Semi-blind Source Extraction Methods: Application to the measurement of non-contact physiological signs

Richard Macwan

▶ **To cite this version:**

Richard Macwan. Semi-blind Source Extraction Methods: Application to the measurement of non-contact physiological signs. Computer Vision and Pattern Recognition [cs.CV]. Université de Bourgogne - Franche Comté, 2018. English. NNT : . tel-02080653

HAL Id: tel-02080653
https://theses.hal.science/tel-02080653

Submitted on 26 Mar 2019

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.

UBFC
UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ

UNIVERSITÉ DE BOURGOGNE

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

RICHARD MACWAN

Semi-blind Source Extraction Methods

Application to the measurement of non-contact physiological signs

Thèse présentée et soutenue à Dijon, le 14 Décembre 2018

Composition du Jury :

PRUSKI ALAIN	Professeur des Universités, Université de Lorraine	Rapporteur
VEREL SEBASTIEN	Professeur des Universités, Université du Littoral Côte d'Opale	Rapporteur
SAUBION FRÉDÉRIC	Professeur des Universités, Université d'Angers	Président
FOFI DAVID	Professeur des Universités, Université de Bourgogne	Examinateur
TREUILLET SYLVIE	Maître de Conferences, Université d'Orléans	Examinateur
BENEZETH YANNICK	Maître de Conferences, Université de Bourgogne	Encadrant
MANSOURI ALAMIN	Professeur des Universités, Université de Bourgogne	Directeur de thèse

N° | X | X | X |

Title: Semi-blind Source Extraction Methods
Application to the measurement of non-contact physiological signs

Keywords: Semi-blind source extraction methods, Remote photoplethysmography, Constrained Independent Component Analysis, integration of biophysical constraints

Abstract:

Non-contact physiological measurements are highly desirable in many biomedical fields such as diagnosis of infants, geriartic patients, patients with extreme physical trauma, and fitness and well-being. Remote photoplethysmography is increasingly being used for non-contact measurement of heart rate from videos which is one of the most common biomedical property required for most medical diagnosis. One of the common techniques for performing remote photoplethysmography involves using Blind Source Separation (BSS) methods to extract the cardiac signal from video data.

In this context, the objective of this thesis is to develop different methods in the field of extraction and separation of sources by improving upon traditional BSS methods. These novel semi-blind source extraction methods are integrated with biophysical constraints, and applied to the application of remote photoplethysmography measurement. In addition, one of these methods is extended to measure the spatial distribution of photoplethysmographic signals of the skin.

Remote photoplethysmography aims to measure biophysical parameters such as heart rate and heart rate variability by quantifying the periodic changes in skin color due to the rhythmic beating of the heart. These changes manifest in the image data obtained from simple video cameras, which is processed to generate a temporal signal representing the cardiac signal. We have improved existing methods by incorporating the ubiquitous property of quasi-periodicity of biophysical signals such as cardiac and neurological signals. Quasi-periodic signals have higher autocorrelation than non-periodic signals. This observation was combined with independent component analysis techniques and Generalized Eigenvalue Decomposition (GEVD) to develop semi-blind source extraction methods.

Titre : Méthodes d'extraction semi-aveugle
Application à la mesure des signes physiologiques sans contact

Mots-clés : Méthodes d'extraction semi-aveugle, Photopléthysmographie à distance, L'analyse de composantes indépendantes,L'analyse de composantes indépendantes contraint,Intègration des contraintes biophysiques

Résumé :

De nombreuses applications pourraient bénéficier de la mesure de paramètres physiologiques sans contact. On peut citer par exemple le suivi de constantes vitales en milieu hospitalier, en particulier chez les nourrissons, les personnes âgées ou les patients souffrant de traumatismes physiques extrêmes, mais également dans le cadre du sport et du bien-être. La photopléthysmographie sans contact est de plus en plus utilisée pour la mesure de la fréquence cardiaque à partir de vidéos. Les variations périodiques de la couleur de la peau, dues aux battements cardiaques, sont quantifiées pour générer un signal temporel à partir des séquences d'images obtenues par les caméras vidéo.

Une des techniques les plus courantes de photopléthysmographie sans contact consiste à utiliser des méthodes de séparation aveugles de sources pour extraire le signal cardiaque des données vidéo. Dans ce contexte, l'objectif de cette thèse est de développer différentes méthodes dans le domaine de l'extraction et de la séparation des sources améliorant les méthodes aveugles traditionnelles telles que l'Analyse en Composantes Indépendantes. Ces nouvelles méthodes d'extraction de sources semi-aveugle incorporent des contraintes biophysiques et sont appliquées au domaine de photopléthysmographie sans contact. Nous avons utilisé par exemple les propriétés de quasi-périodicité, communes à beaucoup de signaux biomédicaux, et de chrominance liée aux caractéristiques optiques de la peau comme information a priori pour guider les techniques de séparation de sources. De plus, ces méthodes ont été étendues pour également mesurer la distribution spatiale des signaux photopléthysmographiques.

UBFC UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ
Université Bourgogne Franche-Comté
32, avenue de l'Observatoire
25000 Besançon, France

ACKNOWLEDEMENTS

I wish to thank the members of my jury for agreeing to read the manuscript and to participate in the defense of this thesis. I am grateful to Alain Pruski and Sebastien Verel for generously offering their time and efforts for the review of this manuscript. I equally extend my gratitude to the other members of the jury, Frédéric Saubion, David Fofi and Sylvie Treuillet for their constructive and supportive remarks and interest in this thesis.

Firstly, I would like to express my immense gratitude to my supervisor, Dr. Yannick Benezeth for his continuous support during my research, his patience and flexibility with respect to my work process, and most of all, his technical expertise. His constant vigilance with respect to deadlines, even mine, has been vital in realizing the publication of our research.

I am equally grateful to my director, Prof. Alamin Mansouri, who has provided me with all the moral encouragement and help I needed, despite him not being in the same city. He was instrumental in making sure that I do not face any problems during my research, be it arranging for international conferences, or making sure that I was able to fulfill my teaching responsibilities at Auxerre. His responsiveness has been extremely influential in order to ensure the timely advancement of my research.

I also thank Université de Bourgogne for providing me the opportunity to pursue this research, especially the VIBOT programme for introducing me to the exciting world of research.

Last but not the least, I am grateful to my sisters Rashmi, Reena and Priti and my mother Pushpa, for believing in me, for being there for me with their love and emotional support, and their tolerance when I have been far from them and too occupied to remember birthdays and anniversaries. I am also grateful to the other members of my family, especially the Nivens, and friends, especially Roberto, who have supported me all along the way.

Contents

I Context and Motivation 1

1 Introduction 3
1.1 Photoplethysmography and Remote Photoplethysmography 4
1.2 Different paradigms of rPPG measurements 7
1.3 Objectives of the thesis 9
1.4 Plan of the thesis 10

2 State of the art 13
2.1 Source Separation methods 14
2.2 Methods based on physical properties of the skin 16
2.3 Subspace based methods 17
2.4 Smart-ROI based methods 18
2.5 Motion compensation and illumination rectification based methods 20
2.6 Learning based methods 22
2.7 Conclusion 23

II Contribution 25

3 Background 27
3.1 Databases and Experimental setup 30
3.2 Periodicity and Autocorrelation 31
3.2.1 Derivatives of Mean-squared Autocorrelation 34
3.3 Recapitulation 36

4 MAICA for rPPG estimation 37
4.1 Proposed method 38
4.1.1 Objective functions 38
4.1.2 Linear Scalarization 39
4.1.3 Lagrange multipliers and Multi-objective Optimization 40
4.1.4 Parameter free MAICA 40

4.2 System Framework . . . 41
4.3 Results and Analysis . . . 43
4.3.1 Performance analysis of MAICA . . . 44
4.3.2 Performance analysis of parameter-free MAICA . . . 48
4.4 Summary . . . 48

5 Constrained Independent Component Analysis 51
5.1 Proposed method . . . 52
5.1.1 Chrominance based constraint . . . 52
5.1.2 Combination of periodicity and chrominance based constraints . . . 54
5.1.3 Constrained ICA . . . 55
5.1.4 cICA Optimization Algorithm . . . 56
5.2 System Framework . . . 57
5.3 Results and Discussion . . . 57
5.4 Summary . . . 60

6 Periodic Variance Maximization 61
6.1 Related work on periodic signal extraction . . . 62
6.2 Proposed method . . . 63
6.2.1 Iterative Periodic Variance Maximization . . . 64
6.2.1.1 Estimate the basis that maximizes periodicity . . . 64
6.2.1.2 Optimize the periodicity metric $\mathcal{P}(\tau)$ . . . 65
6.2.2 Periodic Variance Maximization applied to rPPG signal estimation . . . 67
6.3 Results and Discussion . . . 67
6.4 Summary . . . 71

7 GEVD in higher dimensions 73
7.1 Spatial rPPG Distribution Estimation . . . 76
7.1.1 Block Diagonalization . . . 77
7.1.2 The QZ Algorithm . . . 79
7.1.2.1 Algorithm details . . . 80
7.1.2.2 Householder Transformations . . . 84
7.2 Proposed higher dimensional versions of the QZ algorithm . . . 87
7.2.1 Block QZ Algorithm . . . 87
7.2.1.1 Block Householder Transformations . . . 87
7.2.2 Tensor QZ Algorithm . . . 90

7.2.2.1 Tensor Householder Transformations 90
7.2.2.2 Stage 1 91
7.2.2.3 Stage 2 91
7.2.2.4 Stage 3 92
7.2.2.5 Stage 4 93
7.3 Experiments and Results 93
7.3.1 Synthetic Example 93
7.3.2 Tensor QZ on the synthetic video 93
7.3.3 Results analysis 95
7.4 Summary 98

III Conclusion 99

8 Conclusions and future work 101

List of Figures 113

List of Tables 115

List of Abbreviations 121

I

Context and Motivation

1

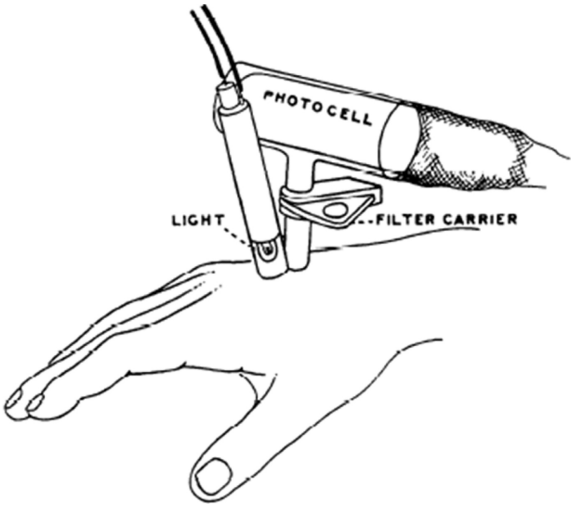
INTRODUCTION

Physiological parameter measurement from biomedical signals is not unbeknownst to applications in fields such as medicine, biomedical health and well being, fitness, and psychology. A large number of signal processing techniques have been employed to analyse such signals.

