2014]).



FIGURE 1.6 – Pulse oximeter (source : BOXYM oFit2 FingerClamp Pulse Oximeter).

are usually very complex and expensive, for instance, the price of a MRI machine is usually higher than 100,000 € and the cost of a single MRI diagnostic procedure costs

more than 1000 €. The optical equipment which simply consists of device that illuminates the body or tissues with a specific light source or ambient light and to receive the photons which are emitted by the first layers of cells is completely harmless and often much easier to set up. Even in the simplest configuration, optical imaging based on the use of a color camera with ambient light as a light source provides very useful medical information. By way of illustration, we can list all the diagnostic aid systems in dermatology based on analysis of RGB or multispectral images. It is possible to find many applications optical imaging in dermatology because it is relatively easy to obtain images of the skin with a camera. These techniques have been used extensively for the detection of melanoma [Capdehourat et al., 2011] and also for the detection and monitoring of other skin lesions such as cutaneous carcinoma [Esteva et al., 2017] and vitiligo [Alghamdi et al., 2012]. These techniques have also been used for the analysis of other human tissues, for example in digestive endoscopy (the stomach [Martinez-Herrera et al., 2016] and the colon [Pigò et al., 2013] [Galeano et al., 2012] and cystoscopy (the bladder [Cauberg et al., 2010]).

Usually the diagnosis is confirmed by the analysis of a sample of tissue (i.e. a biopsy). The use of optical imaging allows the analysis of the images to help the diagnosis in a non-invasive way. Another example where the conventional cameras are used to obtain information for basic medical treatment is the monitoring of physiological parameters in hospital environment, especially in intensive care units. Monitoring of vital signs includes general temperature, cardiac and respiratory rhythms, blood oxygenation rate and blood pressure. This monitoring often involves several sensors attached to the body of the patient and this limits her/his freedom of movement and the quality of his sleep. In some cases, the techniques do not allow these measurements to be carried out, for example on patients with extensive skin lesions. *Video health monitoring* has recently emerged as a promising alternative for some health indicators among those mentioned above. Due to its contactless nature, this technology will significantly improve patient comfort.

The use of cameras as a sensor in health monitoring will make it possible to provide simple, quick and objective information for clinical examination and allow a large number of users to access their health indicators in a very easy way. Our research is placed in this context. It aims to develop new techniques for analyzing images and videos to estimate physiological parameters. The topic concerns contactless photoplethysmography and its applications. Among all the health indicators, cardiac parameters are undoubtedly key factors. As explained in the previous subsection, PPG is a technique that allows to monitor the proper function of the cardiovascular system by measuring changes in blood volumes in the tissues. This process is now well known and very often adopted with measurements on the surface of the skin. It has been shown recently that it is possible to measure the PPG signal by simply using ambient light as the light source and a camera as photoreceptor [Poh et al., 2010]. This technique is called remote photoplethysmography (rPPG).

The basic principle of rPPG follows directly from the principle of contact PPG in reflectance mode. The light reaching the camera is modulated by the blood pulsations of the skin. Thanks to its non-contact nature, the rPPG technology has many advantages. First of all, it limits the physical restrictions and wiring associated with patient monitoring. In addition, it avoids the deformation caused by the compression of the spring clips of the PPG sensors and therefore provides, in theory, more reliable signals for clinical applica-

tions. Then the use of a sensor imaging offers additional functionality over a point sensor. This technique can analyze a large tissue surface and the measurement of motion artefacts using various motion compensation techniques. Fig. 1.7 shows the rPPG experimental model with a the camera of a laptop.

The rPPG is a powerful tool which paves the way to many new applications, such as the remote health monitoring and emotion detection, however, there are some limitations, for instance, the subject movement, the light variation, dark skins and low frame rate can significantly increase the noise of the method and reduces the performance. We will give many details about rPPG in section 2.1. In the PhD project we have tried to alleviate some of these limitations.



FIGURE 1.7 – The rPPG measurement.

1.2/ HEART RATE VARIABILITY AND PULSE RATE VARIABILITY

Many existing research studies have examined the influence of emotions on the ANS utilizing the analysis of the heart rate variability which serves as a dynamic parameter of the autonomic function and balance [Sztajzel et al., 2004]. While the rhythmic beating of the heart at rest was believed to be monotonously regular, we now know that the rhythm of a healthy heart under resting conditions is actually irregular. Heart Rate Variability (HRV), derived from the ECG, is a measurement of these naturally occurring, beat-to-beat changes in heart rate (Fig. 1.8).

Biomedical studies (e.g. [Sztajzel et al., 2004]) show that HRV is an important indicator of both physiological resiliency and behavioral flexibility, reflecting the individual's capacity to adapt effectively to stress and environmental demands. It has become apparent that while a large degree of instability is detrimental to efficient physiological functioning, too little variation in heart rhythms can also be pathological. An optimal level of variability within an organism's key regulatory systems is critical to the inherent flexibility and adaptability that epitomize healthy function. In healthy individuals, the heart remains similarly

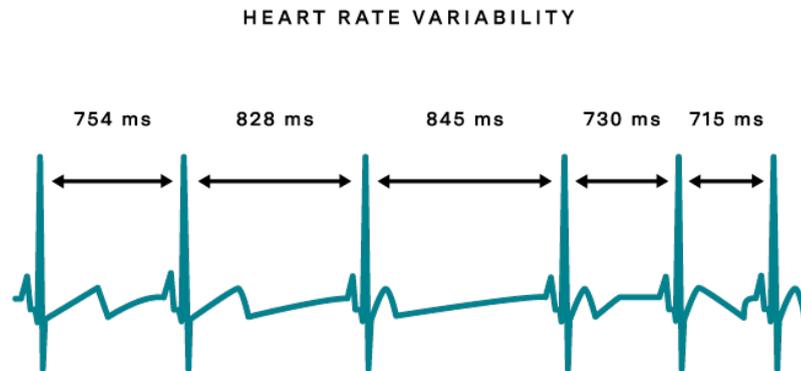


FIGURE 1.8 – An example of the HRV measurement. The HRV data is calculated as the difference of time of the signal peaks in this case (source :<https://blog.ouraring.com/what-is-heart-rate-variability/>).

responsive and resilient, primed and ready to react when needed.

The normal variability in heart rate is due to the synergistic action of the two branches of the ANS, which act in balance through neural, mechanical, humoral and other physiological mechanisms to maintain cardiovascular parameters in their optimal ranges and to permit appropriate reactions to changing external or internal conditions. In a healthy individual, thus, the heart rate estimated at any given time represents the net effect of the parasympathetic (vagus) nerves, which slow heart rate and the sympathetic nerves, which accelerate it. These changes are influenced by emotions, thoughts and physical exercise. The changing heart rhythms affect not only the heart but also the brain's ability to process information, including decision-making, problem-solving and creativity. They also directly affect how we feel. Thus, the study of heart rate variability is a powerful, objective and noninvasive tool to explore the dynamic interactions between physiological, mental, emotional and behavioral processes.

It has been shown in a number of studies (e.g. [Salahuddin et al., 2006] [Taelman et al., 2009] [Castaldo et al., 2016]) that during mental or emotional stress, there is an increase in sympathetic activity and a decrease in parasympathetic activity. This results in increased strain on the heart as well as on the immune and hormonal systems. Increased sympathetic activity is associated with a lower ventricular fibrillation threshold and an increased risk of fibrillation, in contrast to increased parasympathetic activity, which protects the heart.

The HRV data can be transformed into Power Spectral Density (PSD) by Fast Fourier Transform (FFT). The PSD is used to discriminate and quantify sympathetic and parasympathetic activity and total autonomic nervous system activity. Power spectral analysis reduces the HRV signal into its constituent frequency components and quantifies the relative power of these components. The power spectrum is divided into three main frequency ranges. The very low frequency range (VLF) (0.0033 to 0.04 Hz), representing slower changes in heart rate, is an index of sympathetic activity, while power in low frequency

range (**LF**) (0.04 to 0.15 Hz), representing slower changes in heart rate, is also an index of sympathetic activity, while power in high frequency range (**HF**) (0.15 to 0.4 Hz), representing quicker changes in heart rate, is primarily due to parasympathetic activity. The frequency range around the 0.1 Hz region reflects the blood pressure feedback signals sent from the heart back to the brain, which also affect the HRV waveform.

Usually, this physiological parameter is measured by ECG. According to Malik et al., HRV is defined as the variation of the inter-beat intervals measured by the distance of the R-peaks in the ECG signal [Malik, 1996]. Pulse Rate Variability (**PRV**) is defined as the variation of the peak intervals of the PPG signal. Some previous research shows that the PRV can be a surrogate measurement of HRV in some conditions [Gil et al., 2010]. For example, Fig. 1.9 shows the pulse transit time measured by ECG signal and PPG signal. It can be seen that there is a constant shift between the two signals possibly caused by the locations of measurement. Poh et al. proved that high degrees of agreement could be achieved between the PRV measured by a contact PPG sensor and a webcam [Poh et al., 2010]. This means that PRV measured by rPPG has great potential for multimedia applications such as remote assessment of pain, stress and emotion, although it cannot replace HRV in critical medical analysis due to the low frame rate, the noise and errors of rPPG method and completely different experimental conditions. For instance, a long-term security system for monitoring drivers' heart conditions and emotion can only be implemented with PRV and rPPG.

It is worth noting that it is difficult to precisely measure the PRV remotely because the rPPG signal is usually noisy due to light variation, face movement, sensor noise, etc. Therefore the peak detection on the rPPG signal is a challenging task. We will describe more details about PRV in section 2.3. In this thesis, improving the performance of PRV measurement is one of the key contributions.

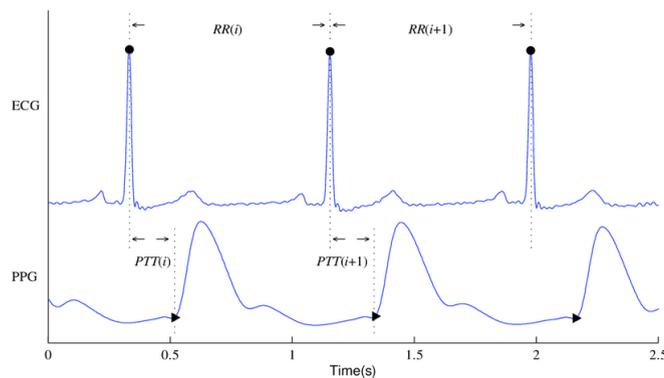


FIGURE 1.9 – ECG signal and PPG signal (image taken from [Zhang et al., 2016]).

1.3/ OBJECTIVES

As explained in the previous sections, the rPPG has great advantage over the contact PPG and ECG measurement as it is low-cost, non-contact, and much more convenient

in long-term monitoring. Based on the rPPG framework, the remotely measured PRV is potentially an effective physiological sign for various medical and psychological applications, as the PRV features can reflect the ANS activities in some conditions. Therefore, we study the rPPG framework and aim to improve the remote PRV measurement in the PhD project.

The remote PRV measurement is a camera-based technique in practice and the rPPG signals are extracted from the facial skin pixels, therefore the first critical task is to get an effective and precise region of interest from the camera videos. This step is challenging because the human face images have noise caused by beard, eyebrow, hair, glasses, light illumination variation, background noise and so forth, which would obviously affect the rPPG signals extraction. Novel algorithms have to be proposed and implemented to discard the non-skin pixels and get useful pixels as many as possible and reduce the noise of these pixels. Therefore, the first objective is to improve the region of interest segmentation.

Unlike the contact method, the rPPG signal extracted by the remote method is much more noisy than the ECG signal, due to the low frame rate, sensor noise, head movement, environmental influence such as light reflection and so forth, it is difficult to precisely detect the peaks on the rPPG signal and measure the PRV. In practical applications, the peaks must be automatically detected with high precision. So the second task is to propose robust methods to detect the peaks and measure the PRV on rPPG signal.

1.4/ ORGANIZATION

This thesis consists of three main parts, namely the context, the contribution and the conclusion. The first part has two chapters, namely the introduction and the state of the art. The introduction chapter describes the scientific background and the objectives of the project. In the state of the art chapter, we describe the existing related research works for the rPPG framework, such as the region of interest segmentation, RGB color channel selection, physiological parameters extraction, the experimental devices, and the PRV measurement. We explain why the state of the art methods are not sufficient and should be improved.

Then, the second part consists of three chapters which describe the three main contributions in the PhD project. These three contributions are: 1) comparative study of existing rPPG methods, which aims to find out the best combination of methods for region of interest segmentation and rPPG signal extraction, 2) model-based region of interest segmentation for rPPG, which is a novel method to address the issue of precise region of interest detection, 3) remote PRV measurement, in which some novel algorithms are proposed to improve the peak detection on rPPG signal and PRV measurement.

The third part is the conclusion. In this chapter, we summarize the main points of the PhD project and discuss the potential future work.

STATE OF THE ART

This chapter describes the state of the art in remote measurement methods of cardiac variability features. First, we will describe the camera-based measurement of rPPG signals by presenting the algorithms and the associated devices. More precisely, we will first describe the seminal works that validated the possibility of measuring remotely the photoplethysmogram signal. Then, we will present the key steps of the regular algorithmic pipeline for video-based rPPG methods: the region of interest segmentation, channel signal combination, and the physiological parameter estimation. The different materials and devices that can be used to make this measurement will also be presented.

Next, we will describe how the cardiac variability signal is estimated from the PPG or rPPG signals and what are the main features of this signal. We will present the guidelines for HRV/PRV measurement, compare the HRV, PRV and remotely measured PRV signals and then describe in details the peak detection and filtering problem that is critical for PRV measurement.

Contents

2.1 Remote photoplethysmography framework	16
2.1.1 Overview of rPPG technology pipeline	16
2.1.2 Region of interest selection	17
2.1.3 Color channel signal analysis	20
2.1.4 Physiological parameter estimation	24
2.2 RPPG with novel devices	25
2.3 Pulse rate variability measurement	27
2.3.1 PRV measurement and applications	28
2.3.2 Peak detection of Blood Volume Pulse signal	29
2.4 Conclusion	31

2.1/ REMOTE PHOTOPLETHYSMOGRAPHY FRAMEWORK

In this section, the research works which focus on rPPG technology for physiological parameter measurement is described and discussed. The related articles are introduced in four parts, the overview of the rPPG technology pipeline, the region of interest segmentation, the color channel signal combination and physiological parameters measurement.

2.1.1/ OVERVIEW OF rPPG TECHNOLOGY PIPELINE

Figure 2.1 presents the general principle of measuring physiological parameters by non-contact PPG. Most methods are based on the same main steps where the region of interest (usually the face) is first detected and tracked over time, a spatial average is then performed for each image and the RGB color channels are then combined to obtain the rPPG signal. This signal is filtered and then used to estimate various physiological parameters such as heart and respiratory rates or even pulse rate variability. Several works have been published on the subject [Kublanov et al., 2015] [Wiede et al., 2016] [van Gastel et al., 2016] [Fouad et al., 2019] [Artemyev et al., 2020]. In the rest of this section, we briefly present some seminal works that introduced this pipeline and in the next sections we will detail the different methods for each of these steps.

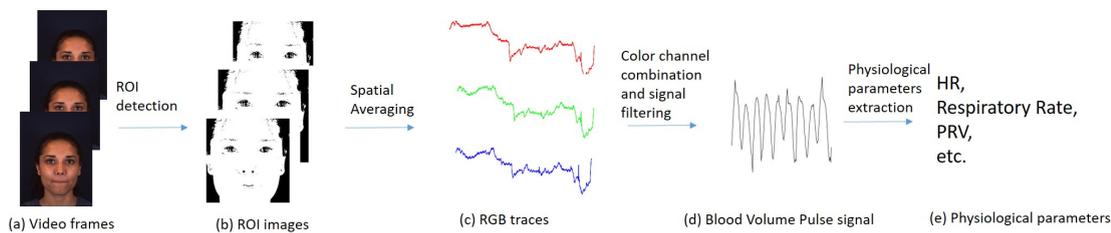


FIGURE 2.1 – Classical rPPG framework: (a) the original video frames, (b) the detected skins (the white part is the detected skin pixels and the black part is the non-skin pixels), (c) spatially averaged RGB signals, (d) Blood Volume Pulse signal, (e) physiological parameters measurement (e.g. HR, respiratory rate, PRV, etc).

In 2009, Zheng et al. [Zheng et al., 2009] introduced the basic principles of camera-based rPPG and compared it with the contact point measurement techniques such as pulse oximetry probes. In the experiments, they used a CCD camera with a frame rate of 30 fps, and the distance of the human face and the camera is about 13 cm. They have shown that the PPG signal can be estimated from the green channel of the camera. Moreover, the time-frequency spectrum of these pulsatile components derived from the rPPG signals could be identified, indicating that both the HR and the shape of the plethysmogram were determined. Although the errors did not reach an ideal level, and the experimental set up was not convenient enough, this research work implied that camera based rPPG has great potential for future applications.

In 2010, Poh et al. [Poh et al., 2010] presented the precise framework for camera-based rPPG for the first time such as in Fig. 2.1. In their work, they adopted several effective signal and image processing methods to reduce the noise. For instance, they proposed

to automatically identify the coordinates of the face location in the video recording so that they were able to track the region of interest (ROI) accurately. Then they used Independent Component Analysis (ICA) to process and combine the RGB channels and get the Blood Volume Pulse (BVP) waveform. The idea is based on the assumption that the cardiac signal is mixed linearly in the RGB time traces. Other sources such as noise, light and motion disturbances are statistically independent of the heart signal. It is therefore possible to separate these sources with metrics based on statistical independence. The experimental conditions were also loosened with respect to the work of Zheng et al. [Zheng et al., 2009], for example, the distance between the camera and the face was about 0.5 meter, unlike Hu et al.'s experiment in which the distance is 13 cm. Then they assessed the heart rate measurement by comparing the results with contact sensor quantitatively. Several metrics such as mean error, standard deviation, root mean square error, correlation coefficient were used to evaluate the performance. And the results showed that the relative errors for metrics such as mean absolute error are usually less than 10% on their dataset.

Inspired by this work, many researchers have focused on improving the framework by using better ROI detection, channel signal processing and combination and better physiological signal extraction algorithms.

2.1.2/ REGION OF INTEREST SELECTION

As shown in Fig. 2.1, the ROI detection is the key step for rPPG signal extraction. In the work of Moreno et al. [Moreno et al., 2015], the researchers manually selected the ROI with a rectangle. Obviously, the manual approach is not convenient, therefore the automatic ROI segmentation methods have been investigated. Since the BVP signal is directly extracted from the facial skin pixels, the objective of this step is to get as many skin pixels as possible and discard as many background pixels and non-skin pixels as possible. There are basically two methods, skin and face detection and landmarks segmentation.

Skin and face detection. Since most of the useful pixels of rPPG are facial skin pixels, therefore the most straightforward methods for ROI detection are face and skin detection methods. Several research works have focused on improvement of facial region detection and skin/non-skin classification since the earliest works where manual selection was used.

The classical method for face detection is the well-known Viola-Jones face detector [Viola et al., 2001] implemented by the computer vision toolbox of MATLAB and OpenCV. In order to avoid spurious movements of the detected face, the Kanade-Lucas-Tomasi tracking [Lucas et al., 1981] algorithm is also be used in many papers [Yang et al., 2016] [Hassan et al., 2017] [Po et al., 2018]. An example of detection is given in Fig. 2.2(a). As suggested by Poh et al. [Poh et al., 2010] and Liu et al. [Liu et al., 2014], it is also possible to simply crop the ROI selecting the center 60% width and height of the box as ROI. This first strategy, very simple, allows to remove many pixels from the background. An example is given in Fig. 2.2(b). However, the eye and mouth regions are often in motion in videos and are therefore likely to introduce noise into the rPPG signal. So, a technique called *Grabcut* [Rother et al., 2004] has been used to

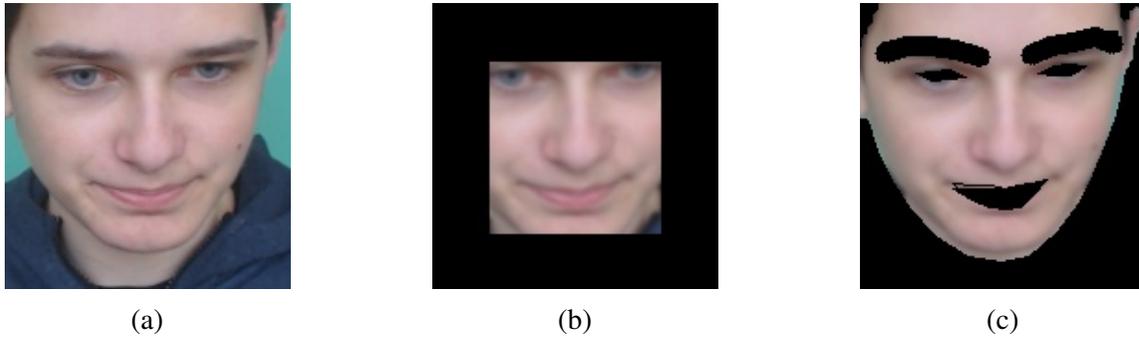


FIGURE 2.2 – ROI segmentation result examples using (from left to right) (a) the detected face, (b) a cropped version of the face and (c) removing some parts of the face using a segmentation strategy.

further improve the ROI selection by segmenting the face into different parts (eyes, mouth and eyebrows) [Stricker et al., 2014]. This *Grabcut* technique uses Gaussian Mixture Models for modelling the distribution of data and Graph cut for segmentation. An example of segmentation is given in Fig. 2.2(c).

Besides the face detection, a more precise way to get the ROI is skin/non-skin classification since in the context of rPPG, we are only interested in skin pixels. Many review papers have been published presenting various color-based skin detection rules (e.g. in [Kakumanu et al., 2007]). A common method [Kovac et al., 2003] is this:

$$\begin{aligned}
 & (R,G,B) \text{ is classified as skin if} \\
 & R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and} \\
 & \max\{R, G, B\} - \min\{R, G, B\} > 15 \text{ and} \\
 & |R - G| > 15 \text{ and } R > G \text{ and } R > B
 \end{aligned} \tag{2.1}$$

with RGB the pixel values for the red, green and blue channels of a camera. An example of skin detection is given in Fig. 2.3(a).

Instead of using fixed predefined thresholds for detecting skin pixels, it is possible to derive the range of skin pixel values using specific ROI of the face. For example, a small ROI in the center of the face detection bounding box is used to create a reference skin color specific to the detected face. Similar pixels in the detected face ROI are then segmented to create the ROI mask. The main advantage of this method [Taylor et al., 2014] is that it is person-specific and does not rely on global threshold values (Fig. 2.3(b)). A more common method for skin detector is based on a thresholding of a non-parametric histogram trained using manually annotated large dataset of images with skin and non-skin pixels [Conaire et al., 2007]. This method is eventually based on a Look-Up-Table and is thus very fast. An example of result is shown in Fig. 2.3(c).

Some papers proposed advanced method for skin detection and face alignment, for instance, Wang et al. [Wang et al., 2018] developed a system using joint face detection and alignment and Delaunay Triangulation method to minimize tracking-artifacts arising from

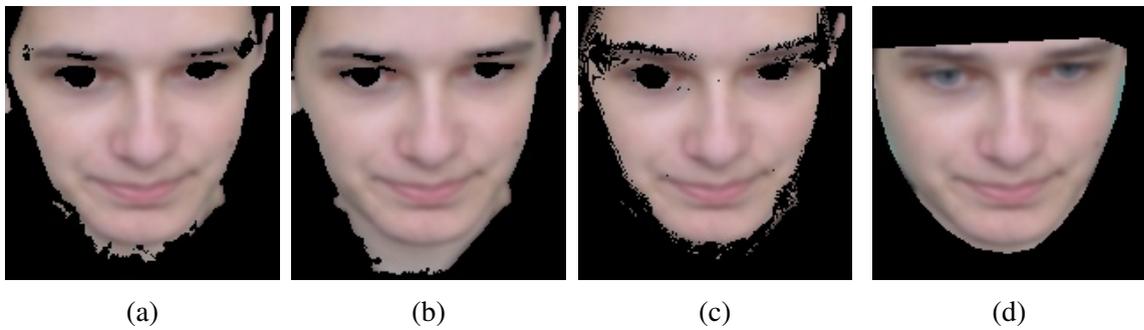


FIGURE 2.3 – ROI segmentation result examples using (from left to right) (a) rule-based skin detection, (b) adaptive skin detection, (c) histogram-based skin detection and (d) landmarks detection.

large head motions and facial expressions and produce alignment-friendly facial bounding boxes with reliable initial facial shapes, facilitating accurate and robust face alignment even in the presence of large pose variations and expressions. Werner et al. [Werner et al., 2014] proposed an approach using fusion of multiple ROI signals to deal with motion. It firstly tracked the key points on the face. And then it divided face with several regions. Instead of using one single ROI, the algorithm used the fusion of different parts of the face such as forehead and cheeks.

Landmarks detection. Another approach to get the ROI is to make use of the landmarks obtained by the facial features [Kazemi et al., 2014]. With the landmarks, the ROI can be selected as a polygon. Fig. 2.4 shows the model for the landmarks. One example of ROI detection with landmarks is given by Fig. 2.3 (d).

Based on this idea, Haque et al. [Haque et al., 2016] introduced a Face Quality Assessment (FQA) method that prunes the captured video data so that low-quality face frames cannot contribute to erroneous results. The quality was assessed by four metrics: resolution, brightness, sharpness, and out-of-plan face rotation. They then extract the parts with good quality and combine them with facial landmarks. A combination of these two methods for vibration signal generation lets them obtain stable trajectories that, in turn, allow for a better estimation of HR.

Based on a similar idea, several researchers improved the facial ROI detection. Xu. et al. [Xu et al., 2017] proposed a method based on Partial Least Squares (PLS) and Multivariate Empirical Mode Decomposition (MEMD) for facial region detection. The framework is based on the assumption that both facial ROI and background ROI have the similar illumination variations and the background ROI can be treated as the denoising reference by using PLS to extract the underlying common illumination variation sources existing in both ROIs. Then, a number of intrinsic mode functions were decomposed by applying MEMD to the illumination-variation-suppressed facial ROI and the HR is evaluated.

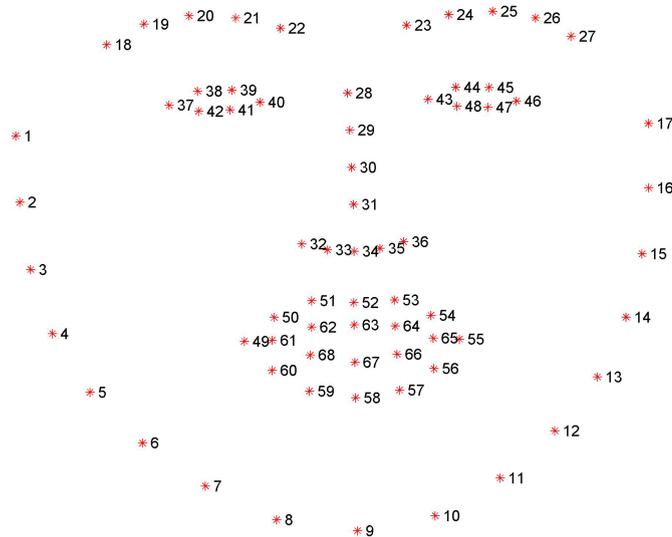


FIGURE 2.4 – 68 facial landmarks.