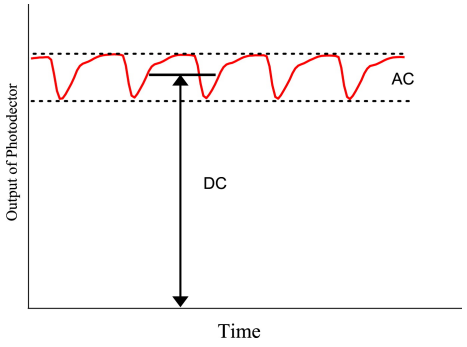
A biomedical signal can be defined as a temporal representation of some activity associated with the human body, and has a characteristic frequency, phase and amplitude. Typical modalities include electrocardiograms (ECG) or photoplethysmograms (PPG) used to measure heart rate (HR) and heart rate variability (HRV), electroencephalograms (EEG) used to measure neural signals, electromyograms (EMG) used to measure muscle electrical activity. Works such as the extraction of fetal ECG from a maternal ECG signal [50], separating artifacts and brain sources from EEG data [37], prosthetic enhancements using EMG signals [52] are few of the many instances of these modalities.

In modern medicine and hygiene, it is becoming increasingly desirable to have less instruments that do more. Non-invasive methods of diagnosis are being researched and are solicited in not just serious medical scenarios but also in day to day fitness and well-being. These requirements have prompted research in non-contact physiological signals measurement to serve as aids for diagnostics and monitoring, and possibly in the future, become the principal means for the same. For instance, much research has been going on in measuring physiological signs such as heart rates and breathing rates from videos by using signal and image processing techniques. Different methods are being developed in order to extract these physiological properties, the main challenges being the extraction of accurate and robust signals on par with contact based sensors such as ECG and finger sensors.

In this thesis, we develop several new techniques to extract the cardiac pulse signal from video frame data. We formulate novel Semi-blind Source Extraction (semi-BSE) methods for the measurement of these *Remote Photoplethysmography (rPPG)* signals which can be useful in various scenarios such as monitoring of geriatric patients, infants, physical trauma victims, fitness and wellbeing, and more. A brief overview of Photoplethysmography (PPG) and rPPG is presented in the next subsection.



(a) The photoelectric plethysmograph Hertzman's original setup from 1938



(b) The signal measured by the photocell

Figure 1.1: PPG setup (left) and the corresponding waveform (right)

1.1/ Photoplethysmography and Remote Photoplethysmography

Photoelectric plethysmography or photoplethysmography (PPG) was first introduced in 1937 by Hertzman where variations in the light absorption of human skin was measured by a photoelectric cell [1] placed over a skin region (typically of the hand) illuminated by a light source above it as seen in figure 1.1. Since this seminal work, much research has been done in understanding this photoplethysmographic waveform whose modulation is not only due to arterial blood pulsation as initially presumed. This presumption on which the pulse oximeter was developed in the 1970s, was later found to be misleading; the PPG waveform is actually more complex and appears to be the sum total of arterial and venous blood interaction with the cardiac, respiratory and autonomic systems. Physiologically, the PPG amplitude is a result of a complex interaction of stroke volume, vascular compliance, and tissue congestion effects [60]. Despite these auxiliary sources of blood pulsation, the effect of arterial pulsations is exploitable enough for modern pulse oximeters to be able to extract and display the oxygen saturation as well as the heart rate derived from the PPG measurements at different wavelengths. This mechanism is explained as follows.

In the original design of Hertzman [1], the light source and sensor were placed obliquely on the flesh tissue, where the amount of light reaching the photocell is reflected by the flesh tissue and the blood pulse therein, as seen in figure 1.1. This is characteristic of *reflective* photoplethysmography. On the other hand, modern pulse oximeters are based on *transmissive* photoplethysmography where the light source and detector are separated by flesh tissue, typically a finger or a toe, where the light measured by the sensor passes through the flesh tissue and thus is modulated by the underlying blood pulse as seen in figure 1.2. This residual light reaching the detector after being absorbed by the finger can be thought of being made up of two components. First, a stationary component that is due to the skin, flesh and non-arterial tissue present between the source and detector termed as the DC component, and second a pulsatile component primarily due

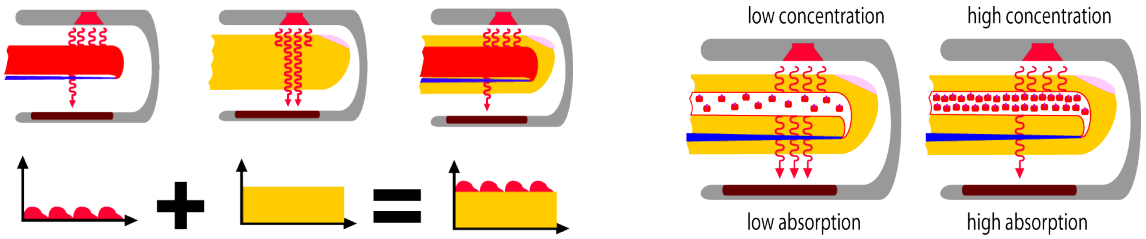


Figure 1.2: PPG measurement using a pulse oximeter. The PPG waveform is made up of AC and DC components (left). The AC component is due to the blood pulse volume (right). Image courtesy of [56]

to the arterial Blood Volume Pulse(BVP), termed as the AC component. This pulsatile component is the direct result of the light absorbed by hemoglobin in the blood which is present mainly in arterial blood. According to Beer's law, the light reaching the detector is inversely proportional to that absorbed by the finger. This combination of AC and DC components results in a signal that can be used to measure heart rate.

Pulse oximeters are classically used to measure oxygen saturation, the cardiac signal can be thought of as a side effect of this measurement. Oxygen saturation is measured based on the fact that different wavelengths affect oxygenated and deoxygneated blood differently. Penetration of light in the skin also depends on the wavelength of the incident light; longer wavelengths penetrate deeper. This means blue light at a wavelength of 400 nm won't penetrate as far as infrared light at a wavelength of 1000 nm. Most pulse oximeter use two sources, a red source at 650 nm and an infrared source at 950 nm, since these two wavelengths provide a difference large enough to warrant adequate disparity in the absorption by oxygenated and deoxygenated hemoglobin that can be exploited to measure the oxygen saturation. Figure 1.3 shows the absorption of light be oxygenated and deoxygenated hemoglobin with respect to different wavelengths.

However, although PPG measurements are non-invasive and relatively easy and portable as compared to ECG signals, they do have certain shortcomings such as extreme susceptibility to motion induced signal corruption, especially since PPG is generally measured at fingers and toes which are prone to random movements more than, for instance, the chest area in case of ECG. Furthermore, non-contact measurements of PPG are highly desirable in scenarios involving infants, older patients, patients with extreme physical trauma such as burn victims, and fitness based scenarios. Recently, the non-invasiveness of PPG has been superseded by that of remote photoplethysmography (rPPG) which is a technique that aims to perform PPG measurements at a distance, without contact.

The basic principal behind remote photoplethysmography derives from reflective photoplethysmography where the light reaching a camera is modulated by the blood pulsations of the skin. The rhythmic beating of the heart results in the pulsating blood volume which in turn results in minute changes in the color of the skin which can be quantified using signal processing to generate a cardiac signal.

Definition 1: Remote photoplethysmography (rPPG)

Remote or imaging photoplethysmography is defined as the technique of using imaging devices such as generic cameras to quantify the blood pulsations into a cardiac signal.

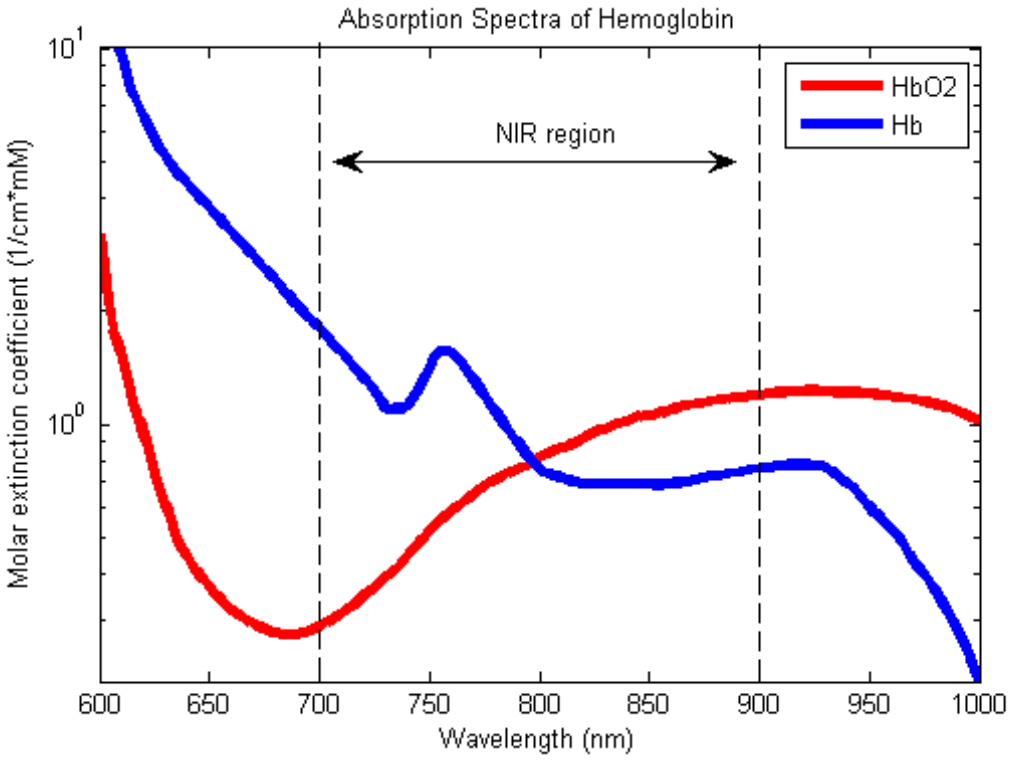


Figure 1.3: Absorption spectra of oxygenated hemoglobin (HbO2) and deoxygenated hemoglobin (Hb) for red and infrared wavelengths. The Molar extinction coefficient is the measure of how strongly a chemical species attenuates light at a given wavelength.

Remote photoplethysmography aims to alleviate the problems faced by photoplethysmography such as single point measurements, sensitivity to motion and the fact that it is contact based, by using inexpensive video cameras. It is also worth mentioning here that there is another technique to obtain the heart rate remotely using a camera by performing ballistocardiography, as in [35], where the minute motions resulting from the blood pulsations are amplified and quantified to form a cardiac signal. This thesis focuses on remote photoplethysmography which quantifies only the *color* changes in the skin due to the blood pulsations. Nevertheless, non-contact photoplethysmography comes with its own set of challenges, especially, noise artefacts owing to motion and illumination perturbances. These problems are not trivial issues, although they are independent enough that they can be solved separately using relevant algorithms and incorporated into an rPPG measurement framework.

Figure 1.4 shows a generic framework for a typical rPPG measurement setup. Video data containing skin pixels from a camera is processed to obtain temporal traces, corresponding to various color bands, typically from the RGB channels. The use of RGB channels is due to the ubiquity of this color space in generic video cameras. However, the formulation of rPPG is decoupled from which physical channels, corresponding to different wavelengths, are actually employed. In consequence, the color space could very well be the Hue-Saturation-Value (HSV) color space, or even have more than 3 color channels, for instance, as in [64].

Let $\mathbf{x} \in \mathbb{R}^{m \times n}$, be a set of temporal signals of length n, where $m = 3$ for RGB channels, represent the temporal signal obtained by frame-wise spatial averaging of the video data.

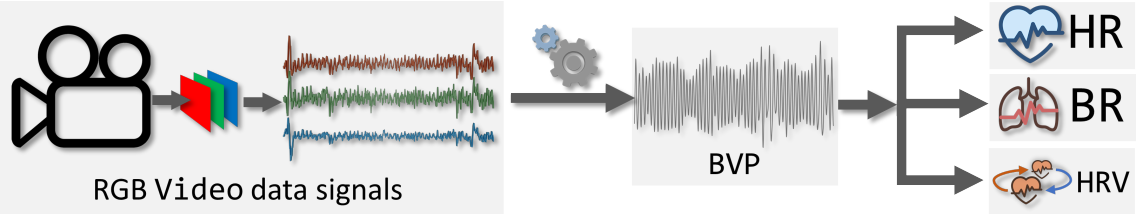


Figure 1.4: The generic framework for physiological parameters estimation using remote photoplethysmography

These temporal traces are then used to extract the BVP signal by using any of the existing or novel rPPG measurement algorithms, some of which are developed in this thesis and will be the subject of discussion in the coming chapters. Furthermore, as is visible in the schematic, physiological parameters other than heart rate, such as heart rate variability and Breathing Rate (BR) can also be extracted from the BVP signal. Heart-Rate Variability (HRV) characterizes the variation in heart rate which is closely related to the breathing rate. A normal healthy human exhibits longer Inter-Beat Intervals (IBI) during exhale as compared to the IBIs during inhale. These physiological parameters are the basis of extensive medical research and diagnosis and their remote, and possibly continuous, measurement can be highly beneficial for diagnosis of heart and breathing related maladies. As a result, recently much research has been focused towards developing efficient algorithms to extract the BVP signal from video cameras. Although, an extensive review of these works will be presented in chapter 2, it is worthwhile to take a bird's eye view of the different paradigms of rPPG measurement which are presented in the next subsection.

1.2/ DIFFERENT PARADIGMS OF RPPG MEASUREMENTS

The problem of rPPG measurement is not trivial, but fortunately owing to the increasing interest therein, it has been attacked from different directions in order to obtain a feasible solution. As a result, numerous algorithms have been developed which can be grouped into different classes. We present a generic categorization of these algorithms as follows. It is to be noted that this is not a complete and exhaustive list of rPPG algorithms, and they will be discussed in more detail in chapter 2.

Source Separation methods In the initial research related to rPPG measurement, the blood volume pulse (BVP) signal was usually extracted from the RGB temporal traces using blind source separation (BSS), of which Independent Component Analysis (ICA) is one of the more commonly used methods.

Definition 2: Independent component analysis (ICA)

Independent component analysis is a signal processing technique for separating a multivariate signal into its constituent components by using metrics based on statistical independence [18].

ICA itself can be performed using different algorithms such as maximization of mutual

information and non-gaussianity [26], and Joint Approximation of Diagonal Eigenmatrices (JADE) [23]. The basis for extracting the BVP signal from the source signals using this method lies in the hypothesis that the original cardiac pulse signal is linearly mixed into the RGB temporal traces, with noise sources representing illumination and motion perturbations. This linear mixture can be separated by using the measures of independence, thus supposedly separating the cardiac signal and the other sources of noise.

Apart from ICA, Principal Component Analysis (PCA) [53] and Non-linear Mode Decomposition (NMD) [89] are other BSS methods that have been used to extract the BVP signal in the context of rPPG measurements. These methods are decomposition techniques that separate and eventually extract the desired signal from the input mixture based on implicit properties. For instance, PCA finds a subspace in which the covariance between the color channels is maximized, which should extract the most significant signal components. In case of NMD, the desired signal is extracted by decomposing the signal into several Time-Frequency Representation (TFR) modes, much like Fourier decomposition.

Chrominance based methods Another class of methods focuses on exploiting the physical characteristics of skin tissue and its distinctive interaction with light. Methods such as CHROM [58], Plane-Orthogonal-to-Skin (POS) method [83] and Blood-Volume Pulse vector (PBV) method [61] fall into this category and are based on a characteristic formulation of models based on physiological properties of the skin. The principal advantages of these methods lies in their computational simplicity, owing to their analytical formulation based on specific skin characteristics.

Subspace based methods Methods that are formulated from the perspective of subspace pursuit fall into this category. This subspace pursuit can be considered analogous to a change of basis, on which the projections of the original RGB signals intensifies the rPPG information, or in other words, amplifies the underlying BVP signal. Such a subspace is eventually similar to a weighting matrix, however it is the manner in which this basis is searched for and the formulation of the methodology focussing on the search of such a subspace, which separates these methods from methods that also search for a weighting matrix, for instance, BSS methods. Methods such as the Spatial-Subspace Rotation (2SR) [83] belong to this class, which exploit the relation of the BVP signal with RGB signals and attempt to define a subspace that fulfils certain criteria which are eventually based on the physical properties of skin and the light reflected therefrom. In this respect, they are similar to chrominance based methods, the difference being in their formulation and treatment of the problem, specifically manipulation and search of subspaces.

Smart-ROI based methods Methods that fall into this class exploit the spatial information from the video data which are overlooked in the previous methods which perform spatial averaging of the skin region e.g. [76], [81] and [67]. This inclusion of the spatial information allows these methods to exploit the face and skin specific information and handle the issue of inhomogeneity in the PPG information across the skin. Moreover, because of their application on the spatial video data, such methods, or parts of such methods can be coupled with other methods such as CHROM (as in [76], [82]) to improve the final BVP signal estimates.

Motion compensation based methods Motion and illumination disturbances are a major issue in rPPG measurements, especially in fitness-based scenarios with repetitive motion, and is an active area of research. Methods that specifically focus on eliminating these issues fall into this category such as the Normalized Least Mean Square (NLMS) adaptive filtering methods proposed by Li *et al.* [63], motion compensated rPPG signal estimation using Active Appearance Models (AAM) by Tasli *et al.* [66], and Sub-band rPPG signal estimation for fitness videos by Wang *et al.* [87].

Learning based methods Methods that take either a probabilistic learning approach such as in [75] and [69] or a traditional machine learning based approach such as in [62] fall under this category. Although a machine learning based approach seems counter intuitive for a signal processing based problem, methods based on learning are not unheard of. However, such methods have yet to provide comparable results to the more robust methods from the aforementioned classes.

1.3/ OBJECTIVES OF THE THESIS

Now that we have a general idea about remote photoplethysmography and the related classes of methods, we are in a position to outline the principal objectives of this thesis. The global objective of this thesis was to advance research in remote photoplethysmography measurements by developing new semi-blind source extraction algorithms. These semi-BSE algorithms were then to be validated against relevant video databases in order to assess their performance. An added advantage would be the adaptability of the developed methods to biomedical signal scenarios other than remote photoplethysmography.

The preliminary objective was to formulate a generic framework that can be used for rPPG signal extraction. This framework is supposed to be the starting point for testing and analyzing different algorithms. It defines a common pipeline based environment for acquiring video data, common preprocessing steps and points at which different rPPG extraction algorithms can be plugged in. This allows for a consistent comparison between different methods.

In addition to the framework, a video database tailored specifically for the analysis of the rPPG extraction algorithms was indispensable. This outlined the second objective, creation of such a video database that can be used for rigorous algorithm analysis. To fulfill this requirement, the database should provide PPG ground truth heart rates as well as the PPG signal waveforms.

Next, the main objective was to develop novel algorithms for extraction of rPPG signals. We choose to develop semi-blind source extraction methods owing to their adaptability with respect to different signal extraction application areas. These novel methods improve upon the traditional blind source separation methods by incorporating biophysical constraints like periodicity information. They can eventually be used in a real-time environment after some computational optimizations.

Finally, a more sophisticated method aiming towards incorporation of spatial relationships between the underlying data was deemed indispensable because of two reasons. First, most rPPG signal extraction methods work on frame-wise averaged signals, ignoring the spatial information present in the frames, retaining which, interesting information

about the underlying data manifest. Second, a spatially adapted method can be useful in applications other than rPPG signal extraction, such as analysis of ECG and neural signals where the separate distinct temporal signals consist of crucial information. To remain in the domain of rPPG measurement, the application of spatial PPG distribution estimation was chosen to validate the high dimensional algorithm.

1.4/ PLAN OF THE THESIS

We present here the outline of the thesis that enabled us to fulfill the objectives set in the previous section.

In the next chapter, the detailed state of the art in rPPG measurement algorithms using the classification presented subsection 1.2 as a basis is presented. The chapter discusses different methods that have been used to perform rPPG signal measurements, highlighting the core algorithms along with the relevant advantages and limitations. The history and background of rPPG signal measurement, is followed by some common source separation methods used thereof. These are followed by methods based on the physical properties of the skin which are in turn followed by methods that extract special subspaces that enhance the desired rPPG signals based on their implicit properties. Next, methods that extract rPPG signals by manipulating spatial regions of the skin content in video frames are discussed followed by methods dealing with motion and illumination compensation, finally followed by learning based methods. Part II then details the main contribution of this thesis. It starts with an introductory chapter presenting the theory behind ICA and autocorrelation, representing the periodicity a priori information, which are significant in the discussion of the methods developed in the chapters that follow. The second part of this chapter describes the experimental setup and the inhouse UBFC-RPPG database which have been quintessential in the analysis of the methods developed in this thesis.

The description of our novel rPPG algorithms starts with Multi-objective optimization using Autocorrelation and ICA (MAICA) in chapter 4 followed by Constrained Independent Component Analysis (cICA) using autocorrelation and chrominance constraint in chapter 5. These algorithms belong to the class of semi-blind BSE methods that augment the ICA BSS method by incorporating a priori information related to the periodicity and physiological properties of the skin in an optimization scheme.

Then follows the Periodic Variance Maximization (PVM) algorithm in chapter 6 that eliminates the expensive optimization scheme with a combination of Generalized Eigenvalue Decomposition (GEVD) by an optimization on a reduced search space to perform the rPPG signal estimation. Finally, a spatial analogue of the PVM algorithm is presented in chapter 7 where the GEVD algorithm is extended to the spatial case, the Tensor GEVD, which can be used in the application of Spatial rPPG Distribution Estimation (SrPDE). It is worth mentioning here that since the PVM algorithm and consequently, the Tensor GEVD algorithm, differs in essence from the cICA and MAICA algorithms, the related state of the art is presented in their respective chapters instead of in chapter 2. Furthermore, the analysis, experiments and performance comparisons related to each method are also presented in their respective chapters for sake of coherence.

Finally, conclusions, limitations and future research paths make up the body of the last chapter where we summarize our contributions, their utility in remote photoplethys-

mography and the possible use of part of our contribution in scenarios other than rPPG measurement.

2

STATE OF THE ART

Remote photoplethysmography has been increasingly used for measuring heart rate, breathing rate and heart rate variability recently because of its non-invasiveness, ease of use and low cost. In section 1.2 we provided a brief overview of different classes of rPPG measurement algorithms. This chapter expands on this summary and provides a detailed review of the existing state of the art for rPPG measurement algorithms, including some related state of the art on generic signal extraction methods.