2.1.3/ COLOR CHANNEL SIGNAL ANALYSIS

As shown in the Fig. 2.1, after segmenting the ROI, temporal RGB signals are obtained by simple spatial averaging of the pixels. To obtain an rPPG signal, it is then necessary to combine the three RGB channels. There are four main categories of commonly used methods for this. The first is based on source separation methods, the second on color space analysis, the third is based on light-tissue interaction models and the fourth is RGB signals decomposition.

Blind source separation. In early work on rPPG measurement, the cardiac signal was extracted from RGB time traces using Blind Source Separation (BSS) and more specifically ICA.

ICA is a signal processing technique for separating a multivariate signal into its constituent components using metrics based on statistical independence. ICA can be performed using various algorithms such as mutual information maximization and nongaussianity maximization or algebraic methods exploiting the structure of the fourth-order cumulants (e.g. JADE for Joint Approximate Diagonalization of Eigenmatrices) [Liao et al., 2002]. Concerning rPPG signals, the idea is based on the assumption that the cardiac signal is linearly mixed in the RGB time traces. Other sources such as noise, light and motion disturbances are statistically independent of the cardiac signal. It is therefore possible to separate these sources with independence measurements.

In addition to ICA analysis, Principal Component Analysis (PCA) [Wedekind et al., 2015] and Non-linear Mode Decomposition (NMD) [Demirezen et al., 2018] are other source separation methods that have been used to extract the cardiac signal. These methods are decomposition techniques that separate and extract the desired signal from the input mixture based on certain properties. For example, PCA looks for directions that maximize the variance of the data, under the orthogonality constraint, which should allow the extraction

of the most important signal components. In the case of NMD, the desired signal is extracted by decomposing it into several modes of time-frequency representation, similar to Fourier decomposition.

This BSS strategy was firstly adopted by Poh et al. [Poh et al., 2010]. The approach could be applied to color video recordings of the human face and is based on automatic face tracking along with BSS of the color channels into independent components. Using Bland-Altman and correlation analysis, they compared the cardiac pulse rate extracted from videos recorded by a basic webcam to an FDA-approved finger BVP sensor and achieved high accuracy and correlation even in the presence of small movements artifacts. Furthermore, they applied this technique to perform heart rate measurements from three participants simultaneously. This was the first demonstration of a low-cost accurate video-based method for contact-free heart rate measurements that was automated, motion-tolerant and capable of performing concomitant measurements on more than one person at a time.

With the similar idea, many researchers furtherly investigated BSS algorithms. Christinaki et al. [Christinaki et al., 2014] evaluated the performances of three BSS algorithms for rPPG measurements. The selected algorithms are the JADE, the FastICA and the RobustICA. Mannapperuma et al. [Mannapperuma et al., 2014] investigated the limits of detection of the HR while reducing the video quality. They compared the performance of three ICA methods (JADE, FastICA, RADICAL) and autocorrelation with signal conditioning techniques and identify the most robust approach. Both researchers found that different BSS algorithms gave similar performance whatever the dataset is used.

Macwan et al. [Macwan et al., 2018] presented an algorithm for measuring rPPG signals using constrained ICA (cICA). They incorporated a priori information into the cICA algorithm to aid in the extraction of the most prominent rPPG signal. The priori information is implemented using two constraints: first, based on periodicity using autocorrelation, and second, a chrominance-based constraint exploiting the physical characteristics of the skin. The method improved performances over traditional BSS methods in terms of accuracy and robustness.

Interestingly, another method is to use joint blind source separation (J-BSS). Qi et al. [Qi et al., 2017] showed that it was promising to incorporate data from different facial subregions to improve remote measurement performance via J-BSS. This was the first time that J-BSS approaches, instead of prevailing BSS techniques such as ICA, was successfully applied in non-contact physiological parameter measurement. Experimental results showed that the proposed J-BSS method outperforms the ICA-based methodologies. Wei et al. [Wei et al., 2017] proposed to use a dual region of interest on facial video image which was selected to yield 6-channels RGB signals and this allows BSS algorithms to work more stable and efficient in separating multiple physiological signals. By applying second-order blind identification algorithm to those signals generated above, they obtained 6-channels outputs that contain BVP and respiratory motion artifact. For the automatic selections of the BVP among these outputs, they devised a kurtosis-based identification strategy, which guarantees the dynamic Respiratory Rate (RR) and HR monitoring available. The experimental results indicated that, the estimation by the proposed method had an impressive performance compared with the measurement of the commercial medical sensors.

Color space analysis. As shown in the previous subsection, most of the rPPG methods process the RGB color-space to obtain the blood volume pulse signal. The rPPG methods which use BSS methods presented above to estimate heart pulse rate to achieve high estimation accuracy require considerable signal-processing power and result in significant processing latency. High processing power and latency are limiting factors when real-time pulse rate estimation is required or when the sensing platform has no access to high processing power. The limitation of the rPPG methods using RGB spaces is that they can suppress at most $(n - 1)$ independent distortions by linearly combining n wavelength color channels [Wedekind et al., 2015]. Their performance are highly restricted when more than $(n - 1)$ independent distortions appear in a measurement, as typically occurs in fitness applications with vigorous body motions. Some researchers have found that simply using a different color space may improve the performance of rPPG measurement and reduced the required computing power.

Tsouri et al. [Tsouri et al., 2015] considered seven different color spaces (sRGB, HSL, HSV, HIS, XYZ, CIE XYZ, and CIE YUV) and compare their performance with state-of-the-art algorithms that use ICA. The comparison is performed over a dataset of 41 video recordings from subjects of varying skin tone and age. Results indicate that the hue channel provides better estimation accuracy using extremely low computation power and with practically no latency. Rumiński [Rumiński, 2016] proposed to use YUV color model to estimate the pulse rate. The experiments showed that using native YUV video formats and analyzing V signals recorded for the forehead ROI in the time domain, it is possible to accurately and efficiently measure the pulse rates.

Alternative color spaces can be used to enhance the BVP signal extracted from image frames. For instance, Suh et al. [Suh et al., 2017] proposed a method for extracting the pulse signals with improved quality from videos recorded with a remote RGB camera using skin color magnification which revealed the hidden physiological components remarkably in the time-series signal. The method is done with three steps, 1) firstly they converted the RGB channel to YCbCr space, 2) secondly, they enlarged the skin color variation by radially expanding n times of the distance between each pixel and the pixel cluster center in Cb and Cr plane (Fig. 2.5), 3) after the distribution of the skin pixels is widened, the resulting frame is reconverted into the RGB color space, and as a result, the pixel changes over time are enlarged and the cardiac pulse information from the image frames are enhanced. Unlike the postprocessing algorithms such as filtering techniques, the method proposed by the authors focused on magnifying cardiac components in the signal.

Light-tissue interaction model. The methods based on light-tissue interaction model focus on exploiting the physical characteristics of the skin and the interaction between light and skin. Indeed, the relative amplitude of the PPG signal (and thus rPPG) in light reflected from the skin varies as a function of wavelength. This is mainly due to the absorption spectrum of hemoglobin shown in Fig. 2.6. Therefore, there is significantly more rPPG information in the wavelengths corresponding to the color green than in the other colors. This is why some of the early techniques were limited to selecting the green channel only [Lempe et al., 2013]. Methods such as Chrom [De Haan et al., 2013], POS [Wang et al., 2016] and PBV [De Haan et al., 2014] use these properties and rely on a

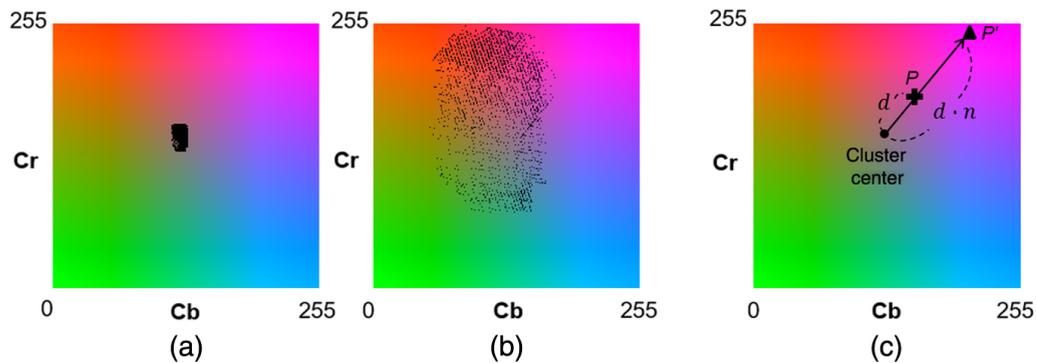


FIGURE 2.5 – Before and after applying the skin color magnification (picture taken from [Suh et al., 2017]). Image (a) and (b) show the skin pixel distributions before and after applying the skin color magnification. Image (c) shows the principle of skin color magnification in a single pixel.

detailed formulation of models based on skin properties. The main advantages of these methods lie in their simplicity of calculation, due to their analytical formulation. These methods can be seen as analogous to a change of basis on which projections of the original RGB signals intensify the rPPG signal. We will discuss the Chrom in details as the representative of light-tissue interaction model and we experimentally compared it with the BSS method in section 3.1.2.

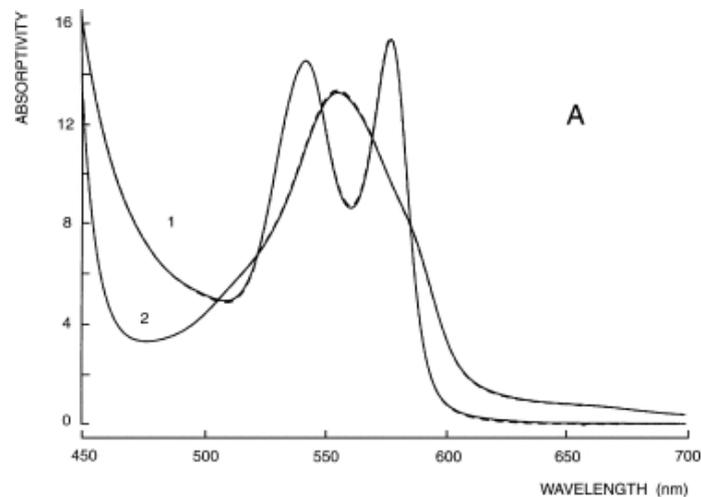


FIGURE 2.6 – Absorption spectra of hemoglobin. Oxyhemoglobin (1) and deoxyhemoglobin (2) (image taken from [Zijlstra et al., 1997]).

RGB signals decomposition and subspace analysis. Wang et al. [Wang et al., 2017] proposed an effective and simple method that algorithmically extends the number of possibly suppressed channels without using more wavelengths. The core idea is to increase the degrees-of-freedom of noise reduction by decomposing the n wavelength camera-signals into multiple orthogonal frequency bands and extracting the pulse-signal per band-basis. This processing can suppress different distortion-frequencies using independent combinations of color channels. The inspiration of the method is based on the observation that dif-

ferent motion frequencies cause apparent distortions in different color variation directions in RGB space. This precludes treating them simultaneously, but treating them independently in different frequency bands may solve this problem. To this end, they decomposed the RGB signals into multiple orthogonal frequency bands for sub-band pulse extraction. Once the different motion frequencies are separated and suppressed in different frequency bands, they synthesized a clean pulse-signal by combining the processing results from the individual sub-bands. This idea leads to a novel pulse extraction method called ‘Sub-band rPPG’.

Wang et al.’s team also proposed a conceptually novel algorithm, namely “Spatial Subspace Rotation” (2SR), that improves the robustness of rPPG [Wang et al., 2015]. Based on the assumption of 1) spatially redundant pixel-sensors of a camera, and 2) a well-defined skin mask, the core idea is to estimate a spatial subspace of skin-pixels and measure its temporal rotation for pulse extraction, which does not require skin-tone or pulse-related priors in contrast to existing algorithms. It consists of two steps: 1) in the spatial domain, a subspace of skin-pixels is constructed in RGB space; 2) in the temporal domain, the rotation angle of spatial subspaces between subsequent frames is measured for pulse extraction. The experiments demonstrated that when multiple pixel-sensors of a regular RGB camera are used for skin sensing, and a well-defined skin mask is available, 2SR outperforms the popular ICA-based method.

2.1.4/ PHYSIOLOGICAL PARAMETER ESTIMATION

As shown in the Fig. 2.1, it is often necessary to filter the resulting signal before using it to estimate physiological parameters. This can be done, for example, by removing those parts of the signal that are not within a frequency range corresponding to a normal heart rate (typically between 50 and 200 beats per minute). Several strategies have been used for this purpose, ranging from simple averaging filtering, such as [De Haan et al., 2013], to the classic Butterworth bandpass filtering, such as [Niu et al., 2017] [Lin et al., 2017], but also adaptive filtering [Huang et al., 2015], wavelet filtering [Huang et al., 2016] and Singular Spectrum Analysis (SSA) [Zhao et al., 2018]. The filtered signal is then used to estimate physiological parameters.

Heart rate is the first physiological parameter that can be estimated from the rPPG signal. To do this, it is possible to perform a temporal analysis of the signal by estimating the duration between successive peaks of the signal using different heuristics [Lewandowska et al., 2011] or to perform a frequency analysis with, for example, a Fourier transform. The dominant frequency peak corresponds to the heart rate.

The rPPG signal also contains three components that can be related to respiratory rhythm: the change in heart rate known as Respiratory Sinus Arrhythmia (RSA), the change in the baseline signal due to the change in intrathoracic pressure, and the change in pulse amplitude caused by a change in cardiac output. Therefore, a few papers have proposed using one or more of these factors to estimate breathing rate from the rPPG signal [van Gastel et al., 2016] [Alekhin et al., 2013].

Using the differences in oxyhemoglobin and deoxyhemoglobin absorption in the red and near-infrared wavelengths, it is possible in principle to estimate the pulsed oxygen satura-

tion SpO_2 . Several works have been published on the subject based on the precursory work of Wieringa et al. [Wieringa et al., 2005] with more or less complex imaging systems. Humphreys et al. [Humphreys et al., 2007] use for example a dedicated imaging system with a camera synchronized with a dual wavelength light emitting diode system in the near infrared while Bal et al. [Bal, 2015] simply use the blue and red channels of a single camera in the visible spectrum. However, much remains to be done to make this technique reliable and robust. The main problem lies in the determination of the calibration function because the environment is much less controlled with non-contact acquisitions than with contact acquisitions.

Variations in the interval between heartbeats, attributed to changes in the parasympathetic/sympathetic balance of the autonomic nervous system, are usually measured with an ECG and determine the HRV. This parameter has been used in numerous research and clinical studies [Malik, 1996]. As explained previously in section 1.2, the PRV measured from the PPG signal (and by extension rPPG) has similar characteristics to the HRV. Consequently, several studies have focused on the estimation and exploitation of the PRV measured by video analysis [Bousefsaf et al., 2013] [McDuff et al., 2014a] [Benezeth et al., 2018] [Sabour et al., 2019] [Parra et al., 2019] mainly for applications in psychometrics (e.g. recognition of emotions, stress, etc.). We will detail the importance of cardiac variability in section 2.3 and describe the methods for estimating it.

2.2/ RPPG WITH NOVEL DEVICES

Interestingly, with exactly the same basic framework, some researchers proposed some novel cameras to get better results, such as the cameras with higher frame rate, the five-band RGBCO camera, the thermal camera and the polarized camera.

Camera with higher frame rate. Compared with the ECG of which the frame rate is usually higher than 250 fps, most rPPG systems have been limited by a low sample frequency which is usually 20-50 fps. This obviously restricts their use clinically in the assessment of PRV due to the low precision. Therefore in the work proposed by Sun et al. [Sun et al., 2012], PPG signals were remotely captured via an rPPG system at a rate of 200 fps. The physiological parameters (i.e., heart and respiration rate and PRV) derived from the rPPG datasets yield statistically comparable results to those acquired using a contact PPG sensor, the gold standard. More importantly, the work presented evidence that the negative influence of initial low sample frequency could be compensated via interpolation to improve the time domain resolution.

Five-band camera. The RGB cameras are the most widely used cameras available in everyday life, however, some researchers suggested that the 5-band RGBCO cameras could work better since more information could be used in the ICA and PCA methods to combine the signals. McDuff et al. [McDuff et al., 2014b] presented a work that adopted the novel 5-band camera and found that the cyan, green, and orange (CGO) bands performed better than RGB bands in measuring the PRV in the frequency domain. The

correlation between the signals measured by the contact sensor and the camera was over 90%. Their further work [McDuff et al., 2016] utilized the CGO bands and conducted an experiment which included two randomly ordered tasks, namely the ball control task and the Berg Card Sorting Task [Berg, 1948]. The results demonstrated that the PRV features extracted by CGO bands could distinguish between the relaxed and stressed modes with an accuracy of 70%–80%. Additionally, they showed that the value of PRV features could capture the changes of stress since the two tasks in the experiments caused different stress levels for the participants and they were detected according to the value change. In realistic applications, the remote PRV measurement may suffer from missing observations caused by subject movement and subject getting occluded by an object, they addressed the issue by proposing an algorithm to fuse partial camera signals generated from an array of cameras and they improved the PRV measurement in the scenarios where significant amount of data is lost [McDuff et al., 2017].

Camera with cross-polarisation filters. Fan et al. [Fan et al., 2017] proposed a system that consists of a CCD camera and a set of optical filters. Trumpp et al. [Trumpp et al., 2017] analyzed the employment of polarization filtration to assess the gain for the signal quality with suppressed specular reflectance and draw conclusions about the rPPG signal’s origin. The benefit of polarization filtration for rPPG using wavelengths in the red, green and blue color range was studied. The results proved that for an optimal illumination, the perpendicular filter setting provides a significant benefit. Fig. 2.7 shows this novel setup.

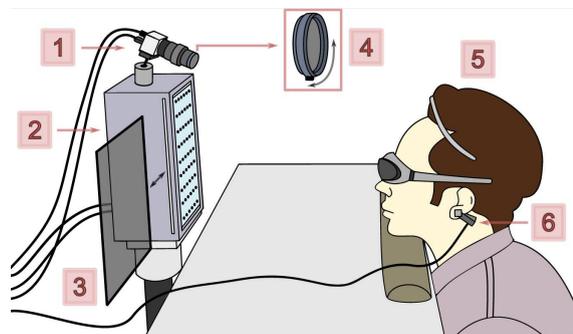


FIGURE 2.7 – The novel setup with polarization device (image taken from [Trumpp et al., 2017]). (1) Camera system, (2) Multi-wavelength illumination matrix, (3) Polarization filter (film), (4) Polarization filter (glass) with an adaptable angle, (5) Test subject with protection goggles and hairband, (6) Reference PPG sensor.

Near infrared and infrared camera. Trumpp’s team [Trumpp et al., 2018] developed a framework which allowed rPPG to be applied in the intraoperative environment using an RGB and a near-infrared (NIR) camera (Fig. 2.8). The method does not rely on the detection of anatomical features or chooses and tracks visible skin regions which are homogeneously illuminated, or solely operates on the image plane without being reliant on the presence of temporal variations related to the cardiac cycle. The performance was evaluated with respect to the quality of extracted PPG signals and correctly detected the

HR. They further evaluated their quality with regard to the extracted rPPG signals. Hu et al. [Hu et al., 2018] leveraged a far-infrared imager and a near infrared camera equipped with a lens and an near infrared lighting array to develop a dual-camera imaging system. With a dual-mode sleep video database which was constructed to evaluate the effectiveness of the proposed system and algorithms. The overall performance of the proposed technique is acceptable for RR and HR estimations during nighttime. Similarly, a method for non-contact measurement of HR using thermal imaging was proposed by Hamedani et al. [Hamedani et al., 2016]. With thermal videos recorded from subjects' faces, the accuracy was 99.42% in the best case and 92.47% in average for 22 subjects. Gupta et al. [Gupta et al., 2016] proposed a system that used a thermal camera, a monochrome camera and an RGB camera to extract the BVP signal. This novel system was proved quite effective to reduce the noise caused by motion and light illumination variation. With this system, one can monitor the HR and PRV and visualize the data in a real-time scenario.

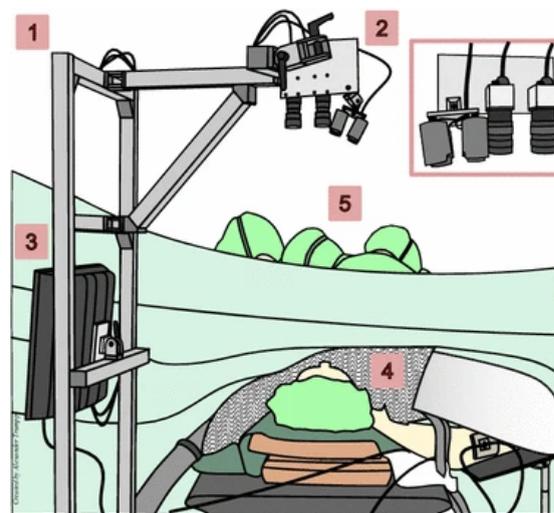


FIGURE 2.8 – The novel intraoperative setup (image taken from [Trumpp et al., 2018]). (1) Construction with adjustable arm for the sensing system. (2) Sensing system (enlarged on the right) including NIR illumination, NIR camera, and RGB camera. (3) Recording PC. (4) Patient (face directed towards the cameras). (5) Surgeons and clinical staff.

These systems are very interesting because they allow to perform monitoring in dark environments, but the possibility to make an accurate measurement with a standard camera would allow a greater use of this technology. This is why we decided to focus in this thesis on regular RGB camera.

2.3/ PULSE RATE VARIABILITY MEASUREMENT

As described in section 1.1 and 1.2, physiological parameters such as HRV and PRV are vital signs for monitoring ANS system. The time and frequency domain features of HRV can reflect the activity of different branches of the ANS. For instance, the HF part of HRV represents the Parasympathetic Nervous System (PNS) while the LF part of HRV

represents the sympathetic nervous system (SNS). With these features, the health and emotion conditions can be estimated. HRV is defined as the variation of the inter-beat intervals measured by the distance of the R-peaks in the ECG signal. PRV is defined as the variation of the peak intervals measured from the PPG signal. Although the remotely measured PRV has great potential in many applications as it can be extracted with low-cost and non-contact equipment, there are several issues here: 1) the equipment that measures the PPG signal and ECG signal are different as we have discussed in section 1.1, therefore the PPG signal may not completely replace the ECG signal and the PRV cannot completely replace HRV as a vital sign for biomedical application. 2) The rPPG signal is measured by camera, and the measurement suffers from the low frame rate of the camera, the noise of the sensors, the light variation, the movement of the subjects and so forth. This can be critical for the measurement of PRV. 3) The PRV measurement relies on precise detection of peaks on PPG signal, and this is particularly difficult for rPPG signal because it is usually noisier than contact based PPG as explained above. Therefore, in this section we discussed the research papers on these issues: firstly we introduced the works that set the guidelines for HRV and PRV measurement. Secondly we described the works that compare PRV measurement and HRV measurement. Then we introduce the possible applications of remotely measured PRV. Finally we describe the improvement of peak detection for HRV/PRV measurement in different conditions.

2.3.1/ PRV MEASUREMENT AND APPLICATIONS

One important task is to study and follow the guidelines of HRV measurement and validate that PRV can be a surrogate measure of HRV in some conditions.

HRV and PRV measurement. The European task force wrote a report [Malik, 1996] that gave the precise definition of HRV, the guidelines for HRV measurement and the relevant applications. The report proposed the standards of HRV measurement, physiological interpretation and clinical use which has been widely accepted by the researchers. Béres et al. [Béres et al., 2019] gave a comprehensive study of the adequate sampling frequency for contact PPG measurement and the detailed instructions for interpolation and sampling. We believe the PRV measurement should follow the standards as well and the rPPG signals should also be sampled and interpolated and this may reduce the errors that are caused by the low frame rate of the camera. Gil et al. [Gil et al., 2010] showed that the PRV extracted from the PPG signal can be used as an alternative measurement of the HRV signal in non-stationary conditions. In their study the classical indices of HRV analysis were compared to the indices from PRV. The conclusion is that the features in both time domain and frequency domain are highly correlated. Futhurly, Iozzia et al. [Iozzia et al., 2016] and Rodríguez et al. [Rodríguez et al., 2015] compared the time and frequency domain variability features obtained by PRV series extracted from rPPG with HRV parameters extracted from ECG signals. Results showed an overall agreement between time and frequency domain indexes computed on HRV and PRV series. However, some differences existed between resting and standing conditions as the experiments provided evidence that some differences exist between variability indexes extracted from HRV and video-derived PRV, mainly in the HF band during standing.

Remotely measured PRV applications. Although the remotely measured PRV are not yet possible to be used in critical medical analysis due to low frame rate and noise, many researchers have found that the remotely measured PRV can be used in stress detection and emotion detection.

Firstly, HRV has been studied by Acharya et al. [Acharya et al., 2006] as a sign of cardiac health and by Kim et al. [Kim et al., 2018] and Levenson et al. [Levenson, 2014] as a sign of ANS which can be used to reflect human stress and emotion. Therefore, the rPPG can be used for remote detection of stress and emotion.

Bousefsaf et al. [Bousefsaf et al., 2013] used this framework to detect the workload changes with interactive Stroop color word test. Similarly, MacDuff et al. [McDuff et al., 2014a] utilized several frequency features of remote PRV such as HF power, LF power and LF/HF to train the model on the dataset which included two emotion states, relaxed and stressed states. It turned out that PRV alone could achieve an accuracy of 70% to distinguish between these two states. Mitsuhashi et al. [Mitsuhashi et al., 2019] defined three different stress levels by the difficulty of the tasks. For instance, easy (mental arithmetic of $5*6$), middle (mental arithmetic of $13*16$), and difficult (mental arithmetic of $114*123$). With the KNN method to classify the stress modes, they showed that the accuracy of the classification is around 66%–83% for different stress levels. Kessler’s team [Kessler et al., 2017] made use of the rPPG framework for pain classification with a support vector machine (SVM) and Random Forest classifier. They showed that the LF part of frequency signal can be used to assess the pain. Belaiche et al. [Belaiche et al., 2019] used both micro-expressions and PRV to predict three emotion states, namely, happiness, disgust and anger. This study found that although the accuracy of the PRV based method is higher than the micro-expressions based method in emotion states recognition but the average accuracy is usually not higher than 60% for the dataset that contains sudden emotion change.

It can be concluded that the results of these applications are fairly good but not precise as the accuracy is lower than 90%. One of the possible reasons is that the peak detection of rPPG is critical as the rPPG signal is much more noisy than ECG signal. Therefore our task is to try to improve the precision of the peak detection as much as possible.

2.3.2/ PEAK DETECTION OF BLOOD VOLUME PULSE SIGNAL

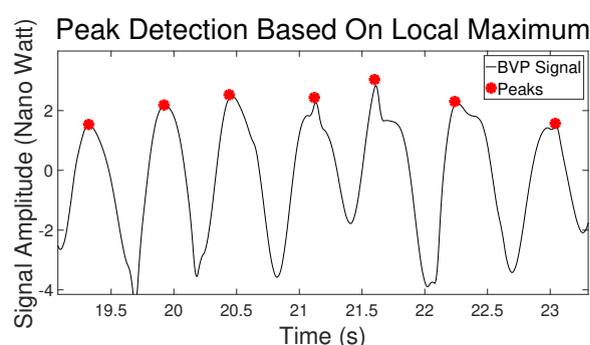


FIGURE 2.9 – An example of peak detection of BVP signal.