In one of the pioneer works, Verkruysse [42] showed that remote PPG signal extraction could be performed using a simple consumer level camera. Their work postulated that the G channel of the RGB temporal traces contained the most prominent photoplethysmographic signal. These RGB temporal traces were obtained by quantifying frame-wise skin pixel data, for instance by spatial averaging and then concatenating them. The prominence of the PPG signal in the green channel was deemed due to two factors. First, the higher absorption of green light by the oxygenated hemoglobin as compared to red and second, deeper penetration into the skin as compared to blue light. However, this does not disqualify the red and blue channels as sources of complementary rPPG information. Furthermore, a significant aspect of their contribution was the substantiation of extracting a usable cardiac signal by using only ambient light, as opposed to the use of a dedicated light source for PPG.

Although the work of Verkruysse *et al.* was not the first to introduce non-contact PPG, it was indeed the first to extract a cardiac signal using just a simple camera and ambient light. Remote PPG was in fact introduced earlier by Wu *et al.* [35], albeit with a complex setup with a custom built hardware and software based CCD imaging system using a Near-InfraRed (NIR) light source. Another similar rPPG measurement setup was presented by Wieringa *et al.* [33] where they propose a contact-less setup to measure oxygen saturation using a CMOS-camera with apochromatic lens and a dedicated light source, a 3λ-LED-ringlight corresponding to three wavelengths λ_1 = 660 nm, λ_2 = 810 nm and λ_3 = 940 nm, having 100 LEDs per wavelength.

The feasibility of rPPG in a medical scenario, for patients in the Neonatal intensive care unit, has also been investigated in [57]. They corroborated the ability of rPPG to obtain a signal strong enough to extract the heart rate of all of the 19 infant subjects from videos recorded under ambient light. The rPPG signal was obtained from manually selected regions of interest (ROIs) by spatial averaging of the green channel, from recordings of 1-5 minutes using a standard color digital camera at a distance of 1 meter through plexiglass or directly with open incubators. The videos were recorded in uncrompressed AVI format at a resolution of 300×300 pixels, at 15 or 30 frames/s. The study was de-

signed so as to be completely non-invasive, making sure that the infants were not touched or repositioned. Not surprisingly, they also faced certain perturbances due to low ambient light level and infant motion. The reference heart rate was measured using ECG for infants with weight <1 kg and using pulse oximeters for infants with weight >1 kg. Keeping in mind that the ROIs were manually selected, this study proves the capability of rPPG to measure heart rates for not only adults but also infants and can highly benefit in their medical diagnosis.

Many new rPPG measurement algorithms have been introduced recently, although under varying names due to lack of definition of a standard terminology. These works however refer to the same technique of obtaining cardiac pulse signals using video cameras but refer to them as Imaging PPG systems [79], PPG imaging (PPGI) [71], DistancePPG [70], camera photoplethysmography [57], video plethysmography [68], and more, in addition to rPPG. There is an ongoing effort to unify these nomenclatures into a standard one, which we shall endorse in this thesis by referring all these techniques as rPPG. In [79], Sun and Thakur provide an overview of a wide range of rPPG measurement (IPPG) systems demonstrating the research on rPPG and showing its ubiquity and widespread acceptance. In a similar work, McDuff *et al.* provide a review on state of the art rPPG (PPGI) techniques considering measurements other than pulse rate under realistic conditions such as presence of motion artifacts [71].

Having looked at the initial works focused on remote photoplethysmography estimation, we present next a detailed review of the state of the art methods categorised by their classes of measurement as discussed in section 1.2.

2.1/ Source Separation methods

One of the widely used paradigms in ongoing research is centered around employing Blind Source Separation (BSS) techniques to extract clean rPPG signals from simple web cameras. Independent Component Analysis (ICA), a very common BSS algorithm has been used in several works [49, 54, 64, 73]. It is a technique used to decompose a multivariate signal into the constituent signals under the assumption that the input signals be uncorrelated [26]. In this context, rPPG signal extraction can be formulated as a signal separation problem where the periodic cardiac pulse, manifested as minute chromatic variations of the skin color, is linearly mixed into the temporal traces with illumination and noise perturbances, obtained from the video data from cameras. ICA performs this separation by maximizing certain metrics of independence e.g. non-gaussianity and mutual information. More details about the formulation of ICA are presented in chapter 3 of part II.

In [49], Poh *et al.* used the joint approximate diagonalization of eigenmatrices (JADE) [23] implementation of ICA. They obtained the rPPG signal simply by choosing the second independent component which coincidentally exhibited the highest SNR of all the extracted components. In a succeeding improvment to their work, they selected the cardiac signal as the component having the highest peak in its power spectrum calculated using the Lomb periodogram.

The advantage of ICA for rPPG measurements over Principal Components Analysis (PCA), autocorrelation and cross-correlation was also investigated in [59]. They compared the superiority of various algorithms of ICA (JADE [23], FastICA [26], Robust Accu-

rate Direct ICA aLgorithm - RADICAL [30]) against the aforementioned algorithms. While on one hand, the poor performance of PCA can be attributed to its inability to extract periodic information, on the other hand, the poor performance of autocorrelation and cross-correlation can be attributed to only extracting periodic information. They show that among the various ICA algorithms, the RADICAL algorithm performed the best with the lowest standard error values.

In [64], McDuff *et al.* investigated the usage of more color channels with a five band (RGBCO) DSLR camera placed at 3m from the subject, which records in the cyan and orange (CO) channels in addition to the regular RGB channels. They postulated that better rPPG signals were obtained using the GCO channels rather than traditional RGB channels. As in [49], the cardiac signal was obtained by performing ICA using the JADE method and choosing the component having the greatest frequency peak within the range 45-180 bpm in the normalized Fast Fourier transform (FFT) of the extracted signals. From the different combinations of the five bands, the orange band featured in the top ten combinations of channels. The orange band is not far from the green band and thus captures significant PPG information, which supports previous work that showed higher PPG information in the green wavelength [42].

Incorporation of a priori information in order to aid the optimization process is also an interesting approach in signal separation and is also a very important idea with respect to this thesis. Lu *et al.* incorporated this a priori information in the form of a reference signal to which the extracted signal should resemble [34]. They validated their method using synthetic experiments, for instance, fMRI time responses emulating activated brain voxels which are confounded by physiological signals such as cardiac, respiratory and blood flow and electric noise from the scanners. The idea behind this *ICA with reference* method is that the objective function of ICA, irrespective of its implementation, can be incorporated into a constrained optimization scheme, such as Lagrange multipliers, with appropriately defined constraints. In [34], they define the constraint as the closeness to the reference signal in the form of correlation. One important feature of this methodology is the flexibility of adapting it to different objective functions as well as different constraints.

In a work inspired by [34], Tsouri *et al.* proposed a method of constrained ICA to extract the BVP signal from RGB temporal traces using a rectangular pulse as a reference signal [80]. The issue with using a reference for extracting the BVP signal lies in choosing the right frequency for said reference signal that allows the optimization process to extract the correct BVP signal. And since the pulse rate of the BVP signal is itself the unknown variable to be determined, this becomes a "chicken and egg" problem. Tsouri *et al.* solve this problem by doing a frequency sweep over the range of the human heart rate, performing constrained ICA with reference signals of different frequencies and choosing the best BVP signal with the highest SNR. Although this approach works, smarter search methods in contrast to a brute force search over the frequency range can provide an improvement. Furthermore this approach is computationally taxing —around 30 times slower than traditional ICA as reported in [80].

Another BSS method was proposed by Demirezen *et al.* using Non-linear Mode Decomposition (NMD), which is a signal decomposition technique similar to Fourier decomposition, with the difference that it is extracted using a time-frequency representation (TFR) [89]. The rPPG signal is extracted by iteratively separating non-linear modes from the fundamental and true harmonics. These non-linear modes favour oscillatory data against noisy and spurious data in the temporal signals.

In a related work, Tarassenko *et al.* perform the spectral analysis of videos of haemodialysis patients using autoregressive (AR) models instead of fourier transforms. This AR model is generated based on the assumption that the value of the current sample of the rPPG signal is a linear combination of its p previous values and the current value of a white Guassian distribution. They used the model to capture the salient rPPG frequencies and eliminate the effect of light flicker on videos of patients undergoing haemodialysis. They also claim that the coefficients of this AR model can be used to construct accurate spatial distribution of heart rate and breathing rate information from temporal sections containing minimum patient movement. [65].

The well-known technique of dimensionality reduction, Principal Component Analysis (PCA), which is essentially a BSS method as well, has also been used in the context of rPPG measurements. Lewandowska *et al.* performed PCA on the covariance matrix $C = \mathbf{x}\mathbf{x}^T$ of the RGB temporal traces $\mathbf{x}$ as $C = \mathbf{U}^T\Lambda\mathbf{U}$. $\mathbf{U}$ represent the eigenvectors corresponding to the eigenvalues λ_1, λ_2 ,and λ_3 in Λ. The eigenvector with the highest eigenvalue represents the subspace on which the projection of the original signals $\mathbf{x}$ exhibit highest variance. Temporal traces corresponding to the face and forehead ROI regions of stationary subjects were processed using PCA to obtain an effective rPPG signal estimation. However, PCA remains a weak rPPG estimation method in presence of motion artefacts and owing to the fact that periodic information is not considered at all.

2.2/ METHODS BASED ON PHYSICAL PROPERTIES OF THE SKIN

In contrast to the blind source separation methods, another class of methods exploits the effect of light on human skin due to its unique physical characteristics resulting in specific absorption and reflection properties. De Haan *et al.* have introduced *color difference* or *chrominance* based methods that are computationally efficient and relatively robust against small movements [58]. The core idea behind the chrominance (CHROM) method originates from the dichromatic reflection model [19] which states that the light reflected from the skin comprises of two parts: a diffuse component that is due to the light that has traveled through the skin and constitutes the color changes due to the cardiac pulse, and a specular component, which is due to the illuminant and no pulse signal. Herein lies the intuition behind the *color difference*, where two orthogonal chrominance signals are built, for instance from normalized temporal signals R_n, G_n and B_n which are then standardized as $X_s = 3R_n - 2G_n$ and $Y_s = 1.5R_n + G_n - 1.5B_n$ to accommodate for non-white illumination which are then used to estimate the cardiac pulse signal as $S = X_f - \alpha Y_f$. Here X_f and Y_f are bandpass filtered versions of X_s and Y_s respectively and $\alpha = \frac{\sigma(X_f)}{\sigma(Y_f)}$ with $\sigma(.)$ as the standard deviation of the respective chrominance signal. This ratio between the standard deviations allows for minimization of motion disturbances since the motion affects both X_f and Y_f equally while the cardiac pulse does not.

De Haan *et al.* further advanced upon their chrominance based methods proving that the various absorption spectra of arterial blood manifest along a specific vector in a normalized RGB space, termed as the BVP vector, P_{bv} [61]. The method is named Pulse from Blood Volume (*PBV*) giving the notation of the BVP vector as P_{bv}. This normalized blood-volume pulse vector is estimated using the model of Hulsbusch [39] as $P_{bv} = \frac{[\sigma(R_n),\sigma(G_n),\sigma(B_n)]}{\sqrt{\sigma^2(R_n)+\sigma^2(G_n)+\sigma^2(B_n)}}$ where $\sigma(\cdot)$ represents the standard deviation of the corresponding temporal trace. The P_{bv} vector is specific to the camera sensor, which is then used to

search the weights, W_{PBV}, that give the pulse signal S, for which the correlation with the color channels $\mathbf{C_n} = [R_n, G_n, B_n]$ equals P_{bv} i.e

$$S\,\mathbf{C_n}^T = kP_{bv} \iff W_{PBV}\mathbf{C_n}\mathbf{C_n}^T = kP_{bv} \tag{2.1}$$

where $[R_n, G_n, B_n]$ are the temporal traces spatially averaged from video frame data as usual. The weighting vector W_{PBV} can thus be expressed as

$$W_{PBV} = kP_{bv}\mathbf{Q}^{-1} \quad \text{with} \quad \mathbf{Q} = \mathbf{C_n}\mathbf{C_n}^T \tag{2.2}$$

with the scalar k chosen so as to make W_{PBV} of unit length.

In a related work, Gastel *et al.* extend the PBV method by identifying that the optimal PBV signature remains same even when the SNR of the signal drops significantly due to motion or limited measurement area [82]. Using this observation they formulate an adaptive PBV method using multi-site measurements, thereby introducing spatial redundancy to improve the SNR of the measured rPPG signal.

More recently, Wang *et al.* improved upon the previous work of De Haan *et al.* and formulated a new mathematical model that incorporates the relevant optical and physiological properties of skin reflection. Using this model they proposed a new algorithm based on the Plane-Orthogonal-to-Skin (POS) which is a plane orthogonal to the skin-tone in the temporally normalized RGB space, suitable for rPPG pulse extraction [83]. The extraction of the POS cardiac signal can be summarized in three steps. First, perform temporal normalization using $\frac{\mathbf{C_n}^i}{\mu(\mathbf{C_n}^i)}$, where $i = [1, 2, 3]$ corresponds to the RGB channels and μ corresponds to the channel-wise mean. This temporal normalization is aimed to eliminate the dependency of the RGB channels $\mathbf{C_n}$ on the average skin reflection color, which is the DC-level in the rPPG signal, including the color of the source light and the intrinsic skin color. Second, project the RGB signals on to the POS using

$$S = \begin{bmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix} \cdot \mathbf{C_n} \tag{2.3}$$

The final step is to perform tuning as $h = S_1 + \alpha S_2$ where $\alpha = \frac{\sigma(S_1)}{\sigma(S_2)}$ as in the CHROM method to eliminate the variation due to motion disturbances as compared to the cardiac pulse.

2.3/ SUBSPACE BASED METHODS

Subspace based methods belong to this class, which exploit the relation of the BVP signal with RGB signals and attempt to define a subspace that fulfills certain criteria which are eventually based on the physical properties of skin and the light reflected therefrom. In this respect, they are similar to chrominance based methods, the difference being in their formulation and treatment of the problem, specifically manipulation and search of subspaces. However, these methods are not formulated using models based on the properties of the skin.

Wang *et al.* propose the Spatial Subspace Rotation (2SR) method which estimates a spatial subspace of skin-pixels which does not require skin-tone or pulse related priors. One of the novelties of this algorithm is that it is formulated while keeping the spatial information intact, in contrast to most rPPG algorithms which obtain temporal RGB traces

by frame-wise spatial averaging. They formulate that the periodic pulsatile variations of the skin color reside in temporally varying subspaces expressed as a combination of scaling and rotation of the eigenvalue decomposition of the spatial RGB correlation $C = \frac{V^T \cdot V}{N}$, where V represents the $N \times 3$ RGB temporal traces, without subtracting the corresponding channel-wise means and N is the total number of skin pixels. The pulse signal is extracted by projecting RGB temporal traces over a sliding window on to the corresponding window-wise temporally varying subspaces. If the eigenvalue decomposition of $C = U^T \Lambda U$ is written as $C = \lambda_1 \cdot u_1 \cdot u_1^T + \lambda_2 \cdot u_2 \cdot u_2^T + \lambda_3 \cdot u_3 \cdot u_3^T$, and considering U^τ, the subspace in the first frame of a temporal stride of length l, as a reference, its rotation and scaling with respect to the subsequent subspaces, $U_{t,t<l}$, in the stride can be expressed as

$$SR = S^{\mathrm{T}} \odot R' \tag{2.4}$$

$$= \overbrace{\sqrt{\frac{\lambda_1^t}{\lambda_2^\tau}} \quad \sqrt{\frac{\lambda_1^t}{\lambda_3^\tau}}}^{\text{scaling}} \odot \overbrace{{u_1^t}^{\mathsf{T}} \cdot u_2^\tau \quad {u_1^t}^{\mathsf{T}} \cdot u_3^\tau}^{\text{rotation}} \tag{2.5}$$

Another interesting subspace decomposition method, although not used for rPPG signal measurement, is worth mentioning here. In [50], Sameni *et al.* use an iterative subspace decomposition technique to extract fetal cardiac signals from maternal abdominal recordings. More precisely, they extract the subspace in which the fetal ECG signals are amplified by iteratively coupling a linear decomposition step using Generalized Eigenvalue Decomposition (GEVD) with a denoising step. This method is in fact the inspiration for the PVM method developed in this thesis which is discussed in chapter 6 of part II.

2.4/ SMART-ROI BASED METHODS

This class of methods focuses on extracting heart rate from specific regions of the face/skin based on criteria that are supposed to enhance the physiological signal. Tulyakov *et al.* used Self-Adaptive Matrix Completion (SAMC) [81] to perform heart rate estimation from face videos by dynamically selecting the face regions. This is done by extracting face regions, followed by computation of chrominance features over R regions, $\mathbf{C} \in \mathbb{R}^{R \times T}$, using the CHROM method [58], followed by SAMC that estimates the low-rank equivalent matrix, $\mathbf{E}$, of the chrominance matrix $\mathbf{C}$, finally followed by heart rate estimation using FFT of the first principal component of the matrix $\mathbf{E}$. The SAMC procedure eliminates the perturbations due to face movements, facial expressions and illumination noise by estimating the low-rank matrix $\mathbf{E}$ by solving the optimization problem: $\min_{\mathbf{E}} \nu \operatorname{rank}(\mathbf{E}) + \|\mathbf{E} - \mathbf{C}\|_{\mathcal{F}}^2$ where ν is a regularization parameter.

Feng *et al.* propose to improve the extracted rPPG signal by performing K-means clustering on a feature space modeled to select skin ROIs corresponding to good rPPG signals [67]. The dynamic ROI which selects the skin regions corresponding to good quality rPPG signals is estimated by first dividing a fixed ROI into non-overlapped blocks after which two features: cross correlation of consecutive wave segments and SNR of the wave segments are used to select candidate blocks exhibiting good rPPG signals. Then, K-means clustering is performed on the feature map to select clusters with higher mean values of the cross correlation coefficient and the SNR. Finally, these dynamic ROIs are averaged and overlap-added to extract the rPPG signal and measure the heart rate.

In a similar work, Bobbia *et al.* followed a slightly different approach by using temporal superpixels (TSP) as dynamic ROIs in contrast to blocks of ROIs in [67]. These TSPs which correspond to structurally and spatially coherent regions, are used to extract candidate pulse signals which are then merged and averaged, weighted by superpixel-wise pulsatility measures, into an rPPG signal [76]. The pulsatibility measure for the i^{th} TSP is measured as:

$$SNR_i = 10\,log_{10}\left(\frac{\int_{f_1}^{f_2} h^i_{signal}(f)|\mathcal{F}\{S_i(t)\}|^2 df}{\int_{f_1}^{f_2} h^i_{noise}(f)|\mathcal{F}\{S_i(t)\}|^2 df}\right) \tag{2.6}$$

where $\mathcal{F}\{S_i(t)\}$ is the Fourier transform of the rPPG signal of the i^{th} TSP, $f_1 = 40bpm$ and $f_2 = 240bpm$ are the maximum and minimum human heart rate frequencies, and h is the double step function corresponding to the first and second harmonics defined by the convolution $h^i_{signal}(f) = [\delta(f - f^i_0) + \delta(f - 2f^i_0)] * \prod(\pm f_r)$ where δ is the Dirac delta function, f^i_0 the peak of the periodogram, convoluted with the *rect* function, $\prod$ of the half width f_r. This pulsatility measure, SNR_i is then used to implicitly select living skin tissue since it will be high for skin tissue and low for background.

Bobbia *et al.* further proposed a new framework to perform real-time unsupervised remote photoplethysmography based on efficient temporally propagated superpixels which reduces the computation time compared to their previous method in [76] by a factor of 8 without deterioration of the segmentation quality [88]. The method comprises of first initializing seeds based on a fixed grid, followed by iterative and implicit boundaries identification on blocks of pixels decreasing the size of the block recursively, followed by post-processing to ensure continuity and minimal size of the superpixels. The intuition behind the iterative and implicit boundaries identification derives from the fact that coherent pixels with similar chromatic properties are assigned to the same superpixel. As soon as pixels in a block end up being assigned to a different superpixel, this assigning process is stopped, the block is divided into 4 blocks and the process is repeated iteratively. This reduces the number of total pixels processed drastically.

In another work, Kumar *et al.* proposed an automatic weighting method for different tracked regions to construct the rPPG signal where the weights depend on the blood perfusion and incident light intensity in the respective region [70]. The method works by dividing the green channel facial frames into ROIs small enough to warrant homogeneous pulsatile information, under the premise that the large intensity changes would only be due to incident light intensity or surface reflectance, and not due to subsurface reflectance which essentially represents the cardiac pulse. Next, the ROIs are averaged using weights determined by maximum ratio diversity [4] which strives to maximize the signal-to-noise ratio component by necessitating that the weights be proportional to the root-mean-squared (RMS) value of the signal component and inversely proportional to the mean-squared noise in that channel. In practice, the signal is considered as the power spectral density (PSD), $\hat{Y}_i(f)$, of the per-ROI temporal averages $y_i(t)$ around a narrow frequency band $[-b, +b]$ around the coarse heart rate HR of the subject while the remaining part of the spectrum, filtered using a bandpass filter of $[B1, B2] = [0.5Hz, 5Hz]$, is considered noise. The coarse pulse rate PR is measured by assuming the weights, G_i as 1.0 and eliminating large intensity changes in the ROI-wise temporal averages. This process can be expressed mathematically as

$$G_i(PR) = \frac{\int_{PR-b}^{PR+b} \hat{Y}_i(f)df}{\int_{B_1}^{B2} \hat{Y}_i(f)df - \int_{PR-b}^{PR+b} \hat{Y}_i(f)df} \tag{2.7}$$

A similar work inspired by [70] extracts the cardiac pulse using Self-Adpative Signal Separation (SASS) instead of maximum ratio diversity [91]. Although the name suggests a signal separation technique, it is actually a smart-roi method very similar to [70]. After face detection and tracking, the facial region is separated into blocks and chrominance features calculated to estimate raw HR signals, after which SASS is performed to separate noiseless HR signals from the raw sub-signals. The main idea behind SASS stems from the quasi-periodicity of the cardiac pulse, specifically that the cardiac pulse over consecutive periods will exhibit higher correlation than the noise signals. If $\mathbf{u}_{HR}^{m}$ represents the temporal CHROM traces corresponding to m blocks, then the SASS problem reduces to finding the weighting matrix that reduces the error e^{m} which is the difference between the CHROM signals at an instant n and those after a delay D. The SASS model can be written as

$$\min_{\mathbf{w}} u^{m}(n+D) - u^{m}(n)\mathbf{w} \tag{2.8}$$

which is then solved using a Least Mean Square self adaptive filter [21].