The PRV measurement is based on the peak detection of BVP signal, therefore the precision of peak detection directly affects the performance of PRV measurement (see Fig. 2.9). As explained before, the rPPG signal suffers from the low frame rate, the noise caused by sensors, the light illumination variation, and human motion. So the peak detection and filtering for rPPG signal is a particularly challenging task. Unfortunately, besides the conventional peak detection with rule-based Local Maximum [Fukunishi et al., 2018], there are not many works that focus on the peak detection and filtering of rPPG signal. So in this subsection, we discuss several research works that focus on peak detection of contact-based PPG signal and ECG signal. It is possible to adopt some of the algorithms for rPPG peak detection.

Zong et al. [Zong et al., 2003] proposed to use Slope Sum Function (SSF) to enhance the rising part of the ECG signal and reduce the falling part so that the shape of the signal is simplified and more clear to conduct peak detection. They used this method to detect the onset of the pulse wave. Jang et al. [Jang et al., 2014] adopted the SSF and used it in the contact PPG. It turned out the SSF could work well on contact PPG signal with some pre-processing and post-processing methods. Fu et al. [Fu et al., 2016] developed a multi-scale Gaussian smoothing-based strategy for accurate peak extraction. The strategy consisted of three stages: background drift correction, peak detection, and peak filtration. Background drift correction was implemented using a moving window strategy. The method is a variant of the system used by the MassSpecWavelet [Du et al.,], i.e., the peaks are found at local maximum values under various smoothing window scales. Since large smoothing window sizes can naturally filter instrumental noise, peak detection can be more accurately implemented compared with that achieved by wavelet analysis. The peak filtering strategy was performed to remove peaks with signal-to-noise ratios smaller than three. Elgendi et al. [Elgendi et al., 2013] proposed an event-related method that utilized the property of the PPG signal that the average height of the systolic peak period (the period where a peak appears) should be higher than the beat period (the period where a heart beat appears) and the systolic peaks are the highest points inside the peak periods. Thus the method detected the BVP peaks and addressed the non-stationary effects caused by severe exercise conditions in hot and humid environment.

Another possible strategy to detect peaks is to detect the zero-crossing points after applying first derivative or Hilbert Transform. With the derivative, the onset of a PPG signal can be related to a zero-crossing point before a maximal inflection, while the systolic peak is related to a zero-crossing point after that inflection. One of the properties of the Hilbert transform is that it is an odd function. That is to say that it will cross zero on the x-axis every time that there is an inflexion point in the original waveform. Therefore a crossing of the zero between consecutive positive and negative inflexion points in the original waveform will be represented as a peak in its Hilbert transformed conjugate.

Li et al. [Li et al., 2010] applied a first derivative estimate to obtain zero-crossing points that are used to evaluate the inflection points in the original PPG signal. Their algorithm firstly seeks the candidate zero-crossing points in the derivative estimate. For detection and denoising, it segments the filtered PPG signal into multiple equal divisions and then applies a selective window to the beginning of each division. Within those windows, the amplitudes and pulse rates are subsequently estimated and averaged as the initial thresholds. Then, the algorithm searches for pairs of inflection and zero-crossing points in

the PPG signal and their derivative in a beat-by-beat manner. Finally, within the beat evaluation phase, the algorithm then re-examines the PPG signals, and identifies candidate onsets and systolic peaks based on both amplitude and interval thresholds determined in the threshold estimation phase. Benitez et al. [Benitez et al., 2001] presented an algorithm for QRS detection using the first differential of the ECG signal and its Hilbert transformed data to locate the R wave peaks in the ECG waveform. Using this method, the differentiation of R waves from large, peaked T and P waves is achieved with a high degree of accuracy. In addition, problems with baseline drift, motion artifacts and muscular noise are minimised. Similarly, Valluraiah et al. [Valluraiah et al., 2015] proposed a method that the signal is preprocessed and is subjected to Hilbert transform along with a window to enhance the presence of the signal, to detect peaks by setting a threshold. The median filter is employed to remove the base line drift along with 35% of amplitude of the signal has been considered under thresholding.

2.4/ CONCLUSION

According to these related works, it can be concluded that the basic framework of rPPG measurement has been intensively studied and widely adopted. There are many existing ROI selection algorithms (face detection, skin classification and landmarks detection), RGB signal channel processing algorithms (BSS, color space analysis, etc.) and HR measurement algorithms. Some computer vision researchers used more complex cameras to improve the rPPG performance, however, the equipment is either expensive or not widely available in everyday life, such as the thermal cameras and polarization devices. The majority of the rPPG works do not focus on improving BVP peak detection, which is critical for PRV measurement. Biomedical and signal processing researchers used some novel algorithms to improve the contact ECG/PPG measurement, and majority of the algorithms have not been adopted by computer vision researchers in remote measurement, possibly due to completely different experimental conditions.

In this thesis, we firstly implement the conventional methods for rPPG methods in order to determine the best combination of the methods in the framework of physiological measurement. Then we propose a novel computer vision algorithm to get better ROI segmentation. After this we study and propose novel and effective algorithms to get precise peak detection for remote PRV measurement.

II

CONTRIBUTION

COMPARATIVE STUDY OF EXISTING rPPG METHODS

As we have described in the previous sections, the rPPG methods share a common pipeline (Fig. 2.1): 1) ROI selection for each frame, 2) RGB signals selection and combination to get the pulse signal, and 3) finally the physiological parameters extraction from the filtered pulse signal. In chapter 2, we have introduced multiple methods based on this pipeline, such as various face detection methods and skin detection algorithms for ROI detection and several channel selection and combination methods which were used to get the BVP signal, such as the blind source separation. It is possible to find several papers (e.g. [Christinaki et al., 2014] and [Li et al., 2018]) that compare either the ROI detection methods or the channel combination methods, but they have not studied the performance of different combination of the methods. In this chapter, the combination of three ROI segmentation methods and three channel selection methods are investigated in the framework of HR measurement. The objective is to find the best combination of the ROI segmentation method and RGB channel combination methods and lay the foundation for further studies.

Contents

3.1	Description of the algorithms implemented in the comparative study	36
3.1.1	ROI detection methods	36
3.1.2	RPPG signal selection and combination methods	37
3.2	Experiments	39
3.2.1	Dataset	39
3.2.2	Software	41
3.2.3	Evaluation metrics	42
3.2.4	Results and discussion	44
3.3	Conclusion	47

3.1/ DESCRIPTION OF THE ALGORITHMS IMPLEMENTED IN THE COMPARATIVE STUDY

3.1.1/ ROI DETECTION METHODS

Based on the state of the art established in the chapter 2, we observed that in most of the articles, ROI segmentation was based either on face detection, skin pixel detection or the use of a region defined by landmarks. It is of course not possible to compare all the variants of these three methods in this chapter, so we have chosen to implement and compare the performances obtained with the following 3 methods which have been used in several articles.

Face detection and tracking. In many papers (e.g. [Wiede et al., 2016], [Yang et al., 2016] and [Kalkhaire et al., 2016]), the regular and well-known Viola-Jones face detector [Viola et al., 2001] is used in conjunction with the KLT feature tracker [Lucas et al., 1981]. This method is now quite classical but presents satisfactory performance in the experimental configuration of the rPPG application where people often remain relatively static in front of a camera. Then, it is possible to observe that their implementations are very easily accessible via Matlab's computer vision toolbox or using OpenCV.

Skin segmentation. As we mentioned in chapter 2, there are many different skin detection methods. Some researchers (e.g. [Macwan et al., 2017]) used Conaire et al.'s [Conaire et al., 2007] method in their rPPG research work. This method adaptively chooses the thresholds for foreground detection for video frames in order to maximize the mutual information between the foreground maps of the images. A dynamic programming algorithm is used to efficiently investigate the search-space of all possible pairs of thresholds. Eventually, this method works effectively and very fast since it is based on a Look-Up-Table.

Landmark-based segmentation. Many scholars have used landmarks detection in rPPG methods (e.g. [Lempe et al., 2013], [Heusch et al., 2017] and [Tasli et al., 2014]). Kazemi et al. [Kazemi et al., 2014] proposed a method based on ensemble of regression trees that performs shape invariant feature selection while minimizing the same loss function during training time as we want to minimize at test time. They presented a natural extension of the method that handles missing or uncertain labels. Quantitative and qualitative results confirm that the method produces high quality predictions while being much more efficient than the best previous method for different datasets.

The examples of the three methods are shown in chapter 2, in Fig. 2.2 (a), Fig. 2.3 (b) and (d) respectively.

3.1.2/ RPPG SIGNAL SELECTION AND COMBINATION METHODS

As for the ROI segmentation, we have implemented three different rPPG signal selection and combination methods for comparison:

Green channel. According to Fig. 2.6, the wavelength around 550nm has the highest absorption coefficient for *HbO* and *Hb*. The wavelength of 550nm is the green channel. So the most straightforward way to estimate the rPPG signal from the RGB temporal traces is to simply make use of the green channel [Lempe et al., 2013].

Principal Component Analysis (PCA). As mentioned in chapter 2, the BSS methods, such as PCA and ICA, are the most widely used method for channel signal combination. In our study, we chose PCA [Wedekind et al., 2015] as the representative of BSS methods for rPPG. The main idea is to estimate three orthogonal linear combination of the three color channel with the PCA algorithm. The component that carries the pulse signal is a priori unknown. Commonly, the selection assumes that the pulse-signal shows the strongest periodicity and is selected in our implementation using the component with the highest signal-to-noise ratio (SNR).

Chrom. Another very popular algorithm is the one proposed by Haan and Jeanne by using the chrominance signals [De Haan et al., 2013]. They proposed to use the "skin-tone standardization", where the normalized skin tone, $[R,G,B]/\sqrt{R^2 + G^2 + B^2}$, is supposed to be the same for all under the white light $[R_s, G_s, B_s] = [0.7682, 0.5121, 0.3841]$.

With the $[R_s, G_s, B_s]$ being the standardized skin tone estimated coefficients. Let R_n, G_n and B_n be the RGB time series obtained after pre-processing. The coefficients correct the potentially nonwhite illumination by dividing the individual color channels by their means and next by multiplying $[R_n, G_n, B_n]$ with $[0.7682, 0.5121, 0.3841]$, i.e., the standardized RGB channels results as $R_s = 0.7682R_n, G_s = 0.5121G_n$ and $B_s = 0.3841B_n$. Then, they compute two intermediate signals X_f and Y_f projected on a plane orthogonal to the specular reflection. The idea being that the specular component of the reflected light does not contain pulsatile information useful for the estimation of the rPPG signal. We have:

$$\begin{aligned} X_f &= 3R_n - 2G_n, \\ Y_f &= 1.5R_n + G_n - 1.5B_n. \end{aligned} \tag{3.1}$$

These signals are then recombined following the alpha-tuning procedure:

$$\begin{aligned} S &= X_f - \alpha Y_f, \\ \text{where } \alpha &= \frac{\sigma(X_f)}{\sigma(Y_f)}. \end{aligned} \tag{3.2}$$

This method is based on a simple analytical formulation of the RGB combination and is thus very efficient. Moreover, thanks to the projection on a plane perpendicular to the

specular component of the reflected light, this method is quite robust to disturbances due to movement.

We have $3 \times 3 = 9$ combinations of methods to get the BVP signal. It is worth mentioning that it is not necessary to implement all the ROI detection and channel selection methods. For instance, the CROP methods proposed by Poh [Poh et al., 2010] and Liu et al. [Liu et al., 2014] were not as effective as other ROI segmentation algorithms because it eliminates the forehead which is useful for rPPG signal extraction. For the landmarks detection methods, we chose Kazemi's [Kazemi et al., 2014] algorithm because many other effective approaches do not provide the facial contour, such as the methods proposed by such as Zhang et al. [Zhang et al., 2014] and Asthana et al. [Asthana et al., 2014]. Fig. 3.1 shows the Kazemi's algorithm implemented by Dlib library [King, 2009] and Fig. 3.2 shows Asthana's algorithm. The latter gives the landmarks of eyes, nose and mouth but does not give the facial contour, so the ROI cannot be obtained.



FIGURE 3.1 – Landmarks detection by Kazemi (Source: blog.dlib.net/2014/08/real-time-face-pose-estimation.html)



FIGURE 3.2 – Landmark detections with no facial countour (Source: sites.google.com/site/chehrahome/home/)

Similarly, we chose the PCA to represent all the BSS methods, because Lewandowska et al.'s rPPG study showed that the PCA method could reduce the computation complexity significantly compared with the ICA methods with similar accuracy [Lewandowska et al., 2011]. Due to the limited experimental time in real life applications, the computational efficiency is taken into consideration. Therefore the ICA-based

methods are not implemented. Overall, only the methods that are most effective, efficient and representative are considered here.

3.2/ EXPERIMENTS

In this section, the dataset, experimental set up, software, evaluation metrics and results are described in details. Similar experimental procedures are adopted in the next chapters.

3.2.1/ DATASET

For the experiments, we used two datasets. One is the openly published MMSE dataset ¹, and the other one is the UBFC-RPPG ² dataset collected by ourselves.

MMSE Dataset. MMSE dataset was created by Zhang et al. [Tulyakov et al., 2016]. For the experiments, we used the 2D data of the MMSE dataset which consists of about 100 videos with the frame rate of 25 Hz. We adopted the dataset for several reasons: first of all, the dataset is large enough for testing. Secondly, the dataset includes different peoples, such as Europeans, Middle Easterners, South Asians, South Americans and East Asians which potentially makes it more challenging for the RGB signal processing. Thirdly, the contact PPG sensor was used to record the pulse signal as the ground truth which was synchronized with the videos so that it can be used to train the data and quantitatively assess the results. Lastly, with emotion elicitation in the experiments, the volunteers showed some movement of faces and heads which make the dataset more complicated and closer to the realistic scenarios. Similar to all the other rPPG experiments, the volunteers were asked to sit at a fixed distance from the web camera with a background board. Several sample images are shown in Fig. 3.3.



FIGURE 3.3 – Some sample images from the MMSE database.

1. <http://www.stulyakov.com/papers/cvpr2016.html>
2. <https://sites.google.com/view/ybenzeth/ubfcrppg>

UBFC-RPPG dataset. The UBFC-RPPG dataset [Bobbia et al., 2017] was created by Richard Macwan³ of ImViA lab with a contact PPG sensor, a web camera and a laptop:

- The contact PPG used in the experiment is a CMS50E Pulse Oximeter made by CONTEC company which utilizes the photoelectronic oxy-hemoglobin detection technology to measure human oxygen saturation and pulse rate. The data were collected from the volunteers' fingers. The timestamps (in millisecond), Heart Rate, the oxygen saturation and PPG signal are recorded by this contact sensor as the ground truth. The device is shown in Fig. 3.4.
- The web camera is HD pro webcam C920 made by logitech. It can be connected to a computer by the USB 2.0. In our experiment, the resolution is 640×480 and the frame rate is 30 fps. The web camera is shown in Fig. 3.5.
- The laptop is used to record and synchronize the webcam and contact PPG.



FIGURE 3.4 – The contact PPG sensor.



FIGURE 3.5 – The web camera.

55 students participating in the experiments. The students were made to sit one meter from the web camera. The focal length of the camera was adjusted to get the clear image. The index finger of each participant was stably put into the contact PPG sensor to get the ground truth. The experiment was done with ambient light. For each student, she/he was asked to play a time-sensitive game and a video of about one minute was recorded. Fig. 3.6 shows the experimental set up. Fig. 3.7 shows two examples of the videos recorded in the experiment. Although no dedicated light sources were adopted here, it can be observed that there is no shadow on the faces.

3. https://scholar.google.com/citations?hl=fr&user=JOrxAzUAAAAJ&view_op=list_works

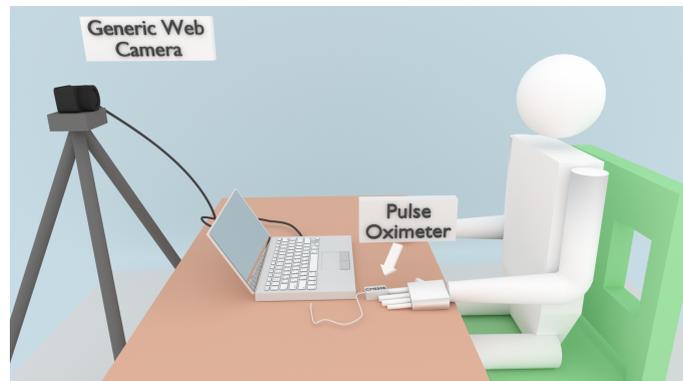


FIGURE 3.6 – Experimental set up.



FIGURE 3.7 – The examples of UBFC-RPPG dataset.

3.2.2/ SOFTWARE

System framework. As shown in Fig. 2.1, for each video frame, the segmented ROI is spatially averaged to obtain the RGB values. The result of this process is a RGB time signal. The RGB temporal traces are then pre-processed by zero-mean normalization, de-trended using smoothness priors approach and bandpass filter with cutoff frequencies 0.7 Hz and 3.5 Hz. The rPPG BVP signal is then extracted using the channel signal selection and combination methods. With this processing, the beats of the BVP versus time can be found by peak detection. The HR is usually computed in two different ways: 1) Set a small window (T seconds) and move it along the BVP signal. Count the number (N)

of the beats within this window. Then HR is calculated as $60 \times N/T$. 2) Get the PSD of the BVP signal and detect the peak of the PSD. The frequency of the peak is considered as the HR. For instance, if the frequency of the peak of the PSD of BVP is 1.5 Hz. The HR is $1.5 \times 60 = 90$ beats/minute. Practically, the peak detection is a difficult task for rPPG signal due to limited frame rate, human movement, light variation and sensor noise, we estimated HR by Welch's method and estimated the periodogram over a 20 seconds moving window, with a step size of one second. The same HR estimation procedure are used on the PPG signal recorded with the contact sensor and on the rPPG signal given by different combination of methods.

The implementation of the algorithms. Most of these algorithms are implemented with Matlab. It is worth mentioning that the skin/non-skin detection has been implemented by the authors [Conaire et al., 2007]⁴. The Viola-Jones algorithm with tracking is proposed by Matlab library⁵. The facial landmark detection proposed by [Kazemi et al., 2014] was implemented with the Dlib C++ library⁶. This Dlib implementation was selected, because it processes much faster than other implementations, and it provides the facial contour that many other implementation do not have. There are 68 landmarks (Fig. 2.4), with which the desired ROI can be found within the facial contour without the eyes and the mouth. Fig. 2.3 (d) shows the ROI of the student obtained based on the landmarks. The black parts are the pixels that are eliminated. It can be seen that the background, eyes and the mouth were discarded.

3.2.3/ EVALUATION METRICS

To evaluate the performance of the rPPG method compared with the ground truth, the signals of the web camera and contact PPG were synchronized. We propose to use some evaluation metrics: Pearson correlation coefficient (R), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and mean Signal-to-noise ratio (meanSNR), etc.

1. Pearson correlation coefficient (R). The R is the correlation between the actual outcomes and the predicted outcomes. It can be expressed as:

$$R = \frac{\sum_{i=1}^n (HR_{PPG}(i) - \overline{HR_{PPG}})(HR_{rPPG}(i) - \overline{HR_{rPPG}})}{\sqrt{\sum_{i=1}^n (HR_{PPG}(i) - \overline{HR_{PPG}})^2} \sqrt{\sum_{i=1}^n (HR_{rPPG}(i) - \overline{HR_{rPPG}})^2}} \quad (3.3)$$

where n is the number of estimations collected from the entire dataset, HR_{PPG} is the ground truth HR, and HR_{rPPG} is the HR calculated by the rPPG signal. This feature can be used to evaluate the correlation between the remote PPG result with contact PPG measurement. Fig 3.8 displayed one example of the correlation assessment of the HR from remote method compared with the ground truth. An ideal algorithm should have all the points onto the diagonal line of 45 degrees and the slope coefficient should be very close to 1.

4. <http://clickdamage.com/sourcecode/index.php>

5. fr.mathworks.com/help/vision/examples/face-detection-and-tracking-using-the-klt-algorithm.html

6. <http://blog.dlib.net/2014/08/real-time-face-pose-estimation.html>

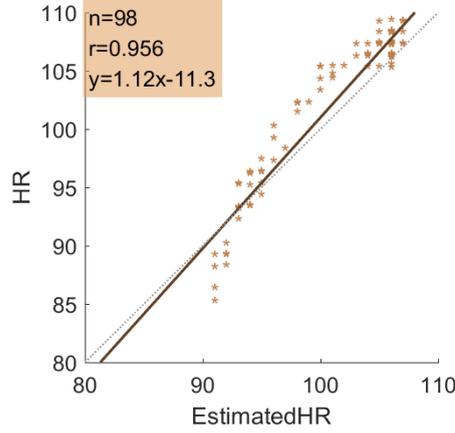


FIGURE 3.8 – Correlation plot between the HR estimated from the rPPG signal (EstimatedHR) and from the contact sensor (HR). n is the number of points.

2. MAE (in second). The Mean Absolute Error (MAE) is the average of the absolute differences between the estimated values and the ground truth where all individual differences have equal weight:

$$MAE = \frac{1}{n} \sum_{i=1}^n |HR_{PPG}(i) - HR_{rPPG}(i)| \quad (3.4)$$

where HR_{PPG} is the ground truth HR, HR_{rPPG} is the estimated HR value from rPPG signal, and n is the number of points of the rPPG signal.

3. MAE5 (in second). The MAE is not a perfect metric because actually it is strongly affected by outliers. A high-quality signal with some extreme outliers may have a worse MAE value than a low-quality signal with fewer extreme outliers. MAE5 is a metric which discards all the outliers with the error bigger than 5. It can be expressed as:

$$MAE5 = \frac{1}{n - n_{out}} \sum_{i=1}^{n-n_{out}} |HR_{PPG}(i) - HR_{rPPG}(i)|, \text{ with } |HR_{PPG}(i) - HR_{rPPG}(i)| < 5 \quad (3.5)$$

where n_{out} is the number of outliers.

4. Mean5. This metric is proposed to calculate the percentage of estimations where the absolute error is under 5 bpm. It is used to assess the quality of a signal. The high-quality signal should have fewer outliers and have a bigger Mean5 value. The estimations with error bigger than 5 are considered as outliers.

5. Mean2.5. Similarly, the Mean2.5 is the percentage of estimations where the absolute error is under 2.5 bpm. The estimations with error bigger than 2.5 are considered as outliers.

6. RMSE (in second). RMSE is Root Mean Square Error. It is the square root of the average of squared differences between prediction and actual observation. It is expressed

as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (HR_{PPG}(i) - HR_{rPPG}(i))^2} \quad (3.6)$$

7. MeanSNR (in dB). MeanSNR represents the average of all the SNR defined as the ratio of the power of the main pulsatile component and the power of background noise, computed in dB due to the wide dynamic range of the signals. The SNR is estimated with:

$$SNR(i) = 10 \log_{10} \left(\frac{\int_{f1}^{f2} h_{signal}(f) |\mathcal{F}\{S_i(t)\}|^2 df}{\int_{f1}^{f2} h_{noise}(f) |\mathcal{F}\{S_i(t)\}|^2 df} \right) \quad (3.7)$$

where $\mathcal{F}\{S_i(t)\}$ is the Fourier transform of the i^{th} rPPG signal $S_i(t)$, $h_{signal}(f)$ and $h_{noise}(f)$ are the signal parts and noise parts respectively, $f1$ and $f2$ are the lower and upper limit of the integral defined by the possible physiological range of the heart rate (40 to 240 bpm in our case). The signal part $h_{signal}(f)$ is estimated using the peak of the periodogram of the contact PPG signal and $h_{noise}(f) = 1 - h_{signal}(f)$. As shown in Fig. 3.9, the red part (main pulsatile) is considered as the signal and the blue part is considered as noise.

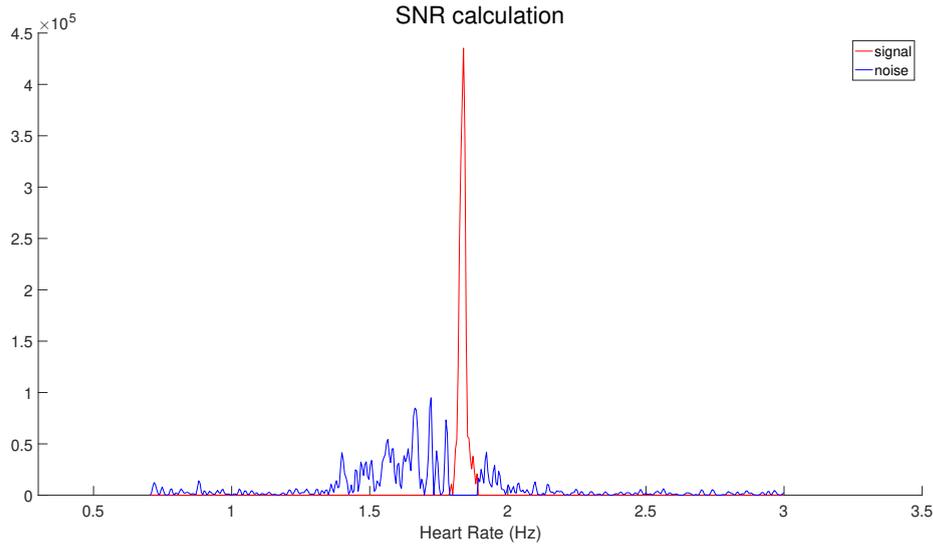


FIGURE 3.9 – The PSD of the rPPG signal. The red part (main pulsatile) is the signal and the blue part is the noise.

3.2.4/ RESULTS AND DISCUSSION

The experimental data was tested with the combination of the three ROI segmentation methods and the three channel combination methods. The average values of metrics for MMSE dataset and UBFC-RPPG dataset are shown in Tables 3.1 and 3.2 respectively.

According to both tables, it can be seen that the results are heterogeneous, the skin detection is best for ROI detection, Chrom is the better method for channel selection method

TABLE 3.1 – The average evaluation values of MMSE dataset.

Methods	R	Mean2.5	Mean5	MAE	MAE5	RMSE	MeanSNR
Skin/Chrom	0.725	0.611	0.858	4.42	3.15	6.10	2.94
Skin/Green	0.402	0.533	0.604	9.22	3.34	12.11	0.72
Skin/PCA	0.459	0.482	0.594	9.58	4.01	13.12	0.24
Face/Chrom	0.508	0.573	0.672	7.36	3.27	9.42	1.32
Face/Green	0.252	0.414	0.536	17.84	3.98	21.12	1.41
Face/PCA	0.151	0.372	0.481	23.13	2.07	22.26	2.12
Landmarks/Chrom	0.583	0.502	0.718	6.21	4.26	8.13	1.04
Landmarks/Green	0.184	0.320	0.441	16.51	5.43	20.82	0.91
Landmarks/PCA	0.202	0.363	0.484	14.16	4.92	18.18	1.29

TABLE 3.2 – The average evaluation values of UBFC-RPPG dataset.

Methods	R	Mean2.5	Mean5	MAE	MAE5	RMSE	MeanSNR
Skin/Chrom	0.800	0.722	0.928	2.88	1.48	3.81	4.63
Skin/Green	0.460	0.603	0.790	7.73	1.62	10.83	1.01
Skin/PCA	0.481	0.579	0.763	8.61	1.72	11.53	0.49
Face/Chrom	0.614	0.644	0.843	5.82	1.58	8.61	1.97
Face/Green	0.296	0.456	0.612	15.05	1.62	18.33	1.77
Face/PCA	0.139	0.356	0.500	20.98	1.64	24.73	3.16
Landmarks/Chrom	0.719	0.697	0.909	3.23	1.54	4.98	3.08
Landmarks/Green	0.375	0.461	0.623	12.65	1.71	16.53	2.74
Landmarks/PCA	0.315	0.451	0.629	13.36	1.85	17.66	1.56

in both dataset and the combination of skin + chrom performs the best among all the combination of methods. The results of both dataset show similar tendency. For instance, the MAE and RMSE of the Skin/Chrom methods are the lowest in both dataset. This method has the highest Mean2.5 and Mean5 of all, which means it has very few outliers. The landmarks detection with Chrom also have a quite low MAE and RMSE for UBFC-RPPG dataset, but it has a lower Mean2.5 and Mean5. This means there are more outliers of the landmarks detection than the skin detection with Chrom method. The Skin/Chrom combination also has the highest correlation R , which is much closer to 1 than other methods, and this means the correlation of Skin/Chrom methods between the remote method and the ground truth is very high. The mean SNR of the Skin/Chrom method is the highest as well, which means the performance of the remote PPG is very close to the ground truth

in the frequency domain. According to the results of both datasets, the underperformed combination is the face detection with PCA. The MAE5 of Face/PCA is low, but the MAE is very high and the Mean2.5 and Mean5 are very low. This means the Face/PCA method has a quite large number of outliers. Fig. 3.10 shows an example of relatively accurate HR measurement obtained with Skin/Chrom from one video of the UBFC-RPPG dataset. The remote signal is very close to the ground truth with very few outliers. Fig. 3.11 shows an example of imprecise HR signal obtained by Skin/Green for the same video, and it can be seen that the remotely measured HR is completely different with the ground truth from 36 to 56 second.

It is interesting that the results of UBFC-RPPG are better than those of MMSE dataset. This may be because the MMSE dataset has more head movements and the signal is much more noisy.

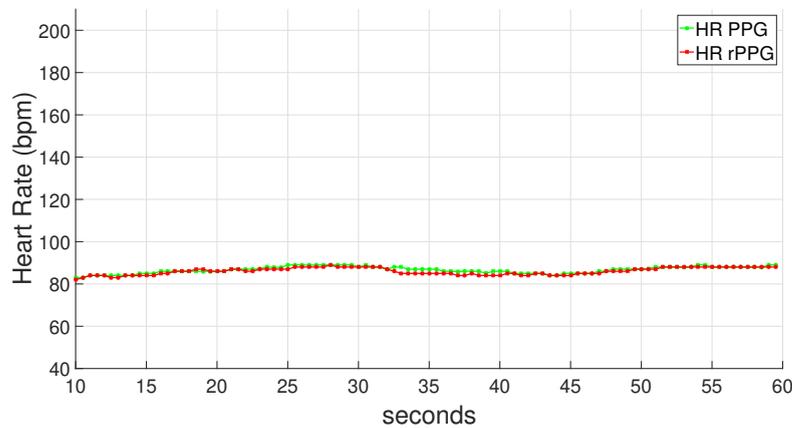


FIGURE 3.10 – An example of relatively accurate measurement generated from UBFC-RPPG dataset (the red signal is the HR obtained by rPPG and the green signal is the ground truth).

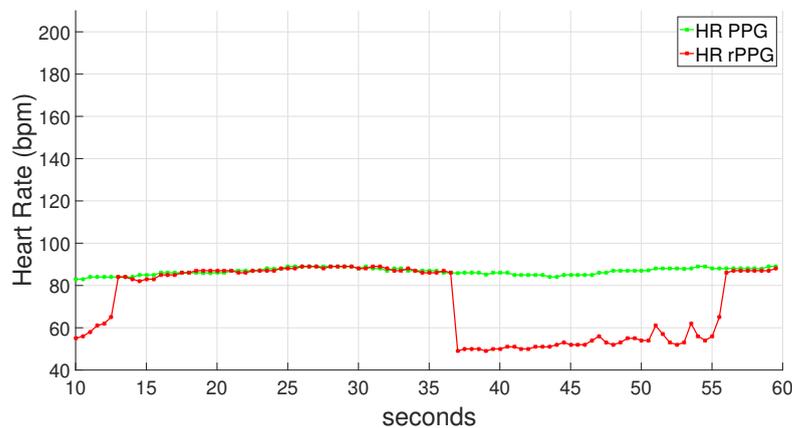


FIGURE 3.11 – An example of imprecise measurement generated by UBFC-RPPG dataset (the red signal is the HR obtained by rPPG and the green signal is the ground truth).

3.3/ CONCLUSION

The rPPG for physiological parameters measurement is much more complicated than the contact PPG methods because there is signal noise and external influences from the environment. Therefore accurate ROI detection and effective channel signal selection methods are very critical for the rPPG applications. Previous researchers have already proposed some methods for these tasks, however, the performance of the combination of the different ROI detection and channel selection methods are not investigated. To obtain the best combination and lay a solid foundation for the further studies, we have studied the algorithms and implemented three state-of-the-art methods for the ROI detection and three rPPG channel selection and combination methods. To test the performance of these methods, we adopted two experimental datasets which were set up with a low-cost web camera, a contact PPG sensor and a laptop. Several evaluation metrics were used to assess the performance of different methods. The results show that Conaire's skin detection [Conaire et al., 2007] with Chrom method [De Haan et al., 2013] is the best among all. It has the highest SNR value and lowest error and obtains high-quality signal with few outliers. The facial landmarks detection with Chrom is not incorrect, but the error is greater than the Skin/Chrom combination and has more outliers. As we have mentioned, the two main objectives of the PhD project are: 1) improve the ROI detection for rPPG framework, 2) improve the peak detection for BVP signal to get better PRV signal. This chapter has laid a solid foundation to achieve the two objectives. Since the Chrom performs better than other channel selection methods, we will use it as the channel signal selection methods in the next steps. The same experimental datasets and similar evaluation metrics are adopted in the next steps as well.

MODEL-BASED REGION OF INTEREST SEGMENTATION FOR rPPG

Existing research works have indicated that the ROI detection has significant effect on the performance of rPPG methods [Bousefsaf et al., 2013] [Bobbia et al., 2016]. This is because the numerical error of the signals will increase if the number of the effective pixels, i.e. skin pixels, is too small. Conversely, the number of non-skin pixels also has a significant negative influence on signal quality. Moreover, it has been shown that rPPG signal is not distributed homogeneously on the face. Some skin regions contain more rPPG signal than others. For example, the cheeks and forehead contain much more effective information than other areas of the face [Kwon et al., 2015]. Therefore, the improvement of ROI selection based on the importance of different regions has great potential. In this chapter, we describe the contribution to improve the ROI segmentation by giving proper weight to each pixel.

Contents

4.1	Introduction	50
4.2	Model-based ROI segmentation	51
4.3	Experiments and results	53
4.4	Conclusion	56

4.1/ INTRODUCTION

In chapter 2, we have introduced several existing ROI segmentation methods. In chapter 3, we have shown that Conaire’s skin detection algorithm [Conaire et al., 2007] is the most effective ROI selection methods among all methods. In the previous rPPG research works, the detected ROI pixels of each frame are spatially averaged to get the RGB channels with which the BVP signal is finally obtained. However, there is one potential problem in this conventional approach, i.e., all the ROI pixels are assumed to contain the same amount of rPPG information with the same quality. This is a wrong assumption as Kwon et al. has indicated that the cheeks and forehead contain much more information than other areas of the face [Kwon et al., 2015]. The causes for this uneven signal distribution may be that the facial blood vessels are not homogeneously distributed, and the concentration of melanin and hemoglobin are not the same in different tissue regions. As illustrated in Fig. 4.1, the facial blood vessels distribution are not homogeneous.

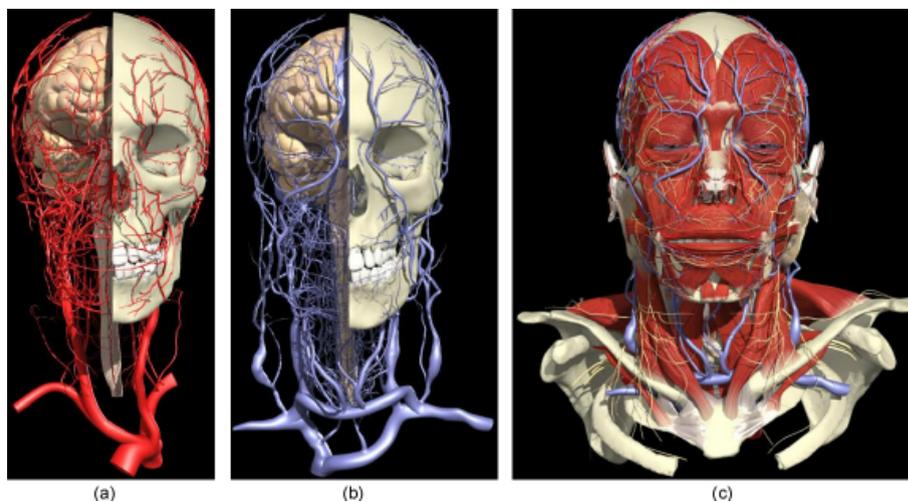


FIGURE 4.1 – Superficial facial blood vessels distribution: a) arteries, b) veins, c) arteries and veins together (image from [Buddharaju et al., 2008]).

From this premise, some studies have suggested to simply select certain sub-regions which contain more information, such as the cheeks and/or forehead, as ROI [Scalise et al., 2012]. This selection has the undeniable advantage of simplicity but it is also possible to weight the skin pixels by their relative importance.

In this chapter, we propose a model-based ROI segmentation that explicitly favors the most important facial regions. The model describing the spatial distribution of rPPG information on the face was trained with ten videos. This model is then used to weight the pixels during spatial averaging. This approach has been validated using our in-house publicly available video dataset which is introduced in the previous chapter. We show that this modification in how the spatial averaging of the ROI pixels is calculated can significantly increase the final rPPG signal quality compared with other state of the art methods such as Conaire’s skin detection.

4.2/ MODEL-BASED ROI SEGMENTATION

The proposed technique for explicitly favoring certain areas of the face during the spatial averaging step of RGB pixels works within the regular rPPG framework presented in Fig. 2.1.

Obtaining the model. The model that encapsulates the spatial distribution of rPPG information was trained using an in-house database of 10 videos recorded under very favorable conditions. For this experiment, we used a EO-23121C camera recording 1024×768 uncompressed images at 30 fps. The average length of each video is about one minute. Subjects sat on a chair with back support. To make sure that the face is fixed in a specified position, we used a shelf and asked the volunteers to put the heads onto the shelf. Fig. 4.2 shows two sample images from the dataset.



FIGURE 4.2 – Two examples of the subjects that were used to get the model.

We estimate the SNR for each pixel of the 10 subjects, thus a SNR map for each subject can be obtained. The model can then be generated as the average of the 10 SNR maps. To do this, the faces of our database should be aligned and resized. In our implementation, the face sequence is aligned based on the location of the eyes. This process relies on the facial landmarks to obtain a normalized rotation and scaled representation of the face sequences which are centered in the image, rotated such that the eyes lie on a horizontal line and scaled so that the size of the faces are approximately identical.

To get the SNR of each pixel, we extracted the rPPG signal of the pixel and obtained the PSD with FFT, and the SNR is estimated as the ratio of the area surrounding the maximum peak in the PSD, divided by the area under the rest of the PSD curve:

$$SNR_{(x,y)} = 10 \log_{10} \left(\frac{\int_{f_1}^{f_2} h_{signal}(f) |\mathcal{F}\{S_{(x,y)}(t)\}|^2 df}{\int_{f_1}^{f_2} h_{noise}(f) |\mathcal{F}\{S_{(x,y)}(t)\}|^2 df} \right) \quad (4.1)$$

where $S_{(x,y)}(t)$ is the signal value in pixel location (x, y) at time t , and $\mathcal{F}\{S_{(x,y)}(t)\}$ is the Fourier transform of the rPPG signal of the pixel in the entire video sequence. Fig. 4.3 shows the two SNR maps of the videos presented in Fig. 4.2. Eventually, the general

model can be obtained as the average of the 10 aligned SNR maps. Fig. 4.4 presents the spatial distribution of the SNR values averaged for all the videos of this dataset. As expected, we can observe that cheeks and forehead have in average, higher SNR (about 6 dB) than other face locations (e.g. 0 dB for the chin).

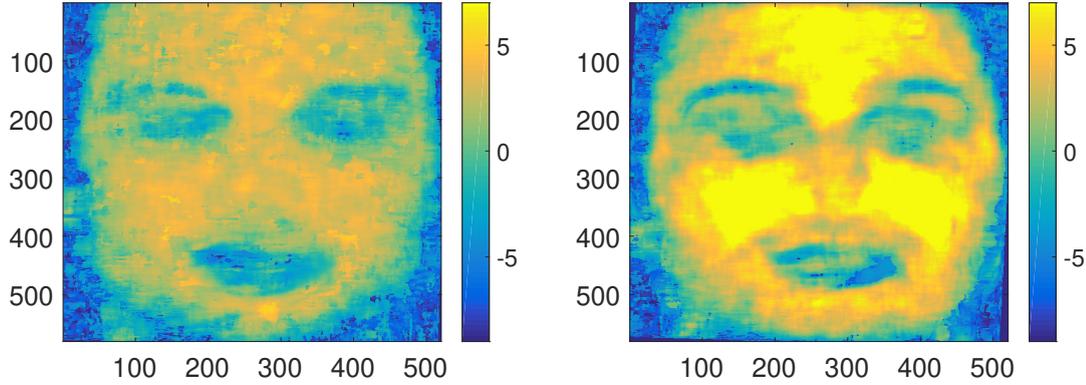


FIGURE 4.3 – The SNR map of the two subjects. The yellow parts have higher SNR and the green/blue parts have lower SNR.

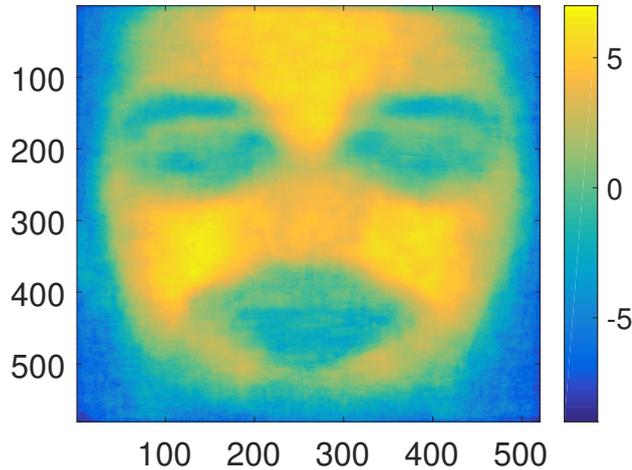


FIGURE 4.4 – The model. The yellow parts have higher SNR and the green/blue parts have lower SNR.

Applying the model on the video sequence. This spatial map is used during the spatial averaging of the pixels of the ROI. Instead of using equal weights for all pixels in the ROI, we use a weighted average where weights are defined based on our model. The weights are calculated as:

$$\omega_{(x,y)} = \frac{a^{SNR_{(x,y)}}}{\sum_{(x,y)} a^{SNR_{(x,y)}}} \quad (4.2)$$

where a is a constant. With higher a , the weights of the region of higher SNR are higher. The optimal selection of a is done empirically. Experiments and details are given in Section 4.3.

Eventually, RGB triplet at time t is obtained for each frame with the following weighted average:

$$S_{RGB}(t) = \sum_{x,y} I_t(x,y) \times \omega_{(x,y)} \quad (4.3)$$

where $I_t(x,y)$ is the RGB value at time t of a pixel at location (x,y) and $\omega_{(x,y)}$ is the corresponding weight of the pixel.

It is interesting to note that this technique can be advantageously combined with all the rPPG methods that perform a spatial average of the ROI pixels.

4.3/ EXPERIMENTS AND RESULTS

As we have shown in the Chapter 3, Conaire's skin detection [Conaire et al., 2007] is the best state-of-the-art ROI segmentation method. Therefore, the proposed model-based ROI segmentation (later called *model*) is experimentally compared with skin detection (*skin*). The result of face detection (*face*) is also presented as a reference. We used the same system framework described in Fig. 2.1, and the channel selection method is Chrom [De Haan et al., 2013] as we have validated that this is the most efficient and effective algorithm for RGB channel combination in Chapter 3. Then the BVP signal are obtained and the HR can be calculated. The same video datasets (UBFC-RPPG and MMSE) and metrics described in Chapter 3 are used in this experiment. Note that the results in this chapter should not be compared with the results in Chapter 3 because some internal parameters have been changed in this implementation.

The results of UBFC-RPPG dataset. First, we empirically select the optimal scalar a of equation 4.2. To do this, we tested some values for a in the system framework with the metrics presented in Chapter 3 and finally $a = 2$ was selected.

Second, the proposed model-based ROI segmentation method is compared with other state-of-the-art methods. The average results are shown in table 4.1. It is obvious that the ROI segmentation algorithms have significant effects on the results. All the metrics are very different with different ROI selection methods. It is also very clear that the proposed algorithm performs much better than all the other algorithms. It has the lowest MAE5 and RMSE. It has the highest correlation R, which means the rPPG method matches the ground truth better than other methods. SNR metric is also the highest which indicates that the method offers the best quality of the signal in frequency domain. The *skin* method is actually very effective, but it is slightly worse than the proposed algorithm.

In the experiments, we noticed some interesting cases. Firstly, the rPPG signal of one video is very clear with very little noise, because the volunteer has no beard, no glasses, did not move and sat at the precise location of the focal length of the camera. The HR of this

TABLE 4.1 – The average evaluation values for different ROI detection methods with UBFC-RPPG dataset.

	R	Mean2.5	Mean5	MAE	MAE5	RMSE	SNR
<i>face</i>	0.606	0.612	0.766	8.72	1.50	12.40	0.57
<i>skin</i>	0.800	0.716	0.851	5.10	1.36	6.78	3.21
<i>model</i>	0.816	0.750	0.872	3.99	1.29	5.55	3.35

video obtained by *face*, *skin* and *model* are shown in Fig. 4.5 from top to bottom respectively, where the red signals are rPPG measured HR and the green signals are ground truth. It can be seen that the rPPG signals from all the three methods matched the contact signal very well. With this video, all three methods gave similarly fine results with the proposed method being slightly better. The MAE is 2.35 and Mean5 is 0.98 for *model*. The MAE is 2.73 and the Mean5 is 0.98 for *skin*. For *face*, the MAE is 2.96 and the Mean5 is 0.95. Fig. 4.6 shows the enlarged details of the results given by *face* and *skin* respectively from 15 to 20 seconds. Fig. 4.7 shows the enlarged details of the results given by *skin* and *model* respectively from 20 to 30 seconds. It is clear that *skin* is more accurate than *face*, and *model* is more accurate than *skin* in these periods.

Secondly, in another video where the volunteer’s head kept moving at the beginning, we observed that the simple *face* detection basically fails while *skin* and *model* succeeds for the same period. This is because the movement of the head significantly reduces the number of the effective skin pixels for face detection, while the negative effects on the skin detection and the proposed method is minimized. Fig. 4.8 shows that in this case the HR signal calculated by face detection (the top image) of the beginning 5 seconds is incorrect. It can be seen that the *model* only gives slightly better results than *skin*, and this is because the beard of the volunteer affects the performance of the proposed method. It implies that the combination of *model* and *skin* may be the ideal solution for ROI detection.

Fig. 4.9, 4.10 and 4.11 show the correlation plot of the HR estimation given by *face*, *skin* and *model* respectively for the entire dataset. It can be seen that the correlations of both *skin* and *model* with the ground truth are much stronger than *face*, and *model* is slightly better than *skin*.

The results of MMSE dataset. Similarly, we empirically selected the optimal scalar $a = 2$ of equation 4.2 for MMSE dataset. The results are shown in Table 4.2. According to the results, it can be seen that both *skin* and *model* performs much better than *face* in this dataset. However, the proposed method is not much better than *skin*. It has higher SNR and Mean2.5 and lower MAE5, but it gives slightly lower correlation R, Mean5 and slightly higher error for MAE and RMSE. This is because in this dataset, there are much more large head movement, therefore the landmarks which are used to align and resize the face sequences may be not precise, and the performance of the proposed method is negatively affected in this process.

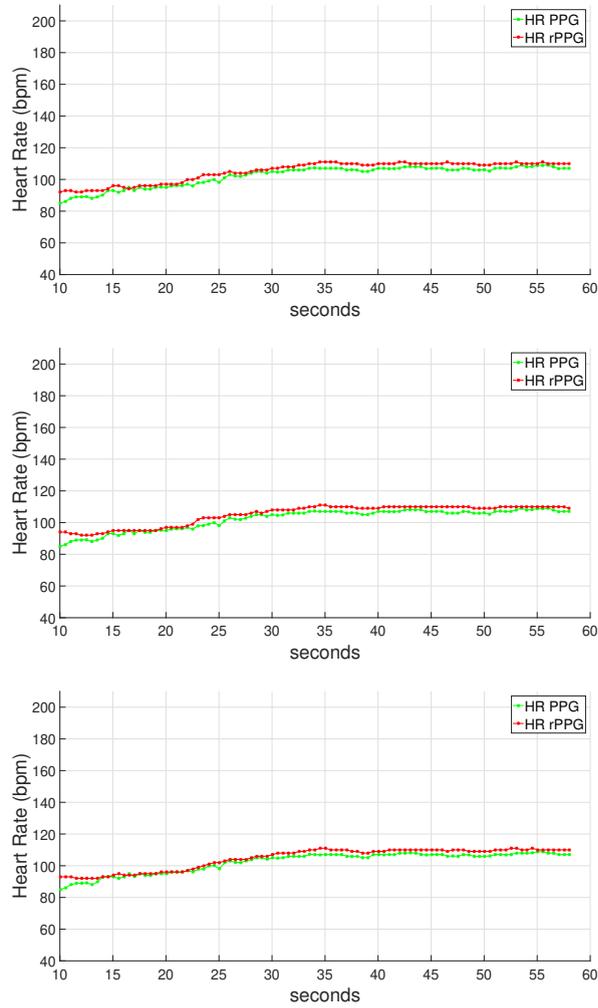


FIGURE 4.5 – The signals obtained by three methods in one case. From top to bottom are *face*, *skin* and *model* respectively. The red signals are rPPG measured HR and the green signals are ground truth.

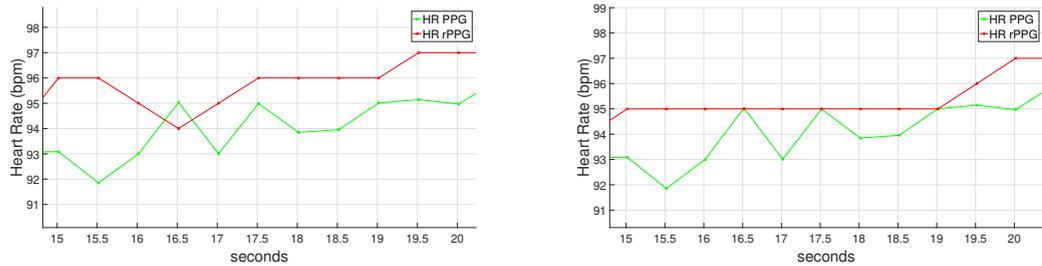


FIGURE 4.6 – The enlarged details of the HR signals obtained by *face* and *skin* from 15 to 20 second. The left image is *face* and the right image is *skin*. The red signals are rPPG measured HR and the green signals are ground truth.

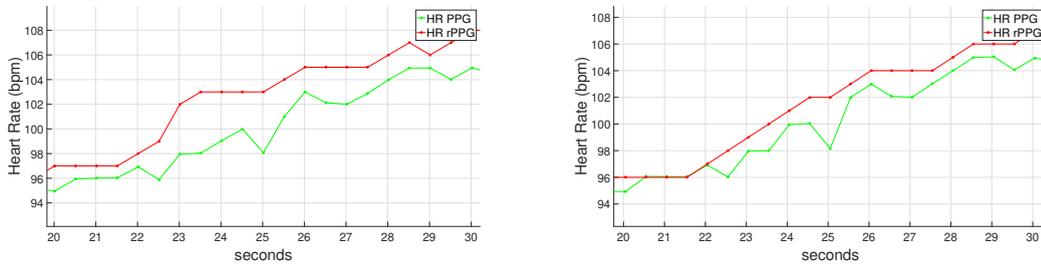


FIGURE 4.7 – The enlarged details of the HR signals obtained by *skin* and *model* from 20 to 30 second. The left image is *skin* and the right image is *model*. The red signals are rPPG measured HR and the green signals are ground truth.

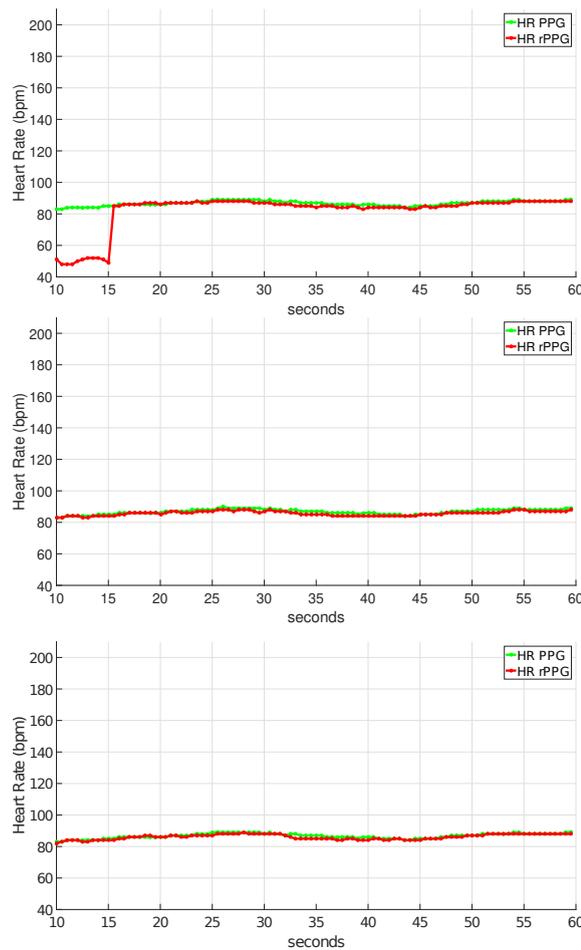


FIGURE 4.8 – The signals obtained by three methods in the second case. From top to bottom are *face*, *skin* and *model* respectively. The red signals are rPPG measured HR and the green signals are ground truth.

4.4/ CONCLUSION

The ROI for rPPG methods must contain as much useful information as possible. Most of the state-of-the-art methods focus on the improvement of the face and skin detection, and

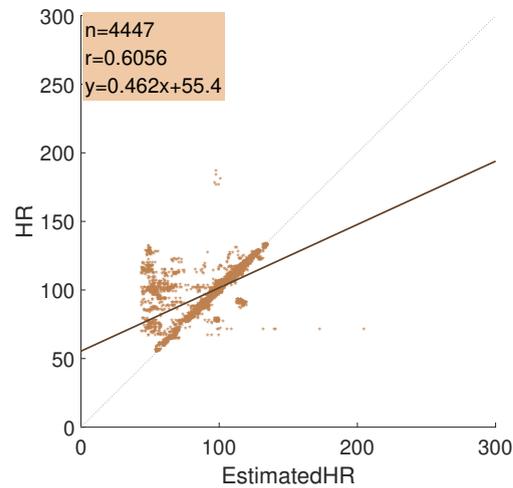


FIGURE 4.9 – The correlation plot of the HR estimation given by *face*. The x axis is the estimated HR and the y axis is the ground truth.