In a slightly different work, Vogels *et al.* used an Near-InfraRed (NIR) camera setup to perform continuous pulse-oximetry measurement of subjects during sleep [90]. The use of the infra-red spectrum is interesting because of its ability to *see* in low light environments and its robustness against illumination disturbances. The method works by preprocessing the video frames into rigid blocks, extracting block-wise parallel pulse signals using the PBV method [61], followed by similarity mapping based on the idea that skin regions belonging to the same subject share pulse-similarities as opposed to those from non-living tissue which exhibit no correlation [74], and finally ROI selection using metrics that mandate high SNR, a reasonable sub-region size, and homogeneity of the skin clusters, finally followed by pulse extraction and Blood Oxygen Saturation (SpO_2) estimation.

2.5/ Motion compensation and illumination rectification based methods

RPPG measurement algorithms are not exempt from the issues related to motion and illumination disturbances which are faced by PPG sensors. Furthermore, the problem is exemplified with the existence of illumination perturbances as well. Fortunately, these disturbances can be modelled and mitigated using compensation techniques, which itself is an important subject of research and is fairly decoupled from the rPPG signal estimation problem. To be more precise, these techniques are generally constituted in the preprocessing stage of different image processing algorithms. However, there have been certain works related to rPPG signal estimation which have motion compensation and illumination rectification techniques ingrained in the respective core algorithms. This class aims to group such methods which focus on combating motion and illumination disturbances in the context of rPPG signal measurement.

Butler *et al.* have assessed the effect of the topology and optical variations of human skin in relation to horizontal movements of the subject and showed that, in presence of motion, the quality of the rPPG signal is determined by the properties of the area of skin chosen [77].

In another work, Wang *et al.* extract the BVP signal based on the ability of a cam-

era to sample multiple skin regions in parallel, thereby delivering the possibility of using these regions as independent pixel based rPPG sensors for pulse measurement, providing redundant information to boost the signal quality. This spatial redundancy of the image sensor is exploited to extract the pulse signal while eliminating motion induced noise. The BVP signal is constructed by using these pixel based rPPG sensors, which are candidate pixels which exhibit high rPPG information. These sensors are extracted by performing motion compensated pixel-to-pixel pulse extraction based on optical flow vectors, spatial pruning by eliminating non-skin pixels and skin pixels with motion-induced disturbances, and temporal filtering. They are then pruned by adaptive band-pass filtering after which PCA is applied to choose the temporal traces exhibiting high rPPG information [73]. ANOVA analysis on their results with post-hoc comparison reports significant improvement on motion robustness.

Li *et al.* use face tracking and Normalized Least Mean Square (NLMS) [21] adaptive filtering methods to compensate against illumination variations. They also perform non-rigid motion elimination by discarding temporal segments of the signal having high standard deviation from the signal mean [63]. The illumination rectification is performed by exploiting the fact that the color component of the green channel of a motionless face, $\mathbf{g}_{face}$ is the sum of the variations due to the blood volume pulse, $\mathbf{s}$ and temporal environmental illumination variations, $\mathbf{y}$ and that the blood volume pulse affects only the face region, and not the background of the video. This fact correlates $\mathbf{y}$ to the mean temporal background signal $\mathbf{g}_{bg}$ using a linear function as $\mathbf{y} \approx h\mathbf{g}_{bg}$, where h represents the coefficient of the linear function. The problem then is reduced to minimizing $\mathbf{y} - h\mathbf{g}_{bg}$ to find the optimal h, which is done by using the Normalized Least Square adaptive filter.

Tasli *et al.* propose a method for rPPG measurement where free movement of the head is allowed, by performing facial appearance modelling for stabilizing color variations in the selected facial region during the signal acquisition phase [66]. For allowing the free head movement facial landmark locations based on Active Appearance Models [22] is performed and the parametric representation of the facial appearance is computed using the FaceReader [31] framework. This localization of the facial landmarks allows the tracking of a selected ROI over the video and thereby obtaining a robust, motion compensated rPPG signal.

Another work that tackles the challenging problem of continuous driver monitoring using rPPG measurement has been proposed by Nowara *et al.* [93]. They argue that illumination variations are significantly reduced in the NIR bandwidth. To solve the problem of continuous monitoring, along with the issue of significantly more motion disturbances and lower SNRs, they propose an rPPG signal tracking and denoising algorithm called sparsePPG based on Robust PCA and sparse frequency spectrum estimation in the NIR band. The idea behind sparsePPG originates from the fact that since the human heart rate spectrum is quasi-periodic, composed of a dominant frequency and its harmonics, its frequency spectrum must be sparse.

Recently, Wang *et al.* have also tried to improve rPPG measurements during fitness exercises, from subjects running on a treadmill [87]. Their proposed method called Subband rPPG, suppresses different distortion-frequencies using independent combinations of color channels, based on the idea that the degrees of freedom of noise reduction can be increased by decomposing the n-wavelength camera signals into multiple orthogonal frequency bands. In another work, they exploit the limited variation of human relative pulsatile amplitude to design a low cost filtering method called amplitude selective filtering

[86]. The spectral amplitude of, e.g. the R channel, is used to select the frequency components in the RGB channels inside the assumed "characteristic pulsatile amplitude range" for pulse extraction, while pruning the rest of components as noise.

In another recent work, Park *et al.* take a slightly different approach by using direct-global separation to suppress direct surface reflection from skin video data in order to improve rPPG estimations [94]. The core idea behind their work is the separation of the two components of the light reflected from the skin of human subjects: direct and global (indirect). The first is due to direct illumination by the light source, and second is predominantly composed of the sub-surface reflection components, i.e. the photons that have penetrated and interacted with the skin tissue. To achieve this, they investigate two techniques, cross-polarization and structured illumination to improve rPPG SNRs. However, their method does require a slightly more complicated setup, with the use of either a camera with a polarizer or a structured (polarized) light source, or both.

2.6/ LEARNING BASED METHODS

Interestingly, machine learning has also been investigated to obtain rPPG measurements. Osman *et al.* have trained a discriminative statistical model to estimate the Blood Volume Pulse (BVP) signal from the human face using ambient light to obtain promising results [72]. They use the variability of the mean of the green channel over a window of $w_s = 0.5s$ seconds as a feature. The windows are discretized to $n = 10$ bins to compensate for variable framerate and their first derivative is then used as the feature to train a discriminative Support Vector Machine (SVM) model. Although they report improvements in speed and accuracy against the ICA method of Poh [49], it is evident that the SVM classifier might not work in a different environment with a different lighting setup. For a more effective classifier, extensive training would be required to detect heart beats over stochastic lighting scenarios, while the problem of motion disturbances still remains to be tackled.

In another work, Alqaraawi *et al.* use a probabilistic approach based on Bayesian learning to obtain better estimates of HRV from PPG signals recorded by wearable devices [75]. They first use the automatic multi scale-based peak detection (AMPD) algorithm which is then enhanced using a Bayesian inference formulation:

$$\text{posterior}_{(i)} = P(\theta_i|\text{APMD}_{\text{output}}(i)) = P(\text{APMD}_{\text{output}}(i)|\theta_i) \cdot P(\theta_i)_{\text{prior}} \tag{2.9}$$

where θ_i is the probability of having a peak at sample i. This formulation is used to estimate the probability of the next peak occurrence over consecutive peak cycles.

Interestingly, a machine learning approach can be beneficial to be applied in the early image processing stage before extracting the temporal signals for improving detections or combating low resolution issues. In this context, Mcduff *et al.* have used deep super resolution for allowing the extraction of physiological information from low resolution videos [92]. They use a deeply-recursive convolutional network (DRCN) [78] to infer the missing values in the up-scaled version of very low-resolution video frames of 41×30 pixels (originally down-sampled from $658 \times 492 pixels$). However, it is observable that this method would only be beneficial for low resolution image data. The performance gains for videos with adequate resolution would be insignificant to warrant its use where the videos have been recorded with an adequate resolution.

2.7/ CONCLUSION

We saw many interesting works in the field of rPPG signal estimation in this chapter. Although this discussion is undoubtedly not exhaustive, especially owing to the growing interest in this field of biomedical signal analysis, it does cover a the majority of methods that we have attempted to classify. We reiterate here that the overview of the state of the art presented in this chapter is related to the application of rPPG signal measurement. On the other hand, the methods developed in chapters 6 and 7 stem from a more generic class of algorithms and thus to maintain coherence in the subject matter, the related state of the art is presented in their respective chapters. We next present the different semi-blind source extraction algorithms that have been developed during the course of this thesis in the next part.

II

CONTRIBUTION

3

BACKGROUND

Semi-blind Source Separation and Extraction methods get their name based on their capacity to extend and improve upon blind source separation methods. They were in fact formulated to overcome the shortcomings of the latter, specifically the *blindness*. The mitigation of this blindness is generally achieved by incorporating some kind of information about the desired signals in a given application scenario. As a result, these methods are often tailored to the application in question. It is owing to this very fact that a generic semi-blind source separation or extraction method is invaluable. It is worth noting here that we restrict ourselves to the application domain of semi-blind source *extraction* methods, where only a specific signal is desired, e.g. extraction of heart rate in a rPPG measurement scenario, or estimation of Fetal ECG from maternal ECG.

To clearly understand the utility of semi-blind source extraction, it is worthwhile at this moment to look at the basic formulation of the most quintessential blind source separation methods, Independent Component Analysis or ICA. Although blind source separation methods are diverse and varied, ICA was chosen owing to its extensive use, especially in the rPPG measurement domain.

ICA is a well-known blind source separation (BSS) technique that aims to solve problems such as the Cocktail Party problem. The cocktail party effect [2] is the phenomenon by which the human brain can seamlessly separate and focus on one of the many audio/vocal sources from a mixture, for instance in a noisy room, which is quite common at a cocktail party. ICA was originally used in audio signal based problems, but has been recently adapted to extract cardiac signals in the context of remote photoplethysmography measurement. The formulation of the ICA problem in the context of remote photoplethysmography comprises of employing temporal traces obtained from video data as the input. The cardiac signal is supposed to have been embedded into the periodic changes of the skin color, mixed with perturbance signals corresponding to illumination, motion and other noise sources.

Let $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ be the time varying color traces from n channels, obtained by linear mixing of m independent source signals denoted as $\mathbf{c} = (c_1, c_2, ..., c_m)^T$. The linear mixing process is then expressed as $\mathbf{x} = \mathbf{Ac}$, where the linear memoryless mixing of the channels is represented by the matrix $\mathbf{A}_{n\times m}$. ICA aims to obtain the linear unmixing matrix $\mathbf{W}_{m\times n}$ to recover all the independent components with minimum knowledge of $\mathbf{A}$ and $\mathbf{c}$. The separated components $\mathbf{y} = (y_1, y_2, ..., y_m)^T$, are obtained by $\mathbf{y} = \mathbf{Wx}$ [34].