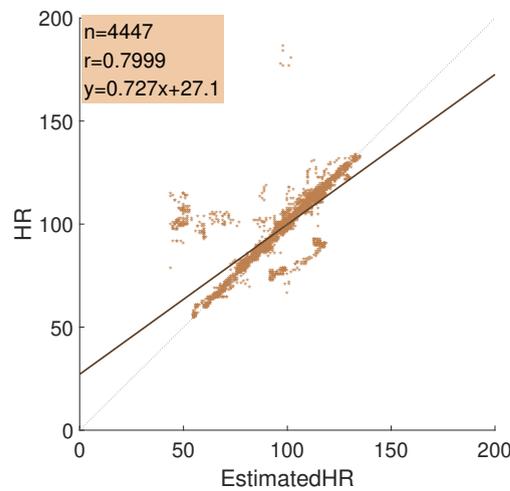


FIGURE 4.10 – The correlation plot of the HR estimation given by *skin*. The x axis is the estimated HR and the y axis is the ground truth.

TABLE 4.2 – The average evaluation values for different ROI detection methods with MMSE dataset.

	R	Mean2.5	Mean5	MAE	MAE5	RMSE	SNR
<i>face</i>	0.473	0.451	0.473	13.91	7.02	17.60	0.33
<i>skin</i>	0.727	0.603	0.799	8.10	4.16	9.32	1.79
<i>model</i>	0.705	0.632	0.781	6.88	5.41	9.97	1.89

their objectives are mostly to get as many skin pixels as possible and to discard as many non-skin pixels as possible. However, the rPPG signals are not distributed homogeneously on the human face due to the distribution of blood vessels, therefore it is reasonable to use weights on the pixels before other processing. In this chapter, we present a model based on SNR weights to improve the ROI precision. To get the model, we did experiments with a

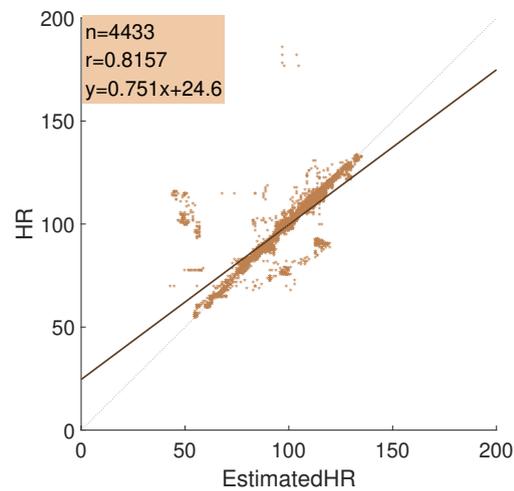


FIGURE 4.11 – The correlation plot of the HR estimation given by *model*. The x axis is the estimated HR and the y axis is the ground truth.

camera and a contact PPG sensor as the ground truth. We recorded 10 high-quality videos with favourable conditions and used them to create the SNR weight map. The datasets introduced in Chapter 3 were tested with this model in the framework of HR measurement. The results showed that the ROI segmentation affects the HR measurement significantly and our new algorithm performed better than the other state-of-the-art methods in UBFC-RPPG dataset. However, the proposed method is not much better than Conaire’s skin detection [Conaire et al., 2007] in the MMSE dataset, because there are significantly more head movements in this dataset. The landmarks are not precise when the head movement is very large and the model generated by the method is not aligned properly with the face sequences, and this has negative influence on the performance.

With the proposed ROI detection method and the Chrom channel combination, we can proceed to do the peak detection on the obtained BVP signal and get the PRV. This will be described in the next chapter.

REMOTE PRV MEASUREMENT

After the model-based ROI detection and Chrom color channel combination, the BVP signal is obtained. The PRV can then be estimated with the BVP peak detection. Unlike the HR, the PRV is sensitive and not precisely measured in remote methods due to the low frame rate of the camera, the movement, the noise of the sensors, the light variation, etc. For instance, in our case, the BVP signal of MMSE dataset is 25 Hz and we interpolated it with the frame rate of 200 Hz. Therefore, improving the precision of the PRV estimation is a significant task. However, the research related to this topic is very limited. The existing works are still based on the conventional Local Maximum and the parameters setting were not discussed in the papers [Fukunishi et al., 2018] [Huang et al., 2016]. There have been more research works for the peak detection with contact equipment, for instance, the famous SSF method [Zong et al., 2003] has been effectively used for contact measurement (e.g. ECG) of HR [Rankawat et al., 2015] [Jang et al., 2014]. But it has not been investigated in the rPPG methods. In this chapter, we focus on the improvement of the BVP peak detection and remote PRV measurement. Firstly we validate that SSF can be used in the rPPG framework. Then we proposed two novel methods, namely adaptive SSF and Adaptive Two-window Peak Detection to improve the performance. Some features and metrics are calculated to assess these methods. We show that the proposed methods are more effective than existing methods.

Contents

5.1	Introduction	60
5.2	Slope Sum Function	60
5.2.1	Algorithm	60
5.2.2	Experiments	61
5.2.3	Results	62
5.3	Adaptive Slope Sum Function	67
5.3.1	Algorithm	67
5.3.2	Experiments and results	68
5.4	Adaptive Two-window Peak Detection	71
5.4.1	Algorithm	71
5.4.2	Experiments and results	74
5.5	Conclusion	77

5.1/ INTRODUCTION

The HRV has been investigated by the medical researchers as a biomaker for the ANS [Evans et al., 2013]. Several features of the HRV can be used to detect the human emotion states. For instance, the LF feature of PRV reflects both sympathetic and parasympathetic activity of the ANS and the HF feature reflects the parasympathetic branch of the ANS. According to the previous research, the PRV measured by PPG techniques can be a surrogate measurement of HRV in some conditions [Gil et al., 2010]. In some rPPG works, the spectral features of PRV were used to detect the cognitive stress [McDuff et al., 2014a]. Some other PRV features obtained by rPPG in time domain such as Standard Deviation and Root Mean Square of the Successive Differences are possible to be used to determine different human emotion states as well [Michels et al., 2013].

One challenge of the rPPG method is to precisely measure the PRV based on BVP peak detection as the performance could suffer from low frame rate, light variation, and human movement under remote conditions. The existing works are very limited for this topic. Majority of them are based on conventional Local Maximum peak detection [Huang et al., 2016][McDuff et al., 2014a][Fukunishi et al., 2018] which may fail in complex signals. To address this issue, we proposed three methods to improve the performance. In this chapter, we firstly describe the original SSF algorithm which has been used in the contact measurement such as the ECG and validate that it can be used to improve the remote PRV measurement. Then we describe the proposed Adaptive SSF method to furtherly improve the performance. After this, we elaborate the novel Adaptive Two-window Peak Detection.

5.2/ SLOPE SUM FUNCTION

5.2.1/ ALGORITHM

Most of the BVP signals are periodic or quasi-periodic signals where rising phases and falling phases appear alternately and sequentially. To reduce the noise of the BVP signal which may affect the detection of systolic peaks, it is possible to consider enhancing the rising trend of the signal and reducing the downward trend of it, so that the signal may be more clear for the peak detection. Based on this idea, Zong et al. proposed to use the SSF [Zong et al., 2003] to detect the onset of arterial blood pressure pulses with contact equipment, and the function is expressed as:

$$S_{new}(i) = \sum_{i-w}^i \Delta S_i \quad \text{and} \quad i = w + 1, w + 2, \dots, N \quad (5.1)$$

and ΔS_i is expressed as:

$$\Delta S_i = \begin{cases} S(i) - S(i-1) & \text{if } S(i) - S(i-1) > 0 \\ 0 & \text{if } S(i) - S(i-1) \leq 0. \end{cases} \quad (5.2)$$

Equations 5.1 and 5.2 show the calculation of the new signal S_{new} transformed from the original signal, where S is the original signal, i represents the time index of the signal

and w is the window size. To maximize the effect of SSF, the window size w should be approximately the same with the length of the rising phase of the original signal.

Fig. 5.1 shows an example of the BVP signal transformed by SSF. The black signal is the BVP signal generated from the MMSE dataset. The blue signal is the new signal after transformation using SSF. It can be seen that after using the SSF, the upslope part is enhanced and it becomes sharper and more straightforward for peak detection.

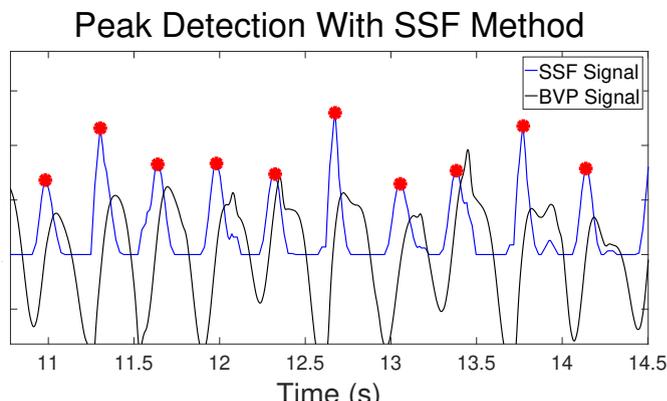


FIGURE 5.1 – An example of the Slope Sum Function (SSF) method. The original BVP signal generated from the video is black and the SSF signal is blue.

5.2.2/ EXPERIMENTS

As discussed, the SSF has been successfully implemented for peak detection for contact signal such as ECG, but it has not been validated on the rPPG signal. Therefore, it is necessary to test if this algorithm can be applied on the rPPG signals to improve the peak detection with the two datasets. The ground truth of the BVP signal and peak locations were given by the contact sensor, so the precision of the peak detection and PRV measurement can be assessed by the errors between the rPPG signals and the ground truth. We proposed to use some metrics to evaluate the performance of both methods. The evaluation metrics of the experiment are shown in Table 5.1, which are classified as three groups of metrics:

1) **Peak Detection Errors.** This group of metrics is used to evaluate the accuracy of peak detection.

- Proportion of correctly/incorrectly detected peaks and missing peaks (**%CP**, **%IP** and **%MP**). Since the ground truth is provided by the contact sensor measured from fingers and the rPPG was measured from faces, there is a time difference between the peaks on the rPPG signal and the contact PPG signal called pulse transit time [Smith et al., 1999] which is possibly caused by the different distance from the heart and the recording sensor. As a result, the search range for the correctly detected peaks was empirically set to 0.2 s. If there is more than one peak in the search range, then the extra peaks are considered as incorrectly detected peaks. If there is no peak, then it is considered as a missing peak. With these conditions,

%CP is calculated as the number of correctly detected peaks over the number of peaks of ground truth. **%IP** and **%MP** are calculated in the same way.

2) **PRV Errors**. This group of metrics is used to evaluate the accuracy of PRV measurement.

- PRV Errors (**PRV_{er}** in seconds) and Inter-beat Interval Errors (**IBI_{er}** in seconds). According to [Malik, 1996], the PRV signal can be represented in two different ways. It is either calculated as a peak interval series versus number of progressive peaks, or a peak interval series versus time, which is obtained as a signal of a function of time by interpolating the discrete event series (DES). In our case, PRV is the peak interval series over time. IBI is the peak interval series versus number of progressive peaks. Both **PRV_{er}** and **IBI_{er}** are calculated as the absolute average difference between the remotely measured PRV signal and ground truth contact PRV signal.
- Relative PRV Errors (**%PRV_{er}**). It is calculated as the average value of **PRV_{er}** over the PRV of the ground truth in each time stamp.

3) **PRV Feature Errors**. This group of metrics is used to evaluate the accuracy of PRV features.

- Errors of Standard Deviation of IBI Series (**STD_{er}** in seconds). This is calculated as the absolute difference between the Standard Deviation (STD) of rPPG measured IBI and the STD of IBI measured by ground truth contact PPG signal.
- Errors of Root Mean Square of Successive Inter-Beat Interval Differences (RMSSD) (**RMSSD_{er}** in seconds). As before, this metric is calculated as the absolute difference between the RMSSD measured by rPPG and the RMSSD measured by the ground truth. The RMSSD was defined as:

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (IBI_{(i+1)} - IBI_i)^2} \quad (5.3)$$

where IBI_i is the i^{th} peak interval value.

The experimental results were the average values of the entire testing dataset.

As mentioned, the window size w used in the equation 5.1 should be approximately the same with the length of the rising phase of the original signal in order to effectively transform the signal with maximum upward trend. In the case of UBFC-RPPG dataset, the window size was selected as the time of 9 samples for the frame rate of 30 fps. For MMSE dataset, the window size was selected as the time of 10 samples for the frame rate of 25 fps. Both parameters were optimized with brute force search by metrics **PRV_{er}**, **IBI_{er}** and **%CP**. We compare this method with the Local Maximum method which can be easily implemented by MATLAB's function "findpeaks". The parameters "MinPeakDistance" and "MinPeakProminence" of the function were optimized using brute force search with **PRV_{er}**, **IBI_{er}** and **%CP** as well.

5.2.3/ RESULTS

Results of MMSE dataset. Tables 5.2, 5.3, and 5.4 are the results of the MMSE dataset with the SSF method and Local Maximum method. First of all, it can be seen that the SSF

TABLE 5.1 – Evaluation Metrics.

Category	Metrics	Denotation	Unit
Peak Detection Errors	Proportion of correctly detected peaks	%CP	Percentage (%)
	Proportion of incorrectly detected peaks	%IP	Percentage (%)
	Proportion of missing peaks	%MP	Percentage (%)
PRV Errors	Inter-beat interval errors	IBI_{er}	Seconds (s)
	PRV errors	PRV_{er}	Seconds (s)
	Relative PRV errors	%PRV_{er}	Percentage (%)
PRV Feature Errors	Errors of standard deviation of IBI signal	STD_{er}	Seconds (s)
	Errors of root mean square of successive inter-beat interval differences	RMSSD_{er}	Seconds (s)

TABLE 5.2 – The average peak detection errors with MMSE dataset.

Methods	%CP	%MP	%IP
<i>Local Maximum</i>	87.8%	3.8%	8.4%
<i>SSF</i>	90.5%	4.0%	5.5%

method gives higher proportion of correctly detected peaks than the conventional Local Maximum method and this means the SSF is an effective method to transform the original signal and detect the peaks. Although the SSF changes the shape of the original signal as shown by Fig. 5.1, the PRV measurement should not be affected if this shift is closer to a constant, as the PRV is the time differences of the peaks.

TABLE 5.3 – The average PRV errors with MMSE dataset.

Methods	IBI_{er}	PRV_{er}	%PRV_{er}
<i>Local Maximum</i>	0.17	0.16	21.7%
<i>SSF</i>	0.15	0.14	21.5%

TABLE 5.4 – The average errors of PRV features with MMSE dataset.

Methods	STD_{er}	RMSSD_{er}
<i>Local Maximum</i>	0.09	0.11
<i>SSF</i>	0.08	0.07

Table 5.3 shows that both the IBI errors and PRV errors obtained by SSF method is smaller

than the Local Maximum method. This means, the SSF method indeed improves the PRV measurement, possibly due to more correctly detected peaks. The shift between the SSF signal and original signal has no significant influence on the PRV measurement with SSF.

Table 5.4 indicates that the PRV features obtained by SSF is better measured than Local Maximum method. This means SSF could improve the performance when these features are adopted in practical applications such as emotion detection.

Fig. 5.2 shows the peak detection with Local Maximum and SSF respectively. It can be seen that both methods work well on this video, however, the Local Maximum method gives an incorrectly detected peak on the remotely measured BVP signal around 22 seconds. While the SSF successfully avoided the error with the transformed signal. Fig. 5.3 shows the clear images of both signals around 22 seconds. It is worth mentioning that the parameters of the Local Maximum method is already fully optimized with the brute force search. This shows the disadvantages of all the rule-based on Local Maximum method such as the one presented by [Fukunishi et al., 2018]. No matter how well the parameters are set, it always fails in some specific cases.

Results of UBFC-RPPG dataset. Similarly, Tables 5.5, 5.6, and 5.7 show the results of the two methods on UBFC-RPPG dataset. Since the head movements of the participants in UBFC-RPPG dataset is much less than those in MMSE dataset, the signal is less noisy and the peak detection is easier for this dataset. We still observe the similar trend of the experimental results, which show the SSF method perform better than the conventional Local Maximum method. For instance, Table 5.5 shows that the SSF gives more correctly detected peaks. Table 5.6 shows that the PRV measured by the SSF method is more precise than Local Maximum method. Table 5.7 indicates that the PRV features is better obtained by SSF than the Local Maximum.

We found examples where rule-based Local Maximum method failed again while the SSF successfully avoided it. Fig. 5.4 shows one of the cases. In this example, the rule-based Local Maximum gave one false peak detection around 5 to 5.5 seconds. The SSF avoided the error because the SSF signal is much more clear in general, and the rules fail with lower probability.

TABLE 5.5 – The average peak detection errors with UBFC-RPPG dataset.

Methods	%CP	%MP	%IP
<i>Local Maximum</i>	94.2%	2.7%	3.1%
<i>SSF</i>	96.4%	1.1%	2.5%

TABLE 5.6 – The average PRV errors with UBFC-RPPG dataset.

Methods	IBI _{er} (s)	PRV _{er} (s)	%PRV _{er}
<i>Local Maximum</i>	0.07	0.06	12.7%
<i>SSF</i>	0.05	0.05	10.7%

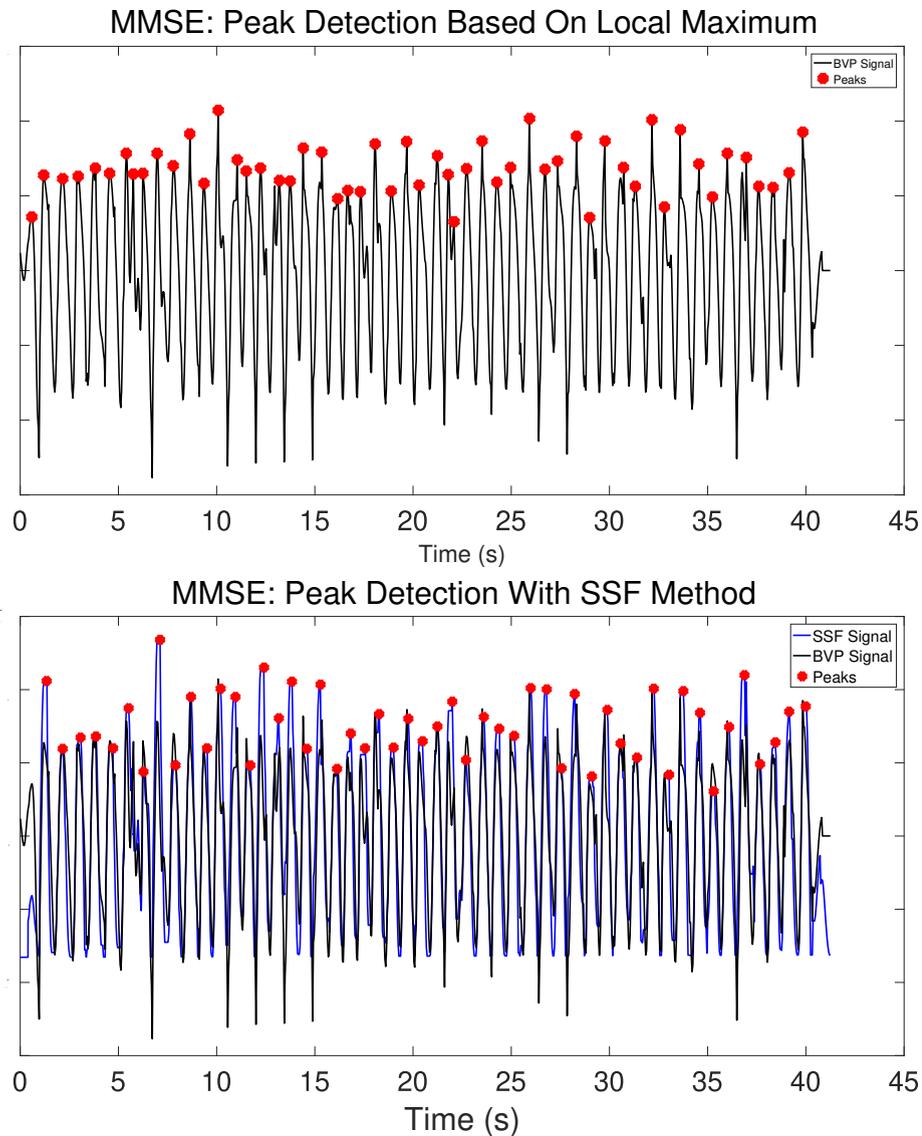


FIGURE 5.2 – The comparison of Local Maximum and SSF methods on the MMSE dataset. The black signal is the remotely measured BVP signal and the blue signal is the SSF signal.

TABLE 5.7 – The average errors of PRV features UBFC-RPPG dataset.

Methods	$STD_{er}(s)$	$RMSSD_{er}(s)$
<i>Local Maximum</i>	0.03	0.09
<i>SSF</i>	0.02	0.04

Conclusion. According to the results generated from both dataset, it is clear that the SSF, which was used for contact signals, is effective for peak detection on rPPG signal, although the experimental conditions of contact and remote methods are completely different. It improves the performance on both dataset and gives better results than the conventional Local Maximum peak detection method. However, the window size w was given manually using brute force search, therefore it is not practical in realistic applica-

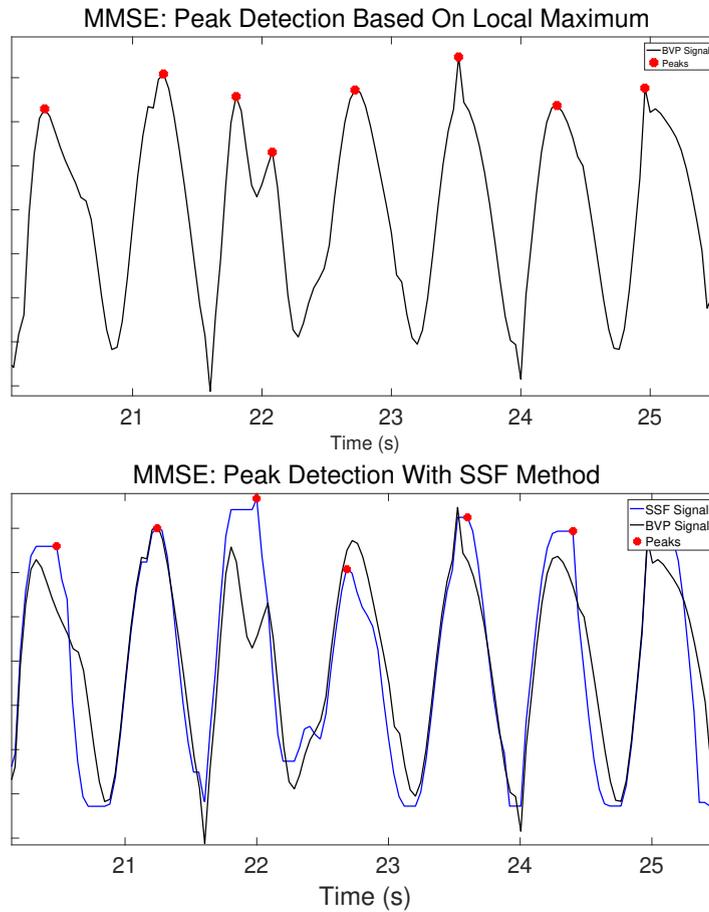


FIGURE 5.3 – The comparison of Local Maximum method and SSF method on the MMSE dataset. The black signal is the remotely measured BVP signal and the blue signal is the SSF signal.

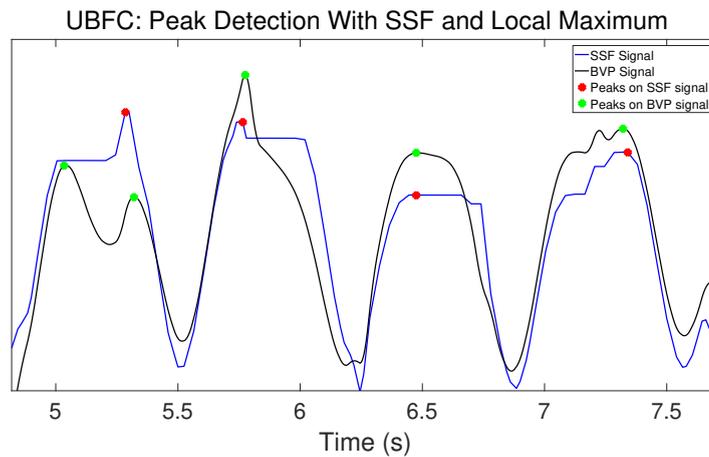


FIGURE 5.4 – The comparison of the Local Maximum method and SSF method on the UBFC-RPPG dataset. The remotely measured BVP signal is black and the SSF signal is blue. The green dots are the peaks detected on the original BVP signal with Local Maximum method and the red dots are the peaks detected with the SSF method.

tions. In the next section, we present Adaptive SSF to address this issue.

5.3/ ADAPTIVE SLOPE SUM FUNCTION

5.3.1/ ALGORITHM

While we have validated that the SSF works better than conventional Local Maximum method in remote PRV measurement on rPPG signal as shown in the previous section, this method has two serious disadvantages. First of all, the window size w is set manually. To maximize the effect of SSF, w should be approximately the same with the length of the rising part. Practically, the window size can be set empirically or using brute force search. For instance, Zong et al. [Zong et al., 2003], the original authors of the SSF method, empirically set w as 128 ms. In our case, the parameter was obtained using brute force search. This restricts the practical applications of the algorithm, because in real-life scenarios the parameters must be given automatically. Secondly, there is a shift between the peaks on SSF signal and original signal. This does not significantly affect the PRV measurement if the shift is a constant, however, it indeed increases the peak location errors as shown in Fig. 5.1. Actually, this issue is related to the window size w as well. When the w is slightly larger than the actual length of rising part, horizontal straight lines may appear in the peak regions of the SSF signal. In this case, the peak detection rule has to be set so that the leftmost points of the horizontal straight lines are considered as the real peaks. If the original signal is noisy and there are more than one local peaks in a certain period, the corresponding peak on the SSF may appear later than the actual peak location. When the w is smaller than the actual length of rising part, the peak locations of SSF signal possibly appear earlier than the actual peak locations if the slopes around the original signal's peaks are smaller than the previous signal parts. As the window size is set manually as a fixed value, it cannot conform with the actual length of rising part. Fig. 5.5 shows an example where the w is too small. It can be seen that the peaks appear earlier when the window size is too small. Fig. 5.6 shows an example where the w is too large. The SSF signal around 13.5 to 14 second and around 17 second are horizontal straight lines, and in both parts, the peaks appear later than the actual peak locations.

To address this issue, we propose to set the w adaptively and automatically. Firstly, we apply FFT on the BVP signal over a 10 seconds' window and get the frequency $F(i)$ for each point in time i . Then the period of each point $P(i)$ is adaptively calculated as:

$$P(i) = \frac{1}{F(i)} \quad (5.4)$$

Let $N(i)$ be the number of the points of the rising part within the period,

$$N(i) = \sum_{i-P(i)/2}^{i+P(i)/2} n_i \quad (5.5)$$

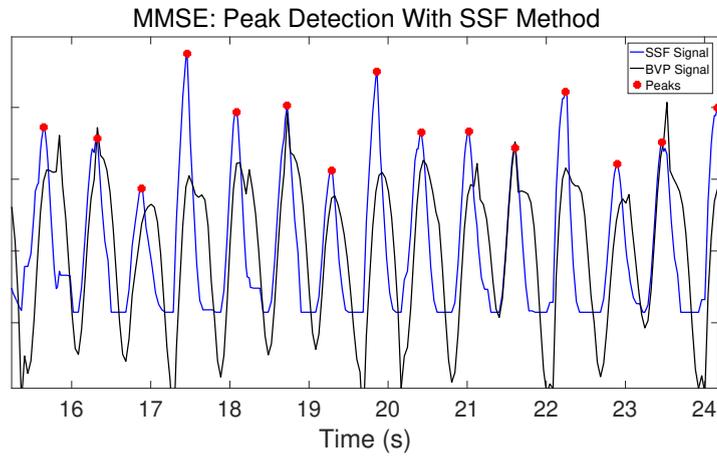


FIGURE 5.5 – An example where the window size is smaller than the length of rising part. The remotely measured BVP signal is black and the SSF signal is blue.

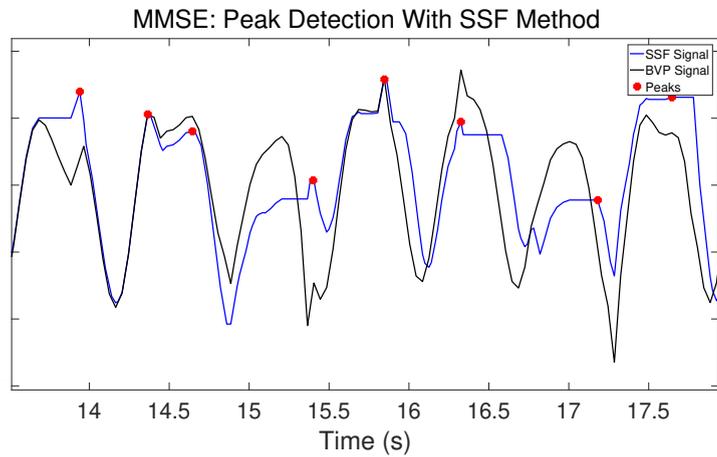


FIGURE 5.6 – An example where the window size is larger than the length of rising part. The remotely measured BVP signal is black and the SSF signal is blue.

and n_i is expressed as:

$$n_i = \begin{cases} 1 & \text{if } S(i) - S(i-1) \geq 0 \\ 0 & \text{if } S(i) - S(i-1) < 0. \end{cases} \quad (5.6)$$

w in time i is finally calculated as:

$$w(i) = N(i) \quad (5.7)$$

The w is the time length of the window, and practically it is the number of the points as the signal is discrete. After the w is adaptively obtained, equation 5.1 is adopted to get the SSF signal.

5.3.2/ EXPERIMENTS AND RESULTS

To evaluate the performance of Adaptive SSF, we used exactly the same metrics which are shown in section 5.2.

Results of MMSE dataset. Tables 5.8, 5.9 and 5.10 are the results of Adaptive SSF and SSF methods. It is clear that the new method indeed improves the performance with the adaptively set window size. The Adaptive SSF gives more correctly detected peaks, and fewer incorrectly detected peaks and missing peaks. It gives more precise PRV measurement and PRV features. This is possibly because the error caused by the peak location shift with imprecise w is reduced. In the experiments, we focus on the comparison of Adaptive SSF and original SSF and do not show the results of Local Maximum method, but it is worth mentioning that both methods performed better than the Local Maximum in all the metrics.

Fig. 5.7 shows the same signal of the same video of Fig. 5.2. It can be seen that the Adaptive SSF method avoided the incorrect peak detection error on 22 seconds just like the original SSF.

Fig. 5.8 shows two examples of SSF and Adaptive SSF respectively on the same signal. It can be seen that both methods work effectively in this case, however, the Adaptive SSF give more precise locations for the three peaks that are located 12 to 12.5 second, 13.5 to 14 second, and 14.5 second. This is because the w set by brute force search is slightly larger than the length of the rising part in these periods, while the adaptively set w is more precise than that of the original SSF method.

TABLE 5.8 – The average peak detection errors with MMSE dataset.

Methods	%CP	%MP	%IP
<i>SSF</i>	90.5%	4.0%	5.5%
<i>Adaptive SSF</i>	91.7%	3.4%	4.9%

TABLE 5.9 – The average PRV errors with MMSE dataset.

Methods	IBI _{er}	PRV _{er}	%PRV _{er}
<i>SSF</i>	0.15	0.14	21.5%
<i>Adaptive SSF</i>	0.14	0.13	18.3%

TABLE 5.10 – The average errors of PRV features with MMSE dataset.

Methods	STD _{er}	RMSSD _{er}
<i>SSF</i>	0.08	0.07
<i>Adaptive SSF</i>	0.06	0.06

Results of UBFC-RPPG dataset. The results of UBFC-RPPG show similar trend of the two methods, as shown in Tables 5.11, 5.12 and 5.13. The Adaptive SSF performs better in all the metrics on this dataset. Again, it is worth mentioning that both methods give better results than the Local Maximum, like those of MMSE dataset.

Similarly, Fig. 5.9 shows that the Adaptive SSF method gets more precise peak locations than the original SSF, because in this case, the manually set w is smaller than the actual length of the rising part of the signal, and as a result, the peak locations on SSF signal appear earlier than the peak locations on the original signal.

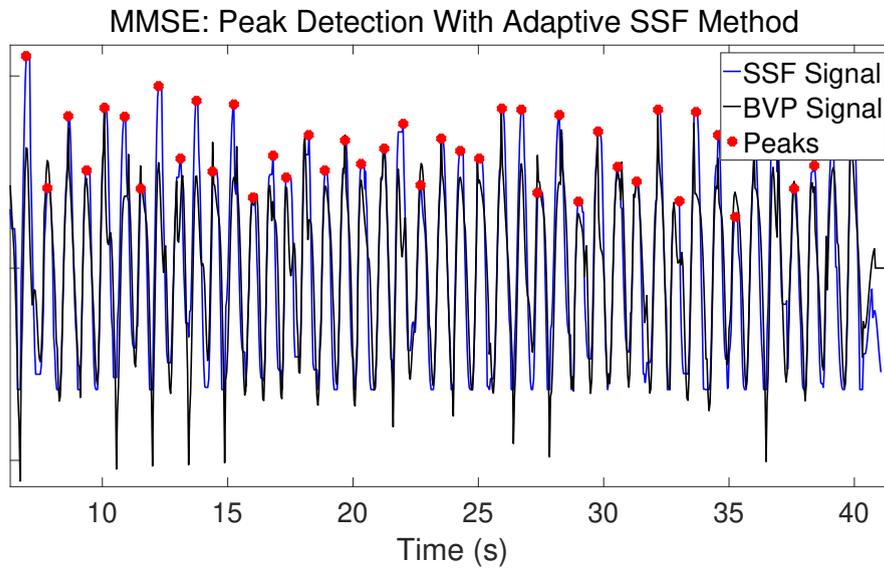


FIGURE 5.7 – An example of the Adaptive SSF method on the MMSE dataset. The remotely measured BVP signal is black and the SSF signal is blue.

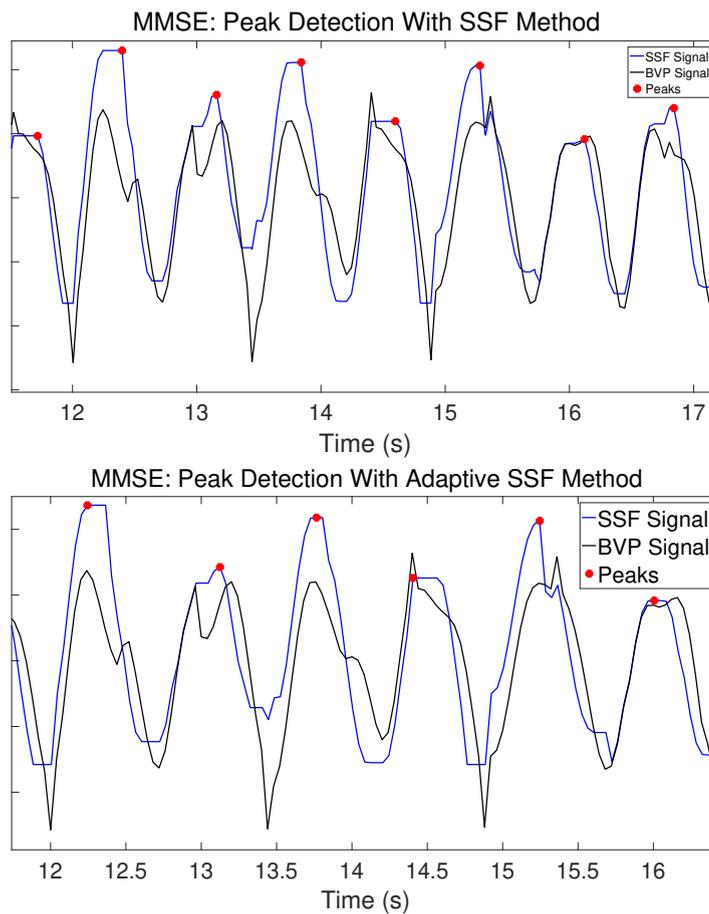


FIGURE 5.8 – The comparison of the Adaptive SSF method and SSF method on the MMSE dataset. The remotely measured BVP signal is black and the SSF signal is blue.

TABLE 5.11 – The average peak detection errors with UBFC-RPPG dataset.

Methods	%CP	%MP	%IP
<i>SSF</i>	96.4%	1.1%	2.5%
<i>Adaptive SSF</i>	96.8%	1.0%	2.2%

TABLE 5.12 – The average PRV errors with UBFC-RPPG dataset

Methods	IBI _{er}	PRV _{er}	%PRV _{er}
<i>SSF</i>	0.05	0.05	10.7%
<i>Adaptive SSF</i>	0.04	0.04	8.7%

TABLE 5.13 – The average errors of PRV features UBFC-RPPG dataset

Methods	STD _{er}	RMSSD _{er}
<i>SSF</i>	0.03	0.05
<i>Adaptive SSF</i>	0.02	0.04

Conclusion. We have shown that the Adaptive SSF performed better than the original SSF with the adaptively set parameter. However, there is still a shift between the original signal and SSF signal and this may affect the precision of PRV measurement. To address this issue, novel algorithms should be proposed to detect the peaks on the original signal without changing the shape.

5.4/ ADAPTIVE TWO-WINDOW PEAK DETECTION

5.4.1/ ALGORITHM

The Local Maximum and SSF-based peak detection methods are essentially one-window methods, which detect the peaks on the BVP signal with one moving window. The SSF-based methods give more precise PRV measurement with higher proportion of correctly detected peaks than the Local Maximum method by transforming the signal into a more clear new signal, but the peak locations are not precise. To improve the precision of BVP peak detection and overcome the disadvantages of SSF-based methods, it is possible to add more than one window and detect the peaks without changing the signal shape.

According to Elendi et al. [Elgendi et al., 2013], the BVP signal is considered to have two important time windows. The first one is the "beat window", which is the entire period of one heart beat. And the other is the "peak window", which is the period where a systolic peak appears. A systolic peak window is inside a beat window, therefore the time length of the peak window is smaller than the beat window. The two windows have two important physical properties: firstly, the average signal amplitude of the peak window is usually higher than the average amplitude of the beat window, and secondly, the systolic peaks are supposed to be the highest points within the peak window. With the definitions of the two windows and these properties, we propose an algorithm that uses two adaptively determined windows and this algorithm was named as "Adaptive Two-Window Peak

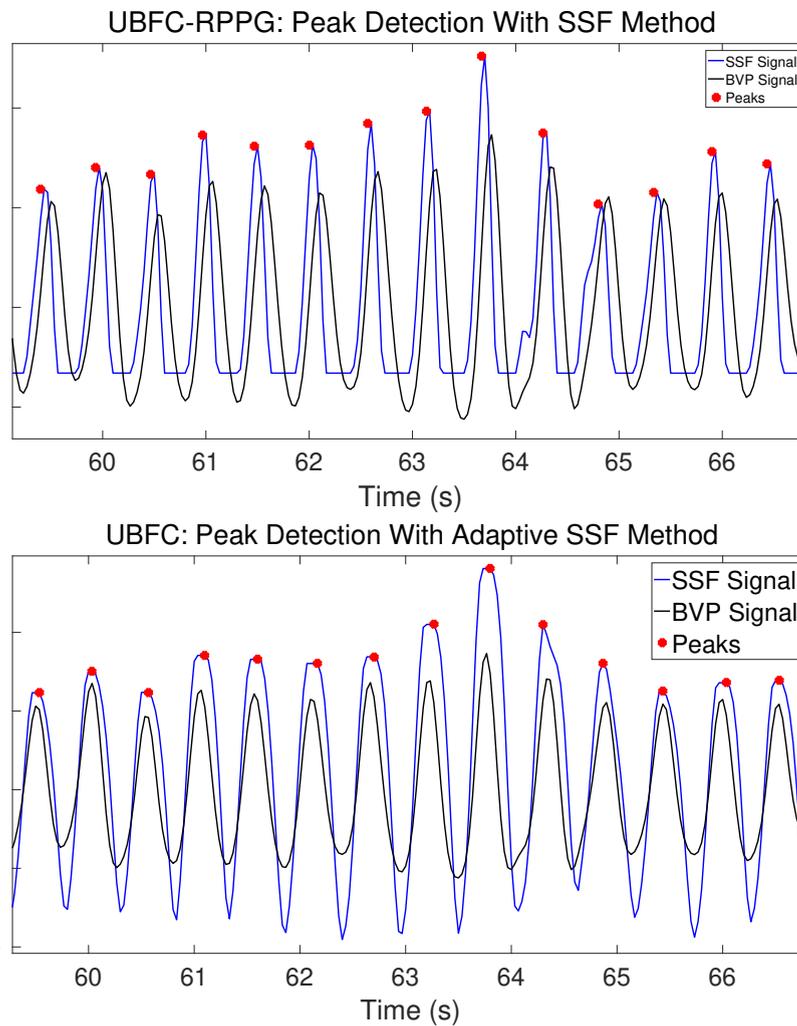


FIGURE 5.9 – The comparison of the Adaptive SSF method and SSF method on the UBFC-RPPG dataset. The remotely measured BVP signal is black and the SSF signal is blue.

Detection”.

The general idea of this algorithm is to define ”block of interest” in which local maxima will be detected. To do this, two moving averages are defined with two different time windows. The first moving average (MA), denoted as MA_b , is calculated with a window of size W_b representing the length of a heart beat. The second MA, denoted as MA_p , is calculated with a window of size W_p delimiting the region of the systolic peak. From the above mentioned properties, it is possible to define the blocks of interest as the periods when MA_p is greater than MA_b . The peaks of the BVP signal will be detected as the local maxima in these blocks of interest. Fig. 5.10 shows how the blocks of interest are defined. In the figure, the black signals is the remotely measured BVP signal and the green and blue signals are MA_b and MA_p respectively. The stripes are blocks of interest as MA_p is larger than MA_b in these areas.

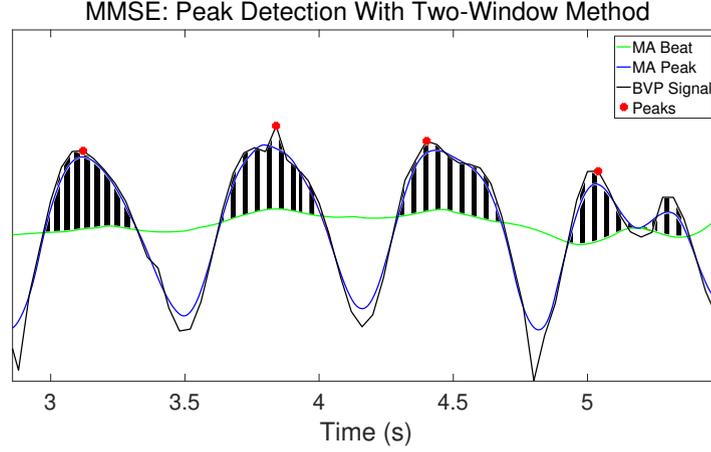


FIGURE 5.10 – An example of block of interest. The black signal is the remotely measured BVP signal. The green and blue signals are MA_b and MA_p respectively. The strips are the blocks of interest.

Based on this idea, the MA of the BVP signal y over the beat window W_b is calculated as:

$$MA_b(i) = \frac{1}{W_b} \left(S(i - W_b/2) + \dots + S(i) + \dots + S(i + W_b/2) \right) \quad (5.8)$$

where $S(i)$ is the BVP value at time i . Similarly, and the MA of the BVP signal over the peak window W_p is calculated as:

$$MA_p(i) = \frac{1}{W_p} \left(S(i - W_p/2) + \dots + S(i) + \dots + S(i + W_p/2) \right) \quad (5.9)$$

and two thresholds are defined as:

$$\tau_1(i) = MA_b(i) \quad (5.10)$$

$$\tau_2 = \overline{W}_p \quad (5.11)$$

where \overline{W}_p is the average window size of the peak windows within the block of interest:

$$\overline{W}_p = (W_p(i_1) + W_p(i_2) + W_p(i_3) + \dots + W_p(i_n))/n \quad (5.12)$$

in which $i_1, i_2, i_3, \dots, i_n$ are the time points of the block interest.

The systolic peaks can be detected with such conditions:

- MA_p signal is compared with τ_1 . Contiguous time periods where MA_p is greater than τ_1 are noted as possible blocks of interest.
- The block of interest is discarded, if the width of the block is smaller than τ_2 .
- The peaks are the maximum values in the blocks of interest.

Since the BVP signal has been very well filtered by the ROI and channel selection methods, the only parameters that have to be determined are the two window sizes. We ap-

plied FFT on the BVP signal over a 10 seconds' window and get the frequency $F_b(i)$ for each point in time, and then the $W_b(i)$ is calculated as:

$$W_b(i) = \frac{1}{F_b(i)}. \quad (5.13)$$

For the detrended BVP signal, the peaks usually appear in the positive part that is approximately half of the signal. The peak window should be within the beat window, therefore we can set the peak window as half of the positive part of the beat window which is 0.25 times of the beat window so the value of $W_p(i)$ can be given as $0.25 \times W_b(i)$. With these rules and parameters, the peaks can be detected.

Fig. 5.11 shows an example of the peak detection of the BVP signal with the Two-Window method. The black signal is the original BVP signal generated from the MMSE dataset. The blue signal is the MA_p . The green signal is the threshold MA_b . It can be seen that in this case, the local peaks around 3.9 second and 4.2 second are eliminated because the MA_p is not higher than the MA_b . So these peaks are considered noise.

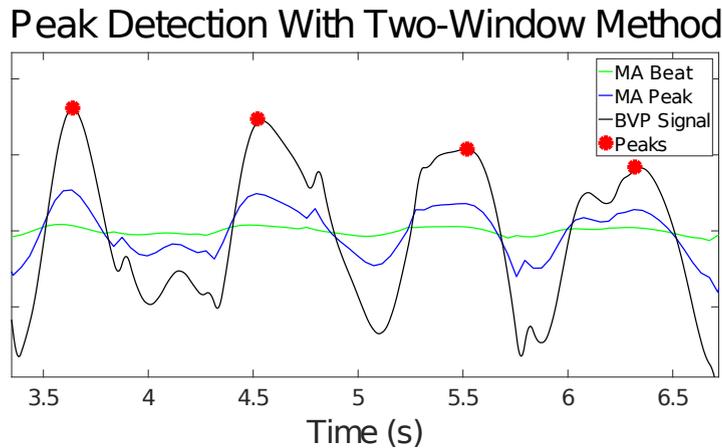


FIGURE 5.11 – An example of peak detection with Two-Window method. The black signal is the remotely measured BVP signal. The green and blue signals are MA_b and MA_p respectively.

5.4.2/ EXPERIMENTS AND RESULTS

To test the algorithm, we compare this method with the Adaptive SSF method and Local Maximum method like what we have done in sections 5.2 and 5.3. Besides the metrics shown in Table 5.1, we add another metric, namely Peak Location Errors (**PLE** in seconds). It is calculated as the average absolute difference between the peak locations detected on the rPPG signal and the annotated peaks of the ground truth.

Results of MMSE dataset. Table 5.14 shows the errors of the peak detection from the three methods with MMSE dataset. The first column is average peak location errors

(PLE). It can be seen that the Two-Window method performed better than Local Maximum with smaller error. The Adaptive SSF has no result in this column because there is a shift between the new signal transformed by SSF and the original signal due to shape change, as we have discussed in sections 5.2 and 5.3. Therefore, it is not fair to compare Adaptive SSF with other methods for location errors. The second, third and fourth columns are the proportion of correctly detected peaks, missing peaks and incorrectly detected peaks respectively. These columns show that the Two-Window method gives more peaks correctly detected than the other two methods.

Fig. 5.12 shows a typical example where the Local Maximum failed but SSF and Two-Window methods were effective. In this specific case, the Local Maximum detected a wrong peak between 14th and 15th seconds and this shows the weakness of this method. On the other hand, the rules given by the Two-Window method have precisely found the block of interest and then detected the maximum value within this block of interest, so it avoided giving a wrong peak.

TABLE 5.14 – The average peak detection errors with MMSE dataset.

Methods	PLE	%CP	%MP	%IP
<i>Local Maximum</i>	0.14	87.8%	3.8%	8.4%
<i>Adaptive SSF</i>	X	91.7%	3.4%	4.9%
<i>Two-Window</i>	0.12	94.0%	2.0%	4.0%

Table 5.15 shows the results of PRV measurement errors. The IBI is the peak intervals versus number of progressive peaks and the PRV is the peak intervals versus time. According to the table, the Two-Window method generated better results for all the three PRV metrics than the other two methods with smaller errors. The table shows that the results of the PRV obtained by the Adaptive SSF method are worse than the Two-Window method although it performed better than the Local Maximum, and this is because the shift caused by SSF transformation is not constant and the location errors are not perfectly reduced for the PRV measurement. As can be seen from the image, the first peak and the fifth peak detected on the SSF signal are not at the actual locations of the original BVP signal. The distance between the first and second peak on the SSF signal and the fourth and fifth peak on the original BVP signal is smaller than the distance between the corresponding peaks on the original signal. As a result, the IBI calculated by these peaks with the Adaptive SSF method is smaller than the real value. On the other hand, the peak detection within the same period performed by Adaptive Two-Window method does not have this problem since it does not change the shape of the signal.

TABLE 5.15 – The average PRV errors with MMSE dataset.

Methods	IBI _{er}	PRV _{er}	%PRV _{er}
<i>Local Maximum</i>	0.17	0.16	21.7%
<i>Adaptive SSF</i>	0.14	0.13	18.3%
<i>Two-Window</i>	0.14	0.12	17.0%

We calculated the errors of the two time domain features which are shown in Table 5.16. Clearly, the Adaptive Two-Window performed the best in both metrics with smaller errors.

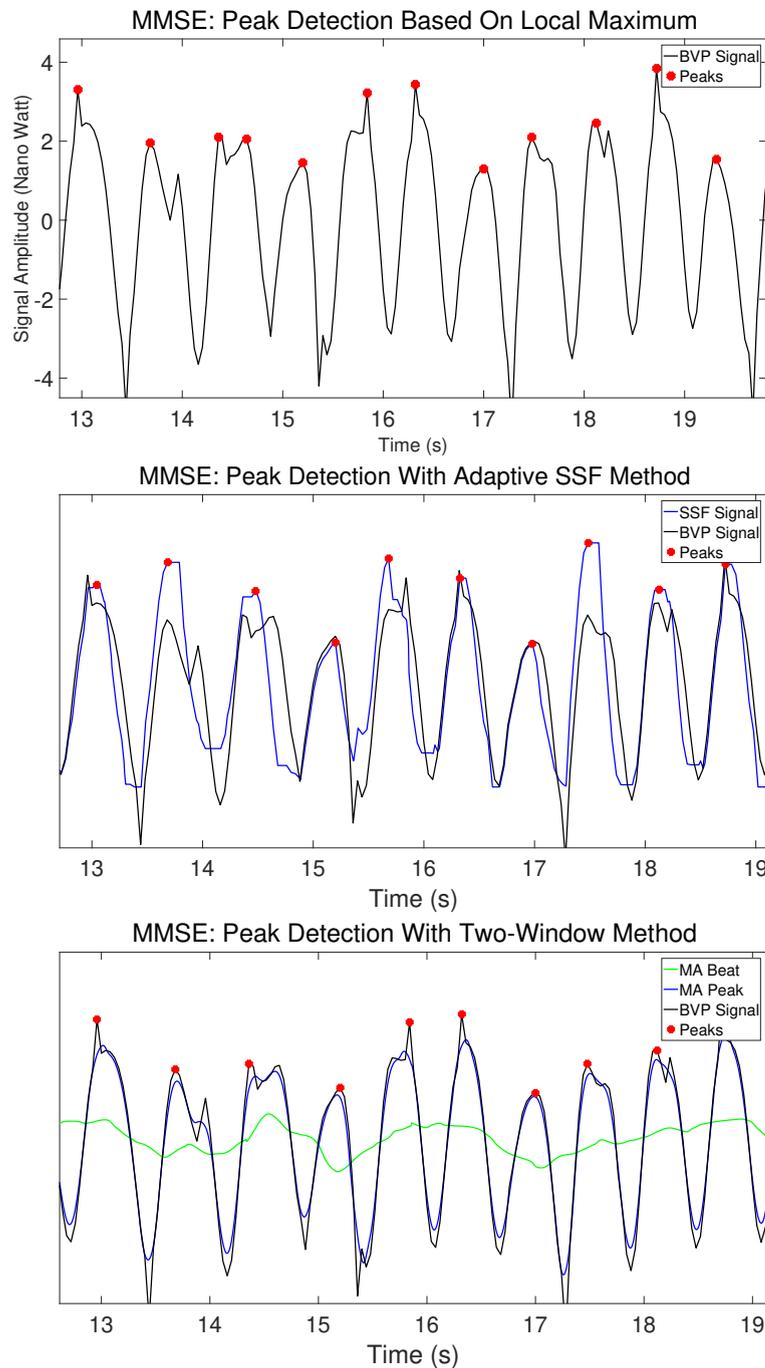


FIGURE 5.12 – The examples of peak detection with Local Maximum, Adaptive SSF and Two-Window methods with a video from MMSE dataset. The black signal is the remotely measured BVP signal.

Results of UBFC-RPPG dataset. Similarly, we calculated the metrics for UBFC-RPPG dataset, and the results are shown in Tables 5.17, 5.18 and 5.19. As expected, the errors for this dataset is smaller than those of MMSE dataset because the signals of this dataset are more clear and less noisy. And for all the metrics, the Two-window method performs better than Local Maximum and Adaptive SSF methods. As shown in Fig. 5.13,

TABLE 5.16 – The average errors of PRV features with MMSE dataset.

Methods	STD _{er}	RMSSD _{er}
<i>Local Maximum</i>	0.09	0.11
<i>Adaptive SSF</i>	0.06	0.06
<i>Two-Window</i>	0.05	0.06

the peak locations on SSF signal around 62 second is not precise, so the IBI calculated between the peak and previous peak is smaller than the actual value, and the IBI calculated between the peak and the next peak is larger than the actual value.

TABLE 5.17 – The average peak detection errors with UBFC-RPPG dataset.

Methods	PLE(s)	%CP	%MP	%IP
<i>Local Maximum</i>	0.04	94.2%	2.7%	3.1%
<i>Adaptive SSF</i>	X	96.8%	1.0%	2.2%
<i>Two-Window</i>	0.01	97.0%	2.0%	1.0%

TABLE 5.18 – The average PRV errors with UBFC-RPPG dataset.