ICA succeeds in separating these independent signals by making some assumptions about the statistical properties of the signals $c_i(t)$. These statistical properties are ex-

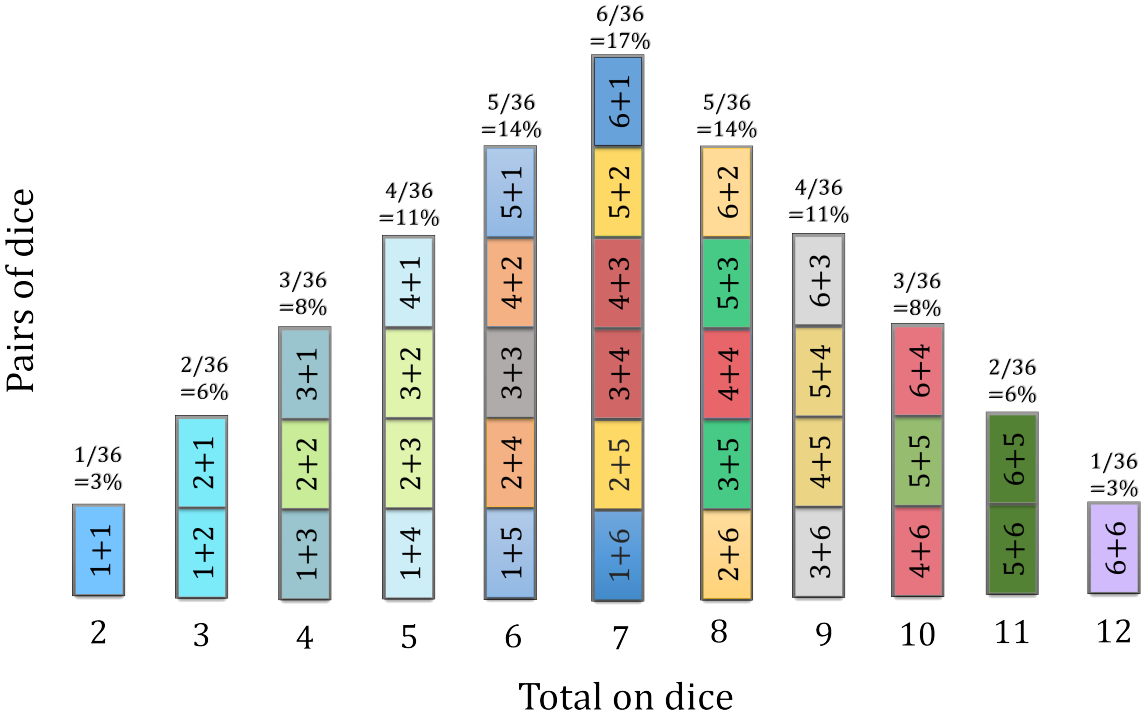


Figure 3.1: Central Limit Theorem. The sum of two random variables tends to be a normal distribution, with a guassian shape.

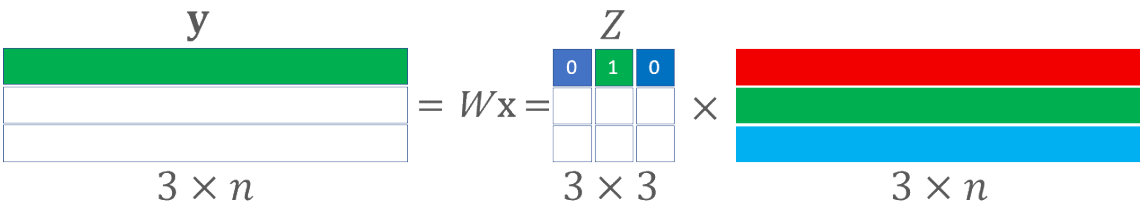


Figure 3.2: Any row of $\mathbf{y} = Z\mathbf{c}$ will be least Gaussian when it is equal to one of $\mathbf{c}_i$, when the corresponding row of Z has only one non-zero element

pressed as a metric that encapsulates the intuitive notion of independence between the signals. The intuition behind independence comes from the central limit theorem, which states that, for a sufficiently large number of samples the sum of independent random variables tends to be a normal distribution, under certain conditions.

Let us consider the well-known dice experiment where the sum of throwing two dice is plotted against their probabilities as shown in figure 3.1. The probabilities of obtaining a sum is proportional to the number of possible ways of attaining that particular sum. For example, there are 6 different combinations of dice values that give the sum 7. A corollary of the central limit theorem is that the sum of two independent random variables tends to be more Gaussian than any of the two original random variables [26]. This means that in the equation $\mathbf{y} = W\mathbf{x} = WA\mathbf{c} = Z\mathbf{c}$, since the separated signals $\mathbf{y}$ are obtained by a linear combination of the original source signals $\mathbf{c}$, any of the separated signals $\mathbf{y}$ shall be (ideally) equal to one of the source signals when the corresponding row of Z has only one non-zero element, in which case, $\mathbf{y}$ is least gaussian. This is depicted in figure 3.2 for $m = 3$. This is the key idea behind ICA, the separated signals

are most independent when they are least gaussian. This property can be exploited in an optimization scheme by using a metric that minimizes (maximizes) the gaussianity (non-gaussianity) of the separated signals $W\mathbf{x}$. Common measures of non-guassianity are kurtosis and negentropy given by equations 3.1 and 3.2 where $E\{\cdot\}$ is the expectation of a discrete random variable, $H(\cdot)$ is the entropy given by $H(Y) = -\sum_i P(Y = a_i)logP(Y = a_i)$ where a_i are the possible values for the discrete random variable Y and $\mathbf{y}_{gauss}$ is a gaussian random variable having the same covariance as $\mathbf{y}$ which is assumed to be a random variable with zero mean and unit variance [26]:

$$kurt(\mathbf{y}) = E\{\mathbf{y}^4\} - 3(E\{\mathbf{y}^2\})^2 \tag{3.1}$$

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \tag{3.2}$$

In practice, approximations of these measures of non-guassianity are used; for simplifying the computational complexity and mitigating non-robustness issues. For instance, the negentropy metric used in our implementation is approximated using

$$J(\mathbf{y}) \propto [E\{G(\mathbf{y})\} - E\{G(\nu)\}]^2 \tag{3.3}$$

where ν is a guassian variable with zero mean and unit variance and G corresponds to special non-quadratic functions that do not grow too fast. Choices of G that have proven useful according to [26] are

$$G_1(u) = \frac{1}{a_1} logcosh(a_1 u), \quad G_2(u) = -exp(\frac{-u^2}{2}) \tag{3.4}$$

In an rPPG signal extraction problem, employing a sufficiently accurate reference signal is prone to the evident complication of choosing its right frequency. This can be done in two possible ways. One alternative is to repeatedly compare the extracted rPPG signal to reference signals of different frequencies, as done by Tsouri *et al.* [80], which as expected is computationally taxing —around 30 times slower than traditional ICA. The other alternative is to update the frequency of the reference signal continuously, in effect making it a parameter to optimize. This increases the complexity of the problem and reduces the probability of convergence. A PPG signal is a very apt reference for rPPG extraction whose synthesis depends critically on the required frequency, even more so than the actual shape of the signal.

This shortcoming or blindness can be overcome by equipping the optimization algorithm with some information about the desired signal. This a priori information is generally aimed to encapsulate the essential properties of the desired signal and if chosen correctly can lead to improved signal extraction. An extremely ubiquitous property of temporal signals is periodicity or quasi-periodicity, which is exhibited by almost all physiological signals, including but not limited to, rPPG, ECG, and EEG [10] . It is owing to this evident universality that we use periodicity as the a priori information which has proven to be valuable resource to improve upon ICA and develop our semi-blind source extraction methods.

The outline of the rest of this chapter is as follows. We start with describing our in-house database that was conceived not only to analyze our rPPG measurement algorithms, but also to serve as a valuable tool for advancing research in this area. Owing to the presence of ground truth PPG waveforms, our database can serve the purpose of

robust algorithm analysis, and has already been requested by several researchers. We also present the common experimental setup that served as a basis for comparing rPPG measurement algorithms developed in this thesis. We then describe the formulation of the periodicity metric in section 3.2, based on autocorrelation and the prerequisites to be able to use it as an a priori information in existing optimization algorithms. We then present the two variants of the first semi-bind source extraction method, Multi-objective optimization using Autocorrelation and ICA (MAICA) in chapter 4 where autocorrelation is combined with the ICA objective function in a multi-objective optimization scheme. Next follow the details about Constrained ICA (cICA) and its improvements in chapter 5 where the autocorrelation function and chrominance are used as constraints to guide the ICA optimization algorithm. In chapter 6 we present another novel method, Periodic Variance Maximization (PVM) which improves upon the optimization based methods in order to tailor the framework to be usable in real-time scenarios. Then, chapter 7 details the extension of the Generalized Eigenvalue Decomposition (GEVD) to the higher dimensional case by formulating a Tensor version of the QZ Algorithm, which is used for performing GEVD of two matrices. This higher dimensional version of the algorithm is capable of performing simultaneous GEVD of pairs of matrices corresponding to multiple data points and can be used to, for instance, estimate the spatial rPPG distribution over the skin by performing a procedure similar to PVM on multiple skin pixels.

3.1/ Databases and Experimental setup

Video databases are indispensable for analysis of rPPG signal measurement algorithms. However, owing to the relative adolescence of this technique, such databases are far from abundant currently. This lack of existing databases tailored towards rPPG measurement analysis, specifically one that can provide accurate PPG groundtruth signals along with the cardiac signals inspired the conception of our inhouse database UBFC-RPPG. Although this database is made public, to assert the efficiency of the algorithms developed in this thesis further, it was also imperative that they be tested against a publicly available database that can be usable for rPPG measurement. Consequently, the MMSE-HR database [84] which is a database originally realized for emotional elicitation, but is useful for rPPG measurement was also used to test our methods.

The UBFC-RPPG database comprises of two datasets which were acquired using the setup shown on the left of figure 3.3. The first, labeled as SIMPLE, comprises of 9 videos (around 21k frames), where the subjects were requested to relax and close their eyes. The second dataset comprises of 46 videos (around 94k frames), labeled as REALISTIC, where the subjects were required to play a time sensitive mathematical game in order to vary the heart rate and also simultaneously emulate the scenario of the typical activity of using a computer. Both the datasets comprise of subjects with skin colors varying from dark to light tones. All the videos were taken under ambient light with limited illumination variations. The video frames were obtained with a custom C++ application using a Logitech C920 web camera placed at a distance of about 1m from the subject with a resolution of 640x480 in 8-bit uncompressed RGB format at 30 frames per second. A CMS50E transmissive pulse oximeter was used to obtain the ground truth PPG data. The experimental setup with sample images from UBFC-RPPG database are shown in figure 3.3 depicting the lighting conditions.