Methods	IBI _{er}	PRV _{er}	%PRV _{er}
<i>Local Maximum</i>	0.07	0.06	12.7%
<i>Adaptive SSF</i>	0.05	0.04	8.7%
<i>Two-Window</i>	0.02	0.01	4.0%

TABLE 5.19 – The average errors of PRV features with UBFC-RPPG dataset.

Methods	STD _{er}	RMSSD _{er}
<i>Local Maximum</i>	0.04	0.09
<i>Adaptive SSF</i>	0.02	0.04
<i>Two-Window</i>	0.01	0.01

5.5/ CONCLUSION

After the BVP signal is obtained with ROI segmentation and color channel signal selection, the next step is to get the precise peak locations on the BVP signal and calculate the PRV values. Unlike the HR, the PRV is much more difficult to measure because the low frame rate, sensor noise, head movement and other noise make the peak detection a very challenging task. Unfortunately, the improvement of peak detection on rPPG signals is overlooked by many researchers, possibly because they can usually get the ideal experimental results by manually and empirically setting the parameters of the conventional Local Maximum methods, such as the works presented in [Huang et al., 2016][McDuff et al., 2014a][Fukunishi et al., 2018]. However, this restricts

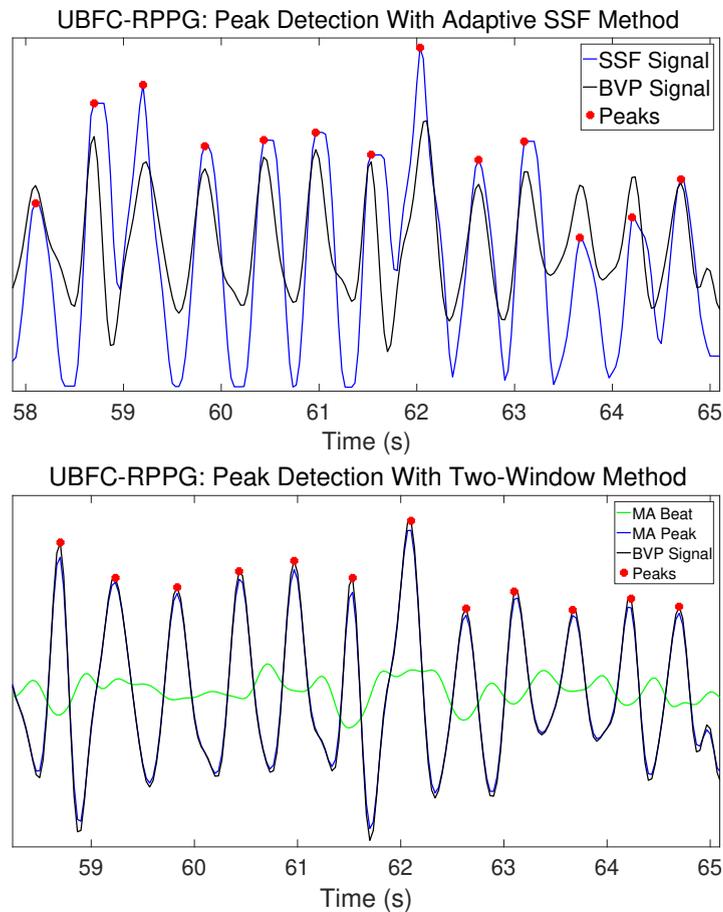


FIGURE 5.13 – The examples of peak detection with Adaptive SSF method and Two-Window method with UBFC-RPPG dataset. The black signal is the remotely measured BVP signal.

the practical applications because it is not convenient to manually set the parameters in commercial applications and the signals in real-life scenarios are usually much more noisy than experimentally obtained signals. Some algorithms [Zong et al., 2003] have been proposed to improve the peak detection of contact measured signals, but they were not used in rPPG research, possibly due to completely different experimental conditions. Another issue is that the precision of remote PRV measurement is not discussed in details in the existing research works.

Therefore, we have focused on improvement of the BVP peak detection and PRV measurement in this chapter. Firstly, we adopted the SSF method [Zong et al., 2003] to improve the BVP peak detection. This method was proposed to detect the onset of the arterial blood pressure pulses with contact medical equipment by enhancing the rising trend of the signal and reducing the decreasing trend. We show that this method can be used to improve the rPPG peak detection in comparison with the conventional Local Maximum method, although the shapes and properties of contact measured signals and rPPG signals are completely different. However, one significant disadvantage of this method is that the window size of the accumulation is manually set, which restricts its utilization in real-life scenarios as the signals in practical applications are usually much more noisy than those

obtained in the lab. In addition, the manually set parameter is a fixed number, thus it may not be precise for all the videos. When the window size is slightly larger than the actual length of the rising part of the signal, the peaks may appear later than the actual locations, and when the window size is slightly smaller, the peaks may appear earlier. To address this issue, we proposed to adaptively and automatically set the window size using FFT. We show that the proposed Adaptive SSF method performs better than the original method in both datasets in all the metrics.

The Local Maximum and SSF-based peak detection methods are essentially one-window methods, which detect the peaks on the BVP signal with one moving window. The SSF-based methods give more precise PRV measurement with higher proportion of correctly detected peaks than the Local Maximum method by transforming the signal into a more clear new signal, but the peak locations are not precise. To improve the precision of BVP peak detection and overcome the disadvantages of SSF-based methods, it is possible to adopt more than one window and detect the peaks without changing the signal shape. Therefore we proposed the Adaptive Two-Window Peak Detection algorithm for BVP peak detection and PRV measurement. This novel method utilizes the properties of the beat windows and peak windows of BVP signals to set the detection rules. The properties are: 1) the average signal amplitude of the peak window is usually higher than the average amplitude of the beat period, 2) the peaks are supposed to be the highest points within the peak windows. We adaptively calculated the sizes of the two windows and detected the peaks using the properties. The experiments showed that this method performed better than the Adaptive SSF and Local Maximum methods.

Our contributions improve the accuracy of the remote PRV measurement, and pave the way for further utilization of the method in commercial applications.

III

CONCLUSION

CONCLUSION AND FUTURE WORK

Biomedical researchers and doctors have put much effort into the cardiac parameters measurement, as these parameters are widely used in health monitoring, disease diagnosis and emotion detection. These physiological parameters are usually measured by inconvenient contact equipment such as ECG, thus the availability and wider application is restricted. Recently, computer vision researchers join this field, because it has been proved that the cardiac parameters can be extracted using low-cost cameras and signal processing algorithms. This has improved the convenience, reduced the cost and offered the potential intelligent and automatic solution for health industry. For instance, the camera-based cardiac measurement technique is possibly much more effective and efficient than conventional medical equipment in the long-term, non-invasive and non-contact health monitoring. However, the computer vision based methods cannot replace the medical equipment in critical medical analysis for now, because the performance suffers from the sensor noise and image noise. Therefore, the method should be intensively studied and improved.

The ANS is the portion of the nervous system that controls the body's visceral functions, including action of the heart, movement of the gastrointestinal tract and secretion by different glands, among many other vital activities. The ANS consists of two branches, the sympathetic and parasympathetic nervous system. The sympathetic nervous system activates the fight or flight response during a threat or perceived danger, and the parasympathetic nervous system controls the state of calm. Therefore, the mental and emotional states directly affect the ANS. It has been shown that the cardiac parameters such as HR and HRV can reflect the activities of the ANS branches. Moreover, PRV measured by PPG can replace the HRV in some conditions. Therefore the remote measurement of HR and PRV with computer technologies has great potential in applications such as remote health conditions monitoring, remote emotion detection, etc. In this thesis, we have focused on the improvement of HR and PRV measurement with rPPG technology.

By far, the basic framework of rPPG has been studied by many researchers. It consists of three main steps, namely the ROI segmentation, RGB channel selection and combination, and physiological parameters extraction. Many state-of-the-art methods have been proposed to improve the performance of these steps, such as the skin and face detection for ROI segmentation, the BSS and color space analysis for RGB channel combination, etc. Some researchers used novel equipment such as thermal cameras, polarization cameras, 5-band cameras, etc. to improve the performance, however, these cameras are either too expensive or not widely used in daily life. We have found several disadvantages of the

existing methods: 1. Although the ROI detection and RGB channel combination methods have been intensively studied, the performance of different combination of the state-of-art methods (such as landmarks detection for ROI and Chrom for channel combination) have not been investigated. 2. In the ROI detection step, the detected pixels are averaged to generate the RGB signals. This assumes all the pixels contribute exactly the same weight to the rPPG signal. The assumption is problematic because the blood vessels are not distributed homogeneously on the face. 3. The remote BVP peak detection and the improvement of PRV measurement are usually overlooked by rPPG researchers, possibly because the signals obtained in the lab are not noisy, thus sometimes the simple Local Maximum peak detection is effective to get the peaks on the BVP signal. However, the parameters of Local Maximum peak detection are manually set, so it cannot be used in realistic scenarios, especially when the signal is noisy.

To address these issues, we have made several contributions. First of all, we selected three ROI detection methods and three channel combination methods and implemented all the nine combinations of the methods in the framework of physiological parameter measurement. The selected ROI segmentation methods are face detection [Viola et al., 2001][Lucas et al., 1981], Conaire's skin detection [Conaire et al., 2007] and landmarks detection [Kazemi et al., 2014]. The selected channel combination methods are Green channel [Lempe et al., 2013], BSS (PCA) [Wedekind et al., 2015], and Chrom [De Haan et al., 2013]. These methods were selected as the representatives of the state of the art, because they work effectively, fast and are widely adopted. To do the experiments, we used two datasets, namely UBFC-RPPG and MMSE datasets. The UBFC-RPPG dataset was made by our lab in relatively favourable conditions, i.e. few head movements and good lighting. The MMSE dataset is closer to realistic scenario with many more head movements. We calculated several metrics such as the mean absolute error, SNR, pearson correlation coefficient, etc. and found that the Conaire's skin detection combined with Chrom as the channel combination method worked the best for both datasets. This comparative study has laid solid foundation for the next step, as the Chrom will be used as the default channel combination method for our research.

Secondly, we proposed a model-based method to improve the ROI detection by giving weight to every pixel before averaging. This is based on the fact that the rPPG information is not distributed evenly on the face. To do this, we recorded ten videos with very high resolution. For each video, we generated a weight map based on the SNR value of each pixel. The SNR value is given by the spectral analysis of the rPPG signal of each pixel in the entire video sequence. With face alignment and resizing, the weight map are properly obtained, and the average of the ten weight maps is the model. We applied this model on the two datasets and compared the performance with the best state-of-the-art method, i.e. Conaire's skin detection. The results showed that the proposed method performed better than skin detection in UBFC-RPPG dataset as expected, and the performance was very close to the state of the art in MMSE dataset. This is because the face alignment and resizing relied on the facial landmarks, and the movements in the MMSE dataset negatively affected the accuracy of the landmarks' locations.

With the BVP signal obtained by model-based ROI detection and Chrom channel combination method, we continue our work on the remote PRV measurement. Unlike the HR, the PRV is much more difficult to measure, because it then requires to detect the peaks

on the noisy BVP signal. In the rPPG framework, this step suffers from low frame rate, sensor noise, head movement, light variation, etc. Unfortunately, the state-of-the-art methods all adopted the conventional Local Maximum method, of which the parameters are usually manually set and the performance could become much worse under realistic scenarios where there is much more noise. To address this issue, we proposed to use the SSF-based methods and Adaptive Two-window Peak Detection. The SSF had been used for the signal generated by contact equipment, and it has not been used in the rPPG, possibly due to the completely different experimental conditions. We firstly validated that the SSF worked more effectively than the Local Maximum for rPPG method, and then we proposed a new method which adaptively set the window size for the accumulation calculation of the SSF. This new method is called Adaptive SSF and we showed that it worked better than the original SSF. However, the SSF has the disadvantages of changing the signal shape and increasing the peak location errors. Therefore, we proposed a novel method which detected the peaks without changing the signal shape, namely Adaptive Two-Window Peak Detection. This novel method utilizes BVP signals' properties that the average signal amplitude of the peak window is usually higher than the average amplitude of the beat window, and the peaks are supposed to be the highest points within the peak windows. We adaptively calculated the sizes of these two windows and detected the peaks using the properties. The experiments showed that this method performed better than the Adaptive SSF and Local Maximum methods in both datasets.

We believe these contributions have improved the feasibility of the rPPG technology in practical applications.

We have several proposals for the future work: 1) It is possible to adopt other parts of the body such as neck, arms and hands as the ROI because people usually do not sit in front of the cameras in realistic scenarios. The methods that detect arms, hands and other parts of the body as the ROI should be studied and investigated. 2) Since the ROI detection algorithm proposed by us is based on a supervised model, it is important to generate a reliable SNR weight map. The model may be different in different environmental conditions. It may fail if the registration of the model is imprecise due to the low accuracy of the landmarks detection. The ROI may have different ways to be resized or reshaped to fit the model. For instance, an elliptical model can be generated with the contour provided by the landmarks detection. Another possibility is to combine the skin detection method with the proposed method to furtherly improve the performance of ROI detection. It may give better results as the ROI detection for realistic datasets with more movements, because the skin detection could reduce the errors caused by incorrect facial landmarks. 3) For the BVP peak detection, the detection rules can be set using machine learning methods. For instance, when the dataset is large enough, a portion of the dataset can be used as the training data to set the detection rules based on the ground truth, and the rules can be tested in the rest of the dataset. 4) We have improved the PRV feature precision but we have not tested these features in the framework of emotion detection. To check if the proposed algorithms improve emotion recognition or stress estimation, it is possible to design some experiments with the aid of psychology researchers. For instance, in the experiments, the participants would possibly be invited to watch relaxed music videos and stressed annoying music videos respectively, and the emotion difference should be detected with the rPPG method since the PRV features can reflect the ANS activities.

LIST OF OUR PUBLICATIONS

INTERNATIONAL CONFERENCES:

[a] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li and F. Yang. (2018). "Comparison of Region of Interest Segmentation Methods for Video-Based Heart Rate Measurements," 18th International Conference on Bioinformatics and Bioengineering (BIBE). IEEE.

[b] Y. Benezeth, P. Li, R. Macwan, K. Nakamura, R. Gomez and F. Yang. (2018). "Remote heart rate variability for emotional state monitoring," EMBS International Conference on Biomedical Health Informatics (BHI). IEEE.

[c] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li and F. Yang. (2019). "Model-based Region of Interest Segmentation for Remote Photoplethysmography," in Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP). SciTePress.

[d] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li and F. Yang. (2019). "An Improvement for Video-based Heart Rate Variability Measurement," 4th International Conference on Signal and Image Processing (ICSIP). IEEE.

INTERNATIONAL JOURNAL:

[a] P. Li, Y. Benezeth, K. Nakamura, R. Gomez, C. Li and F. Yang. (2020). "Video-Based Pulse Rate Variability Measurement Using Periodic Variance Maximization and Adaptive Two-Window Peak Detection," *Sensors*, 20(10):2752.

REFERENCES

- [Acharya et al., 2006] Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., et Suri, J. S. (2006). **Heart rate variability: a review**. *Medical and biological engineering and computing*, 44(12):1031–1051.
- [Alekhin et al., 2013] Alekhin, M., Anishchenko, L., Tataraidze, A., Ivashov, S., Parashin, V., et Dyachenko, A. (2013). **Comparison of bioradiolocation and respiratory plethysmography signals in time and frequency domains on the base of cross-correlation and spectral analysis**. *International Journal of Antennas and Propagation*.
- [Alghamdi et al., 2012] Alghamdi, K., Kumar, A., Taïeb, A., et Ezzedine, K. (2012). **Assessment methods for the evaluation of vitiligo**. *Journal of the European Academy of Dermatology and Venereology*, 26(12):1463–1471.
- [Alian et al., 2014] Alian, A. A., et Shelley, K. H. (2014). **Photoplethysmography**. *Best Practice & Research Clinical Anaesthesiology*, 28(4):395–406.
- [Allen, 2007] Allen, J. (2007). **Photoplethysmography and its application in clinical physiological measurement**. *Physiological measurement*, 28(3):R1.
- [Artemyev et al., 2020] Artemyev, M., Churikova, M., Grinenko, M., et Perepelkina, O. (2020). **Robust algorithm for remote photoplethysmography in realistic conditions**. *Digital Signal Processing*, 104:102737.
- [Asthana et al., 2014] Asthana, A., Zafeiriou, S., Tzimiropoulos, G., Cheng, S., et Pantic, M. (2014). **From pixels to response maps: Discriminative image filtering for face alignment in the wild**. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1312–1320.
- [Bal, 2015] Bal, U. (2015). **Non-contact estimation of heart rate and oxygen saturation using ambient light**. *Biomedical optics express*, 6(1):86–97.
- [Belaiche et al., 2019] Belaiche, R., Sabour, R. M., Migniot, C., Benezeth, Y., Ginhac, D., Nakamura, K., Gomez, R., et Yang, F. (2019). **Emotional state recognition with micro-expressions and pulse rate variability**. Dans *International Conference on Image Analysis and Processing*, pages 26–35. Springer.
- [Benezeth et al., 2018] Benezeth, Y., Li, P., Macwan, R., Nakamura, K., Gomez, R., et Yang, F. (2018). **Remote heart rate variability for emotional state monitoring**. Dans *EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 153–156. IEEE.
- [Benitez et al., 2001] Benitez, D., Gaydecki, P., Zaidi, A., et Fitzpatrick, A. (2001). **The use of the hilbert transform in ecg signal analysis**. *Computers in biology and medicine*, 31(5):399–406.

- [Béres et al., 2019] Béres, S., Holczer, L., et Hejjel, L. (2019). **On the minimal adequate sampling frequency of the photoplethysmogram for pulse rate monitoring and heart rate variability analysis in mobile and wearable technology.** *Measurement Science Review*, 19(5):232–240.
- [Berg, 1948] Berg, E. A. (1948). **A simple objective technique for measuring flexibility in thinking.** *The Journal of general psychology*, 39(1):15–22.
- [Bobbia et al., 2016] Bobbia, S., Benezeth, Y., et Dubois, J. (2016). **Remote photoplethysmography based on implicit living skin tissue segmentation.** Dans *23rd International Conference on Pattern Recognition (ICPR)*, pages 361–365. IEEE.
- [Bobbia et al., 2017] Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., et Dubois, J. (2017). **Unsupervised skin tissue segmentation for remote photoplethysmography.** *Pattern Recognition Letters*, 124:82–90.
- [Bousefsaf et al., 2013] Bousefsaf, F., Maaoui, C., et Pruski, A. (2013). **Remote assessment of the heart rate variability to detect mental stress.** Dans *7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 348–351. IEEE.
- [Buddharaju et al., 2008] Buddharaju, P., Pavlidis, I., et Manohar, C. (2008). **Face recognition beyond the visible spectrum.** Dans *Advances in Biometrics*, pages 157–180. Springer.
- [Capdehourat et al., 2011] Capdehourat, G., Corez, A., Bazzano, A., Alonso, R., et Musé, P. (2011). **Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions.** *Pattern Recognition Letters*, 32(16):2187–2196.
- [Castaldo et al., 2016] Castaldo, R., Xu, W., Melillo, P., Pecchia, L., Santamaria, L., et James, C. (2016). **Detection of mental stress due to oral academic examination via ultra-short-term hrv analysis.** Dans *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3805–3808. IEEE.
- [Caugberg et al., 2010] Caugberg, E. C., Kloen, S., Visser, M., de la Rosette, J. J., Babjuk, M., Soukup, V., Pesl, M., Duskova, J., et de Reijke, T. M. (2010). **Narrow band imaging cystoscopy improves the detection of non-muscle-invasive bladder cancer.** *Urology*, 76(3):658–663.
- [Chaithanya et al., 2014] Chaithanya, M., Kishore, K., et Srinivasulu, A. (2014). **Continues blood pressure measurement and data logging device with sms alert.** *International Journal of Multimedia and Ubiquitous Engineering*, 9(8):25–38.
- [Christinaki et al., 2014] Christinaki, E., Giannakakis, G., Chiarugi, F., Padiaditis, M., Iatraki, G., Manousos, D., Marias, K., et Tsiknakis, M. (2014). **Comparison of blind source separation algorithms for optical heart rate monitoring.** Dans *4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*, pages 339–342. IEEE.
- [Conaire et al., 2007] Conaire, C. O., O’Connor, N. E., et Smeaton, A. F. (2007). **Detector adaptation by maximising agreement between independent data sources.** Dans *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–6. IEEE.

- [De Haan et al., 2013] De Haan, G., et Jeanne, V. (2013). **Robust pulse rate from chrominance-based rppg**. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886.
- [De Haan et al., 2014] De Haan, G., et Van Leest, A. (2014). **Improved motion robustness of remote-ppg by using the blood volume pulse signature**. *Physiological measurement*, 35(9):1913.
- [De Kock et al., 1993] De Kock, J., et Tarassenko, L. (1993). **Pulse oximetry: Theoretical and experimental models**. *Medical and Biological Engineering and Computing*, 31(3):291–300.
- [Demirezen et al., 2018] Demirezen, H., et Erdem, C. E. (2018). **Remote photoplethysmography using nonlinear mode decomposition**. Dans *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1060–1064. IEEE.
- [Du et al.,] Du, P., et Lin, S. **Massspecwavelet-package peak detection of mass spectrum by wavelet transform based methods**.
- [Elgendi et al., 2013] Elgendi, M., Norton, I., Brearley, M., Abbott, D., et Schuurmans, D. (2013). **Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions**. *PLoS One*, 8(10):e76585.
- [Esteva et al., 2017] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et Thrun, S. (2017). **Dermatologist-level classification of skin cancer with deep neural networks**. *nature*, 542(7639):115–118.
- [Evans et al., 2013] Evans, S., Seidman, L. C., Tsao, J. C., Lung, K. C., Zeltzer, L. K., et Naliboff, B. D. (2013). **Heart rate variability as a biomarker for autonomic nervous system response differences between children with chronic pain and healthy control children**. *Journal of pain research*, 6:449.
- [Fan et al., 2017] Fan, Q., et Li, K. (2017). **Noncontact imaging plethysmography for accurate estimation of physiological parameters**. *Journal of Medical and Biological Engineering*, 37(5):675–685.
- [Fouad et al., 2019] Fouad, R., Omer, O. A., et Aly, M. H. (2019). **Optimizing remote photoplethysmography using adaptive skin segmentation for real-time heart rate monitoring**. *IEEE Access*, 7:76513–76528.
- [Fu et al., 2016] Fu, H.-Y., Guo, J.-W., Yu, Y.-J., Li, H.-D., Cui, H.-P., Liu, P.-P., Wang, B., Wang, S., et Lu, P. (2016). **A simple multi-scale gaussian smoothing-based strategy for automatic chromatographic peak extraction**. *Journal of Chromatography A*, 1452:1–9.
- [Fukunishi et al., 2018] Fukunishi, M., Mcduff, D., et Tsumura, N. (2018). **Improvements in remote video based estimation of heart rate variability using the welch fft method**. *Artificial Life and Robotics*, 23(1):15–22.
- [Galeano et al., 2012] Galeano, J., Jolivot, R., Benezeth, Y., Marzani, F., Emile, J.-F., et Lamarque, D. (2012). **Analysis of multispectral images of excised colon tissue samples based on genetic algorithms**. Dans *Eighth International Conference on Signal Image Technology and Internet Based Systems*, pages 833–838. IEEE.
- [Gil et al., 2010] Gil, E., Orini, M., Bailon, R., Vergara, J. M., Mainardi, L., et Laguna, P. (2010). **Photoplethysmography pulse rate variability as a surrogate measurement**

- of heart rate variability during non-stationary conditions.** *Physiological measurement*, 31(9):1271.
- [Gupta et al., 2016] Gupta, O., McDuff, D., et Raskar, R. (2016). **Real-time physiological measurement and visualization using a synchronized multi-camera system.** Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pages 46–53. IEEE.
- [Hamedani et al., 2016] Hamedani, K., Bahmani, Z., et Mohammadian, A. (2016). **Spatio-temporal filtering of thermal video sequences for heart rate estimation.** *Expert Systems with Applications*, 54:88–94.
- [Haque et al., 2016] Haque, M. A., Irani, R., Nasrollahi, K., et Moeslund, T. B. (2016). **Heartbeat rate measurement from facial video.** *IEEE Intelligent Systems*, 31(3):40–48.
- [Hassan et al., 2017] Hassan, M. A., Malik, A. S., Fofi, D., Saad, N., Karasfi, B., Ali, Y. S., et Meriaudeau, F. (2017). **Heart rate estimation using facial video: A review.** *Biomedical Signal Processing and Control*, 38:346–360.
- [Hertzman, 1937] Hertzman, A. B. (1937). **Photoelectric plethysmography of the fingers and toes in man.** *Proceedings of the Society for Experimental Biology and Medicine*, 37(3):529–534.
- [Hertzman, 1938] Hertzman, A. B. (1938). **The blood supply of various skin areas as estimated by the photoelectric plethysmograph.** *American Journal of Physiology-Legacy Content*, 124(2):328–340.
- [Heusch et al., 2017] Heusch, G., Anjos, A., et Marcel, S. (2017). **A reproducible study on remote heart rate measurement.** *arXiv preprint arXiv:1709.00962*.
- [Hu et al., 2018] Hu, M., Zhai, G., Li, D., Fan, Y., Duan, H., Zhu, W., et Yang, X. (2018). **Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation.** *PloS one*, 13(1):e0190466.
- [Huang et al., 2015] Huang, R.-Y., et Dung, L.-R. (2015). **A motion-robust contactless photoplethysmography using chrominance and adaptive filtering.** Dans *Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE.
- [Huang et al., 2016] Huang, R.-Y., et Dung, L.-R. (2016). **Measurement of heart rate variability using off-the-shelf smart phones.** *Biomedical engineering online*, 15(1):11.
- [Humphreys et al., 2007] Humphreys, K., Ward, T., et Markham, C. (2007). **Noncontact simultaneous dual wavelength photoplethysmography: a further step toward non-contact pulse oximetry.** *Review of scientific instruments*, 78(4):044304.
- [Iozzia et al., 2016] Iozzia, L., Cerina, L., et Mainardi, L. (2016). **Relationships between heart-rate variability and pulse-rate variability obtained from video-ppg signal using zca.** *Physiological measurement*, 37(11):1934.
- [Isin et al., 2017] Isin, A., et Ozdalili, S. (2017). **Cardiac arrhythmia detection using deep learning.** *Procedia computer science*, 120:268–275.
- [Jang et al., 2014] Jang, D.-G., Park, S., Hahn, M., et Park., S.-H. (2014). **A real-time pulse peak detection algorithm for the photoplethysmogram.** *International Journal of Electronics and Electrical Engineering*, 2(1):45–49.

- [Kakumanu et al., 2007] Kakumanu, P., Makrogiannis, S., et Bourbakis, N. (2007). **A survey of skin-color modeling and detection methods**. *Pattern recognition*, 40(3):1106–1122.
- [Kalkhaire et al., 2016] Kalkhaire, S. D., et Puranik, V. G. (2016). **Remote detection of photoplethysmographic signal and svm based classification**. Dans *International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)*, pages 128–132. IEEE.
- [Kazemi et al., 2014] Kazemi, V., et Sullivan, J. (2014). **One millisecond face alignment with an ensemble of regression trees**. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874.
- [Kessler et al., 2017] Kessler, V., Thiam, P., Amirian, M., et Schwenker, F. (2017). **Pain recognition with camera photoplethysmography**. Dans *Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–5. IEEE.
- [Kim et al., 2018] Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., et Koo, B.-H. (2018). **Stress and heart rate variability: a meta-analysis and review of the literature**. *Psychiatry investigation*, 15(3):235.
- [King, 2009] King, D. E. (2009). **Dlib-ml: A machine learning toolkit**. *The Journal of Machine Learning Research*, 10:1755–1758.
- [Kovac et al., 2003] Kovac, J., Peer, P., et Solina, F. (2003). **Human skin color clustering for face detection**. Dans *The IEEE Region 8 EUROCON, Computer as a Tool.*, volume 2, pages 144–148.
- [Kublanov et al., 2015] Kublanov, V., et Purtov, K. (2015). **Researching the possibilities of remote photoplethysmography application to analysis of time-frequency changes of human heart rate variability**. Dans *International Conference on Biomedical Engineering and Computational Technologies (SIBIRCON)*, pages 87–92. IEEE.
- [Kwon et al., 2015] Kwon, S., Kim, J., Lee, D., et Park, K. (2015). **Roi analysis for remote photoplethysmography on facial video**. Dans *37th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 4938–4941. IEEE.
- [Lempe et al., 2013] Lempe, G., Zaunseder, S., Wirthgen, T., Zipser, S., et Malberg, H. (2013). **Roi selection for remote photoplethysmography**. Dans *Bildverarbeitung für die Medizin 2013*, pages 99–103. Springer.
- [Levenson, 2014] Levenson, R. W. (2014). **The autonomic nervous system and emotion**. *Emotion Review*, 6(2):100–112.
- [Lewandowska et al., 2011] Lewandowska, M., Rumiński, J., Kocejko, T., et Nowak, J. (2011). **Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity**. Dans *federated conference on computer science and information systems (FedCSIS)*, pages 405–410. IEEE.
- [Li et al., 2010] Li, B. N., Dong, M. C., et Vai, M. I. (2010). **On an automatic delineator for arterial blood pressure waveforms**. *Biomedical Signal Processing and Control*, 5(1):76–81.
- [Li et al., 2018] Li, P., Benezeth, Y., Nakamura, K., Gomez, R., Li, C., et Yang, F. (2018). **Comparison of region of interest segmentation methods for video-based heart rate**

- measurements.** Dans *18th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 143–146. IEEE.
- [Liao et al., 2002] Liao, X., et Carin, L. (2002). **A new algorithm for independent component analysis with or without constraints.** Dans *Sensor Array and Multichannel Signal Processing Workshop Proceedings*, pages 413–417.
- [Lin et al., 2017] Lin, Y.-C., et Lin, Y.-H. (2017). **A study of color illumination effect on the snr of rppg signals.** Dans *39th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 4301–4304. IEEE.
- [Liu et al., 2014] Liu, Y., Ouyang, J., et Yan, Y. (2014). **An improved method of measuring heart rate using a webcam.** Dans *Photonics North*, volume 9288, page 92880S. International Society for Optics and Photonics.
- [Lucas et al., 1981] Lucas, B. D., et Kanade, T. (1981). **An iterative image registration technique with an application to stereo vision.**
- [Macwan et al., 2018] Macwan, R., Benezeth, Y., et Mansouri, A. (2018). **Remote photoplethysmography with constrained ica using periodicity and chrominance constraints.** *Biomedical engineering online*, 17(1):1–22.
- [Macwan et al., 2017] Macwan, R., Benezeth, Y., Mansouri, A., Nakamura, K., et Gomez, R. (2017). **Remote photoplethysmography measurement using constrained ica.** Dans *E-Health and Bioengineering Conference (EHB)*, pages 430–433. IEEE.
- [Malik, 1996] Malik, M. (1996). **Heart rate variability: Standards of measurement, physiological interpretation, and clinical use: Task force of the european society of cardiology and the north american society for pacing and electrophysiology.** *Annals of Noninvasive Electrocardiology*, 1(2):151–181.
- [Mannapperuma et al., 2014] Mannapperuma, K., Holton, B. D., Lesniewski, P. J., et Thomas, J. C. (2014). **Performance limits of ica-based heart rate identification techniques in imaging photoplethysmography.** *Physiological measurement*, 36(1):67.
- [Martinez-Herrera et al., 2016] Martinez-Herrera, S. E., Benezeth, Y., Boffety, M., Emile, J.-F., Marzani, F., Lamarque, D., et Goudail, F. (2016). **Identification of precancerous lesions by multispectral gastroendoscopy.** *Signal, Image and Video Processing*, 10(3):455–462.
- [McDuff et al., 2014a] McDuff, D., Gontarek, S., et Picard, R. (2014a). **Remote measurement of cognitive stress via heart rate variability.** Dans *36th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 2957–2960. IEEE.
- [McDuff et al., 2014b] McDuff, D., Gontarek, S., et Picard, R. W. (2014b). **Improvements in remote cardiopulmonary measurement using a five band digital camera.** *IEEE Transactions on Biomedical Engineering*, 61(10):2593–2601.
- [McDuff et al., 2017] McDuff, D. J., Blackford, E. B., et Estep, J. R. (2017). **Fusing partial camera signals for noncontact pulse rate variability measurement.** *IEEE Transactions on Biomedical Engineering*, 65(8):1725–1739.
- [McDuff et al., 2016] McDuff, D. J., Hernandez, J., Gontarek, S., et Picard, R. W. (2016). **Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera.** Dans *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 4000–4004.

- [Michels et al., 2013] Michels, N., Sioen, I., Clays, E., De Buyzere, M., Ahrens, W., Huybrechts, I., Vanaelst, B., et De Henauw, S. (2013). **Children’s heart rate variability as stress indicator: Association with reported stress and cortisol**. *Biological psychology*, 94(2):433–440.
- [Mitsuhashi et al., 2019] Mitsuhashi, R., Iuchi, K., Goto, T., Matsubara, A., Hirayama, T., Hashizume, H., et Tsumura, N. (2019). **Video-based stress level measurement using imaging photoplethysmography**. Dans *International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 90–95. IEEE.
- [Moreno et al., 2015] Moreno, J., Ramos-Castro, J., Movellan, J., Parrado, E., Rodas, G., et Capdevila, L. (2015). **Facial video-based photoplethysmography to detect hrv at rest**. *International journal of sports medicine*, 36(06):474–480.
- [Niu et al., 2017] Niu, X., Han, H., Shan, S., et Chen, X. (2017). **Continuous heart rate measurement from face: A robust rppg approach with distribution learning**. Dans *International Joint Conference on Biometrics (IJCB)*, pages 642–650. IEEE.
- [Parra et al., 2019] Parra, F., Scherer, S., Benezeth, Y., Tsvetanova, P., et Tereno, S. (2019). **Development and cross-cultural evaluation of a scoring algorithm for the biometric attachment test: Overcoming the challenges of multimodal fusion with” small data”**. *IEEE Transactions on Affective Computing*.
- [Pigò et al., 2013] Pigò, F., Bertani, H., Manno, M., Mirante, V., Caruso, A., Barbera, C., Manta, R., Bassotti, G., Olivetti, G., et Conigliaro, R. L. (2013). **i-scan high-definition white light endoscopy and colorectal polyps: prediction of histology, interobserver and intraobserver agreement**. *International journal of colorectal disease*, 28(3):399–406.
- [Po et al., 2018] Po, L.-M., Feng, L., Li, Y., Xu, X., Cheung, T. C.-H., et Cheung, K.-W. (2018). **Block-based adaptive roi for remote photoplethysmography**. *Multimedia Tools and Applications*, 77(6):6503–6529.
- [Poh et al., 2010] Poh, M.-Z., McDuff, D. J., et Picard, R. W. (2010). **Non-contact, automated cardiac pulse measurements using video imaging and blind source separation**. *Optics express*, 18(10):10762–10774.
- [Qi et al., 2017] Qi, H., Guo, Z., Chen, X., Shen, Z., et Wang, Z. J. (2017). **Video-based human heart rate measurement using joint blind source separation**. *Biomedical Signal Processing and Control*, 31:309–320.
- [Rankawat et al., 2015] Rankawat, S. A., Rankawat, M., et Dubey, R. (2015). **Ecg artifacts detection in noncardiovascular signals using slope sum function and teager kaiser energy**. Dans *International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)*, pages 6–10. IEEE.
- [Rodríguez et al., 2015] Rodríguez, A. M., et Castro, J. R. (2015). **Pulse rate variability analysis by video using face detection and tracking algorithms**. Dans *37th Annual international conference of the engineering in medicine and biology society (EMBC)*, pages 5696–5699. IEEE.
- [Rother et al., 2004] Rother, C., Kolmogorov, V., et Blake, A. (2004). **” grabcut” interactive foreground extraction using iterated graph cuts**. *ACM transactions on graphics (TOG)*, 23(3):309–314.

- [Rumiński, 2016] Rumiński, J. (2016). **Reliability of pulse measurements in videoplethysmography**. *Metrology and Measurement Systems*, 23(3).
- [Sabour et al., 2019] Sabour, R. M., Benezeth, Y., Marzani, F., Nakamura, K., Gomez, R., et Yang, F. (2019). **Emotional state classification using pulse rate variability**. Dans *4th International Conference on Signal and Image Processing (ICSIP)*, pages 86–90. IEEE.
- [Salahuddin et al., 2006] Salahuddin, L., et Kim, D. (2006). **Detection of acute stress by heart rate variability (hrv) using a prototype mobile ecg sensor**. Dans *Proceedings of the International Conference on Hybrid Information Technology, Cheju Island, Korea*, pages 9–11.
- [Scalise et al., 2012] Scalise, L., Bernacchia, N., Ercoli, I., et Marchionni, P. (2012). **Heart rate measurement in neonatal patients using a webcam**. Dans *International Symposium on Medical Measurements and Applications Proceedings*, pages 1–4. IEEE.
- [Sinex, 1999] Sinex, J. E. (1999). **Pulse oximetry: principles and limitations**. *The American journal of emergency medicine*, 17(1):59–66.
- [Smith et al., 1999] Smith, R. P., Argod, J., Pépin, J.-L., et Lévy, P. A. (1999). **Pulse transit time: an appraisal of potential clinical applications**. *Thorax*, 54(5):452–457.
- [Stricker et al., 2014] Stricker, R., Müller, S., et Gross, H.-M. (2014). **Non-contact video-based pulse rate measurement on a mobile service robot**. Dans *The 23rd International Symposium on Robot and Human Interactive Communication*, pages 1056–1062. IEEE.
- [Suh et al., 2017] Suh, K. H., et Lee, E. C. (2017). **Contactless physiological signals extraction based on skin color magnification**. *Journal of Electronic Imaging*, 26(6):063003.
- [Sun et al., 2012] Sun, Y., Hu, S., Azorin-Peris, V., Kalawsky, R., et Greenwald, S. E. (2012). **Noncontact imaging photoplethysmography to effectively access pulse rate variability**. *Journal of biomedical optics*, 18(6):061205.
- [Sztajzel et al., 2004] Sztajzel, J., et others (2004). **Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system**. *Swiss medical weekly*, 134(35-36):514–522.
- [Taelman et al., 2009] Taelman, J., Vandeput, S., Spaepen, A., et Van Huffel, S. (2009). **Influence of mental stress on heart rate and heart rate variability**. Dans *4th European conference of the international federation for medical and biological engineering*, pages 1366–1369. Springer.
- [Tasli et al., 2014] Tasli, H. E., Gudi, A., et Den Uyl, M. (2014). **Integrating remote ppg in facial expression analysis framework**. Dans *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 74–75.
- [Taylor et al., 2014] Taylor, M. J., et Morris, T. (2014). **Adaptive skin segmentation via feature-based face detection**. Dans *Real-Time Image and Video Processing*, volume 9139, page 91390P. International Society for Optics and Photonics.
- [Tremper et al., 1989] Tremper, K. K., et Barker, S. J. (1989). **Pulse oximetry**. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 70(1):98–108.

- [Trumpp et al., 2017] Trumpp, A., Bauer, P. L., Rasche, S., Malberg, H., et Zaunseder, S. (2017). **The value of polarization in camera-based photoplethysmography**. *Biomedical Optics Express*, 8(6):2822–2834.
- [Trumpp et al., 2018] Trumpp, A., Lohr, J., Wedekind, D., Schmidt, M., Burghardt, M., Heller, A. R., Malberg, H., et Zaunseder, S. (2018). **Camera-based photoplethysmography in an intraoperative setting**. *Biomedical engineering online*, 17(1):1–19.
- [Tsouri et al., 2015] Tsouri, G. R., et Li, Z. (2015). **On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras**. *Journal of biomedical optics*, 20(4):048002.
- [Tulyakov et al., 2016] Tulyakov, S., Alameda-Pineda, X., Ricci, E., Yin, L., Cohn, J. F., et Sebe, N. (2016). **Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions**. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2396–2404.
- [Valluraiah et al., 2015] Valluraiah, P., et Biswal, B. (2015). **Ecg signal analysis using hilbert transform**. Dans *Power, Communication and Information Technology Conference (PCITC)*, pages 465–469. IEEE.
- [van Gastel et al., 2016] van Gastel, M., Stuijk, S., et de Haan, G. (2016). **Robust respiration detection from remote photoplethysmography**. *Biomedical optics express*, 7(12):4941–4957.
- [Viola et al., 2001] Viola, P., Jones, M., et others (2001). **Robust real-time object detection**. *International journal of computer vision*, 4(34-47):4.
- [Wang et al., 2016] Wang, W., den Brinker, A. C., Stuijk, S., et de Haan, G. (2016). **Algorithmic principles of remote ppg**. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491.
- [Wang et al., 2017] Wang, W., den Brinker, A. C., Stuijk, S., et de Haan, G. (2017). **Robust heart rate from fitness videos**. *Physiological measurement*, 38(6):1023.
- [Wang et al., 2015] Wang, W., Stuijk, S., et De Haan, G. (2015). **A novel algorithm for remote photoplethysmography: Spatial subspace rotation**. *IEEE transactions on biomedical engineering*, 63(9):1974–1984.
- [Wang et al., 2018] Wang, Z., Yang, X., et Cheng, K.-T. (2018). **Accurate face alignment and adaptive patch selection for heart rate estimation from videos under realistic scenarios**. *PloS one*, 13(5):e0197275.
- [Wedekind et al., 2015] Wedekind, D., Trumpp, A., Andreotti, F., Gaetjen, F., Rasche, S., Matschke, K., Malberg, H., et Zaunseder, S. (2015). **Assessment of source separation techniques to extract vital parameters from videos**. Dans *23rd European Signal Processing Conference (EUSIPCO)*, pages 434–438. IEEE.
- [Wei et al., 2017] Wei, B., He, X., Zhang, C., et Wu, X. (2017). **Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions**. *Biomedical engineering online*, 16(1):17.
- [Werner et al., 2014] Werner, P., Al-Hamadi, A., Walter, S., Gruss, S., et Traue, H. C. (2014). **Automatic heart rate estimation from painful faces**. Dans *International Conference on Image Processing (ICIP)*, pages 1947–1951. IEEE.

- [Wiede et al., 2016] Wiede, C., Richter, J., Apitzsch, A., KhairAldin, F., et Hirtz, G. (2016). **Remote heart rate determination in rgb data**. Dans *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*, pages 240–246. SCITEPRESS-Science and Technology Publications, Lda.
- [Wieringa et al., 2005] Wieringa, F. P., Mastik, F., et van der Steen, A. F. (2005). **Contact-less multiple wavelength photoplethysmographic imaging: A first step toward “spo 2 camera” technology**. *Annals of biomedical engineering*, 33(8):1034–1041.
- [Xu et al., 2017] Xu, L., Cheng, J., et Chen, X. (2017). **Illumination variation interference suppression in remote ppg using pls and memd**. *Electronics Letters*, 53(4):216–218.
- [Yang et al., 2016] Yang, Y., Liu, C., Yu, H., Shao, D., Tsow, F., et Tao, N. (2016). **Motion robust remote photoplethysmography in cielab color space**. *Journal of biomedical optics*, 21(11):117001.
- [Yanowitz, 2012] Yanowitz, F. G. (2012). **Introduction to ecg interpretation**. *LDS Hospital and Intermountain Medical Center*.
- [Zhang et al., 2016] Zhang, G., Liu, C., Ji, L., Yang, J., et Liu, C. (2016). **Effect of a percutaneous coronary intervention procedure on heart rate variability and pulse transit time variability: A comparison study based on fuzzy measure entropy**. *Entropy*, 18(7):246.
- [Zhang et al., 2014] Zhang, Z., Luo, P., Loy, C. C., et Tang, X. (2014). **Facial landmark detection by deep multi-task learning**. Dans *European conference on computer vision*, pages 94–108. Springer.
- [Zhao et al., 2018] Zhao, C., Lin, C.-L., Chen, W., et Li, Z. (2018). **A novel framework for remote photoplethysmography pulse extraction on compressed videos**. Dans *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pages 1299–1308. IEEE.
- [Zheng et al., 2009] Zheng, J., Hu, S., Echiadis, A. S., Azorin-Peris, V., Shi, P., et Chouliaras, V. (2009). **A remote approach to measure blood perfusion from the human face**. Dans *Advanced Biomedical and Clinical Diagnostic Systems VII*, volume 7169, page 716917. International Society for Optics and Photonics.
- [Zijlstra et al., 1997] Zijlstra, W., et Buursma, A. (1997). **Spectrophotometry of hemoglobin: absorption spectra of bovine oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin**. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 118(4):743–749.
- [Zong et al., 2003] Zong, W., Heldt, T., Moody, G., et Mark, R. (2003). **An open-source algorithm to detect onset of arterial blood pressure pulses**. Dans *Computers in Cardiology*, pages 259–262. IEEE.

LIST OF FIGURES

1.1	The two branches of autonomic nervous system (source: http://www.dysautonomiainternational.org/page.php?ID=122).	4
1.2	The QRS complex (image taken from [Isin et al., 2017]).	5
1.3	The 12-lead ECG (source: https://www.vyaire.com/products/vyntus-ecg-12-lead-pc-ecg).	6
1.4	The original PPG presented by Hertzman [Hertzman, 1937].	8
1.5	Principle of contact PPG (image taken from [Chaithanya et al., 2014]).	8
1.6	Pulse oximeter (source: BOXYM oFit2 FingerClamp Pulse Oximeter).	8
1.7	The rPPG measurement.	10
1.8	An example of the HRV measurement. The HRV data is calculated as the difference of time of the signal peaks in this case (source: https://blog.ouraring.com/what-is-heart-rate-variability/).	11
1.9	ECG signal and PPG signal (image taken from [Zhang et al., 2016]).	12
2.1	Classical rPPG framework: (a) the original video frames, (b) the detected skins (the white part is the detected skin pixels and the black part is the non-skin pixels), (c) spatially averaged RGB signals, (d) Blood Volume Pulse signal, (e) physiological parameters measurement (e.g. HR, respiratory rate, PRV, etc).	16
2.2	ROI segmentation result examples using (from left to right) (a) the detected face, (b) a cropped version of the face and (c) removing some parts of the face using a segmentation strategy.	18
2.3	ROI segmentation result examples using (from left to right) (a) rule-based skin detection, (b) adaptive skin detection, (c) histogram-based skin detection and (d) landmarks detection.	19
2.4	68 facial landmarks.	20
2.5	Before and after applying the skin color magnification (picture taken from [Suh et al., 2017]). Image (a) and (b) show the skin pixel distributions before and after applying the skin color magnification. Image (c) shows the principle of skin color magnification in a single pixel.	23
2.6	Absorption spectra of hemoglobin. Oxyhemoglobin (1) and deoxyhemoglobin (2) (image taken from [Zijlstra et al., 1997]).	23

2.7	The novel setup with polarization device (image taken from [Trumpp et al., 2017]). (1) Camera system, (2) Multi-wavelength illumination matrix, (3) Polarization filter (film), (4) Polarization filter (glass) with an adaptable angle, (5) Test subject with protection goggles and hairband, (6) Reference PPG sensor.	26
2.8	The novel intraoperative setup (image taken from [Trumpp et al., 2018]). (1) Construction with adjustable arm for the sensing system. (2) Sensing system (enlarged on the right) including NIR illumination, NIR camera, and RGB camera. (3) Recording PC. (4) Patient (face directed towards the cameras). (5) Surgeons and clinical staff.	27
2.9	An example of peak detection of BVP signal.	29
3.1	Landmarks detection by Kazemi (Source: blog.dlib.net/2014/08/real-time-face-pose-estimation.html)	38
3.2	Landmark detections with no facial countour (Source: sites.google.com/site/chehrahome/home/)	38
3.3	Some sample images from the MMSE database.	39
3.4	The contact PPG sensor.	40
3.5	The web camera.	40
3.6	Experimental set up.	41
3.7	The examples of UBFC-RPPG dataset.	41
3.8	Correlation plot between the HR estimated from the rPPG signal (EstimatedHR) and from the contact sensor (HR). n is the number of points. . . .	43
3.9	The PSD of the rPPG signal. The red part (main pulsatile) is the signal and the blue part is the noise.	44
3.10	An example of relatively accurate measurement generated from UBFC-RPPG dataset (the red signal is the HR obtained by rPPG and the green signal is the ground truth).	46
3.11	An example of imprecise measurement generated by UBFC-RPPG dataset (the red signal is the HR obtained by rPPG and the green signal is the ground truth).	46
4.1	Superficial facial blood vessels distribution: a) arteries, b) veins, c) arteries and veins together (image from [Buddharaju et al., 2008]).	50
4.2	Two examples of the subjects that were used to get the model.	51
4.3	The SNR map of the two subjects. The yellow parts have higher SNR and the green/blue parts have lower SNR.	52
4.4	The model. The yellow parts have higher SNR and the green/blue parts have lower SNR.	52

4.5	The signals obtained by three methods in one case. From top to bottom are <i>face</i> , <i>skin</i> and <i>model</i> respectively. The red signals are rPPG measured HR and the green signals are ground truth.	55
4.6	The enlarged details of the HR signals obtained by <i>face</i> and <i>skin</i> from 15 to 20 second. The left image is <i>face</i> and the right image is <i>skin</i> . The red signals are rPPG measured HR and the green signals are ground truth. . .	55
4.7	The enlarged details of the HR signals obtained by <i>skin</i> and <i>model</i> from 20 to 30 second. The left image is <i>skin</i> and the right image is <i>model</i> . The red signals are rPPG measured HR and the green signals are ground truth.	56
4.8	The signals obtained by three methods in the second case. From top to bottom are <i>face</i> , <i>skin</i> and <i>model</i> respectively. The red signals are rPPG measured HR and the green signals are ground truth.	56
4.9	The correlation plot of the HR estimation given by <i>face</i> . The x axis is the estimated HR and the y axis is the ground truth.	57
4.10	The correlation plot of the HR estimation given by <i>skin</i> . The x axis is the estimated HR and the y axis is the ground truth.	57
4.11	The correlation plot of the HR estimation given by <i>model</i> . The x axis is the estimated HR and the y axis is the ground truth.	58
5.1	An example of the Slope Sum Function (SSF) method. The original BVP signal generated from the video is black and the SSF signal is blue. . . .	61
5.2	The comparison of Local Maximum and SSF methods on the MMSE dataset. The black signal is the remotely measured BVP signal and the blue signal is the SSF signal.	65
5.3	The comparison of Local Maximum method and SSF method on the MMSE dataset. The black signal is the remotely measured BVP signal and the blue signal is the SSF signal.	66
5.4	The comparison of the Local Maximum method and SSF method on the UBFC-RPPG dataset. The remotely measured BVP signal is black and the SSF signal is blue. The green dots are the peaks detected on the original BVP signal with Local Maximum method and the red dots are the peaks detected with the SSF method.	66
5.5	An example where the window size is smaller than the length of rising part. The remotely measured BVP signal is black and the SSF signal is blue.	68
5.6	An example where the window size is larger than the length of rising part. The remotely measured BVP signal is black and the SSF signal is blue. . .	68
5.7	An example of the Adaptive SSF method on the MMSE dataset. The remotely measured BVP signal is black and the SSF signal is blue.	70
5.8	The comparison of the Adaptive SSF method and SSF method on the MMSE dataset. The remotely measured BVP signal is black and the SSF signal is blue.	70

- 5.9 The comparison of the Adaptive SSF method and SSF method on the UBFC-RPPG dataset. The remotely measured BVP signal is black and the SSF signal is blue. 72
- 5.10 An example of block of interest. The black signal is the remotely measured BVP signal. The green and blue signals are MA_b and MA_p respectively. The strips are the blocks of interest. 73
- 5.11 An example of peak detection with Two-Window method. The black signal is the remotely measured BVP signal. The green and blue signals are MA_b and MA_p respectively. 74
- 5.12 The examples of peak detection with Local Maximum, Adaptive SSF and Two-Window methods with a video from MMSE dataset. The black signal is the remotely measured BVP signal. 76
- 5.13 The examples of peak detection with Adaptive SSF method and Two-Window method with UBFC-RPPG dataset. The black signal is the remotely measured BVP signal. 78

LIST OF TABLES

3.1	The average evaluation values of MMSE dataset.	45
3.2	The average evaluation values of UBFC-RPPG dataset.	45
4.1	The average evaluation values for different ROI detection methods with UBFC-RPPG dataset.	54
4.2	The average evaluation values for different ROI detection methods with MMSE dataset.	57
5.1	Evaluation Metrics.	63
5.2	The average peak detection errors with MMSE dataset.	63
5.3	The average PRV errors with MMSE dataset.	63
5.4	The average errors of PRV features with MMSE dataset.	63
5.5	The average peak detection errors with UBFC-RPPG dataset.	64
5.6	The average PRV errors with UBFC-RPPG dataset.	64
5.7	The average errors of PRV features UBFC-RPPG dataset.	65
5.8	The average peak detection errors with MMSE dataset.	69
5.9	The average PRV errors with MMSE dataset.	69
5.10	The average errors of PRV features with MMSE dataset.	69
5.11	The average peak detection errors with UBFC-RPPG dataset.	71
5.12	The average PRV errors with UBFC-RPPG dataset	71
5.13	The average errors of PRV features UBFC-RPPG dataset	71
5.14	The average peak detection errors with MMSE dataset.	75
5.15	The average PRV errors with MMSE dataset.	75
5.16	The average errors of PRV features with MMSE dataset.	77
5.17	The average peak detection errors with UBFC-RPPG dataset.	77
5.18	The average PRV errors with UBFC-RPPG dataset.	77
5.19	The average errors of PRV features with UBFC-RPPG dataset.	77

LIST OF ACRONYMS

- **ANS** : Autonomic Nervous System. 3
- **BSS** : Blind Source Separation. 20
- **BVP** : Blood Volume Pulse. 17
- **ECG** : Electrocardiography. 4
- **FFT** : Fast Fourier Transform. 11
- **HF** : High Frequency. 12
- **HR** : Heart Rate. 3
- **HRV** : Heart Rate Variability. 10
- **ICA** : Independent Component Analysis. 17
- **JADE** : Joint Approximate Diagonalization of Eigenmatrices. 20
- **LF** : Low Frequency. 12
- **MA** : Moving Average. 72
- **PCA** : Principal Component Analysis. 20
- **PPG** : Photoplethysmography. 6
- **PRV** : Pulse Rate Variability. 12
- **PSD** : Power Spectral Density. 11
- **ROI** : Region of Interest. 17
- **rPPG** : Remote Photoplethysmography. 9
- **RR** : Respiratory Rate. 21
- **SNR** : Signal-to-Noise Ratio. 37
- **SSF** : Slope Sum Function. 30
- **VLF** : Very Low Frequency. 11