The UBFC-RPPG database is made publicly available along with the ground truth data

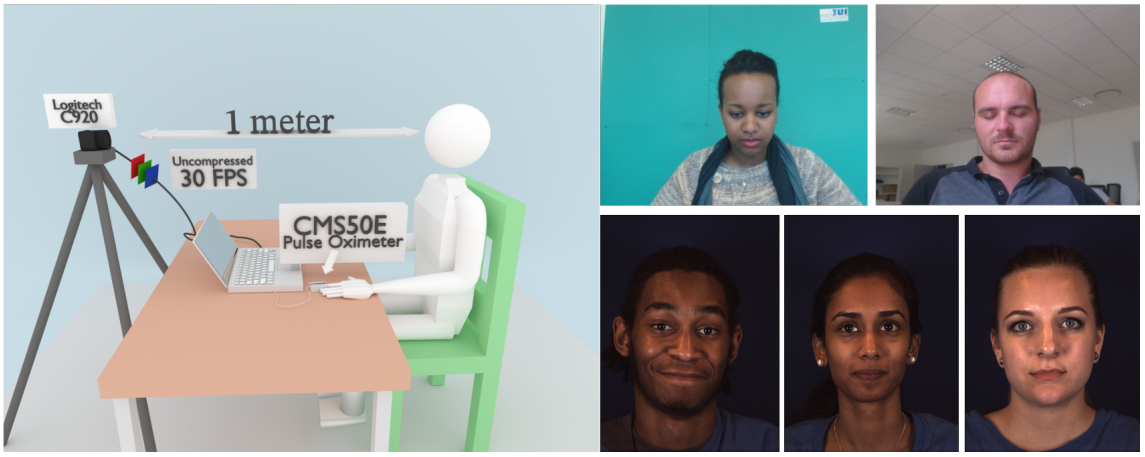


Figure 3.3: The setup for rPPG measurement (left) and sample images from the SIMPLE and REALISTIC datasets from the UBFC-RPPG database (top-right) and the MMSE-HR (bottom-right) database

from the pulse oximeter for rPPG measurement analysis[1].

Presented also in figure 3.3 on the bottom right, are samples from the MMSE-HR database. It comprises of 97 usable videos (about 105k frames) of varying length, at 25 frames per second of varied skin colors. Although, the main objective of the MMSE-HR database was for emotion elicitation, it does provide the video data and HR ground truth data for our validation. It is worth mentioning, though, that the MMSE-HR database only provides the heart rate, and not the actual PPG waveform of the ground truth as compared to our UBFC-RPPG dataset, which proves our earlier argument of databases for rPPG measurement being scarce. Admittedly, when using the UBFC-RPPG database, a more meaningful validation and comparison can be achieved since the same algorithm can be used to extract the heart rate from both the PPG and the rPPG waveforms.

3.2/ PERIODICITY AND AUTOCORRELATION

Quasi-periodicity is a ubiquitous property exhibited by various physiological signals such as electrocardiographic, electromyographic, electroencephalographic and photoplethysmography signals. However, this property has been unfairly ignored by most signal extraction problems, especially in remote photoplethysmography measurement scenarios. This inherent property of biomedical signals can be exploited by using a periodicity metric that can guide the component extraction process to choose the most periodic component.

In order to incorporate periodicity information in existing BSS methods, a metric whose derivative can be calculated analytically would be appropriate. The differentiability of a given constraint is a required condition for most optimization methods which aim to exploit the said constraint. Autocorrelation is such a metric which is simple yet comprehensible, and is analytically well-defined making it feasible to use it in existing optimization schemes as a priori information.

[1] https://sites.google.com/view/ybenezeth/ubfcrppg

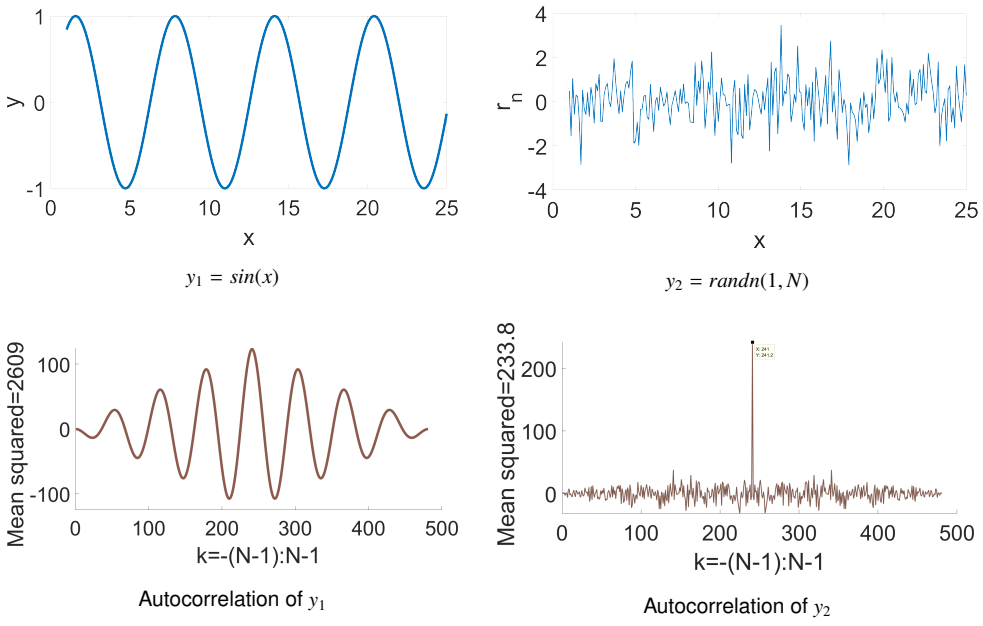


Figure 3.4: Autocorrelation of a sinusoid vs a random signal

Definition 3: Autocorrelation

Autocorrelation is the correlation of a signal with itself at different lag times provided it is sampled at a sufficiently high frequency.

For a time series signal $\mathbf{y} = [y_1, y_2, \ldots, y_N]$ of N elements, its discrete autocorrelation r_k at lags $k \in [-(N-1), \cdots, N-1]$ is given by

$$r_k = \sum_{j=0}^{N-1} y_j \odot \overset{k}{y}_j \tag{3.5}$$

where $\overset{k}{y}_j$ is the j^{th} element of the signal $\mathbf{y}$ lagged (or led if $k < 0$) by k units and padded with zeroes to the left (or right if $k < 0$) and $\odot$ is the element-wise multiplication operator. A periodic signal typically has a higher correlation with itself compared to a non-periodic one which can be quantified by the mean of the squared autocorrelation of the signal and consequently can be used as a measure of the periodicity of a signal. Figure 3.4 depicts the high correlation of a periodic sinusoid compared to that of a uniform random signal with the mean of the squared autocorrelation being much higher than that of the random signal. Our goal is to use this quantification as a priori information for incorporation in existing BSS methods which aim to find an optimum weighting vector $\mathbf{w}^* \in \mathbb{R}^m$ that extracts the desired component from a multivariate signal mixture of m temporal signals.

To make this incorporation feasible, the derivative of this metric needs to be calculated, since derivative based optimization schemes are one of the most common ones, ICA being one of such methods, as seen earlier in this chapter. This derivative needs to be calculated with respect to the weighting vector $\mathbf{w}$. To make this derivation feasible, and simplify its computation, two modifications are advantageous in order to use autocorrelation as a periodicity measure. First, since the autocorrelation function is symmetric, the correlation is only computed for lags $k \in [0, \cdots, N-1]$. Second, since the correlation at

lag 0 is always high, the autocorrelation at lag $k = 0$ is set to *0*, thereby giving the autocorrelation as $\mathbf{r} = [r_1, r_2, \cdots, r_{N-1}]$ comprising of $N-1$ instead of N values, given by equation 3.5 with r_0 set to 0. Keeping in mind that r_k is a scalar, equation 3.5 can be rewritten in matrix notation as

$$r_k = \mathbf{y}[\overset{k}{\mathbf{y}}]^T = \overset{k}{\mathbf{y}}\mathbf{y}^T \tag{3.6}$$

where $\overset{k}{\mathbf{y}}$ is again the signal $\mathbf{y}$ lagged by k units and $k \in [1, \cdots, N-1]$. This formula needs some alteration in order that its derivative be simple to calculate. The complication arises from the fact that the signal is multiplied with its *lagged* version. This product requires special treatment in order that its derivative with respect to other multivariate quantities such as the source signal $\mathbf{y}$ or the weighting vector $\mathbf{w}$ be easily calculable. Considering the definition of autocorrelation in equation 3.6, it can be observed that $\overset{k}{\mathbf{y}}$ can be rewritten as yT_k where T_k is a toeplitz-like matrix that incorporates the lagging at lag k and padding with zeroes of the signal and is given by

$$T_k = \begin{array}{c} \\ k \\ \\ N \\ \\ k \end{array} \begin{array}{c} \begin{array}{ccccccc} k & & & & & N-k & \end{array} \\ \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \end{array} = \begin{bmatrix} 0_{N-k,k} & I_{N-k} \\ 0_{k,k} & 0_{k,N-k} \end{bmatrix} \tag{3.7}$$

T_k is an $N \times N$ matrix composed of the first $N-k$ rows made up of $(N-k) \times k$ zeroes and an identity matrix of size $N-k$, the rest of the elements being zero. Thus, r_k becomes

$$r_k = \mathbf{y}T_k\mathbf{y}^T \tag{3.8}$$

making its differential with respect to $\mathbf{y}$ easier to calculate. To aid the optimization process, a convenient representation of the autocorrelation is the mean-squared autocorrelation which can be expressed as

$$R(\mathbf{y}) = E\{\mathbf{r}^2\} \tag{3.9}$$

with $\mathbf{r} = [r_1, r_2, \cdots, r_N]$ being the autocorrelation as defined in equation 3.6. Using the mean of the squared autocorrelation helps to compensate for the fact that certain values of autocorrelation might be negative, especially if zero mean signals are used. These negative values might possibly nullify the autocorrelation function if the squared or absolute values weren't used.

Furthermore, since $\mathbf{y}$ itself is actually dependent on the ideal weighting matrix $\mathbf{w}^*$, it is the maxima of R with respect to $\mathbf{w}$ that needs to be calculated. For use in optimization schemes of the semi-blind source extraction methods presented in chapters 4 and 5, the first and second derivatives are required. Although, the calculation of these derivatives is not critical to understanding the formulation of the methods developed in this part of the thesis, they are provided in the next section to maintain structure.

3.2.1/ Derivatives of Mean-squared Autocorrelation

Here we present the first and second derivatives of $R(\mathbf{w})$ needed by the Lagrange multipliers method. The derivative of a scalar w.r.t a column vector is a column vector of the same size as that of the vector. The first derivative of $R(\mathbf{w})$ in

$$R(\mathbf{y}) = E\{\mathbf{r}^2\} \tag{3.10}$$

can be obtained as follows considering squared autocorrelation as $\mathbf{r}^2 = [r_1^2\ r_2^2\ ...\ r_N^2]$.

$$R'(\mathbf{w}) = -E\{\frac{\partial}{\partial \mathbf{w}}(\begin{bmatrix} r_1^2 & r_2^2 & \cdots & r_N^2 \end{bmatrix})\} \tag{3.11}$$

where the derivative of the squared autocorrelation $\mathbf{r}^2$ is then obtained using the chain rule of derivatives. Also, we have $\mathbf{y} = \mathbf{w}^T\mathbf{x}$ giving $\frac{\partial \mathbf{y}}{\partial \mathbf{w}} = \mathbf{x}$.

$$\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{w}} = \mathbf{x}\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{y}} \tag{3.12}$$

$$= \mathbf{x}\begin{bmatrix} \frac{\partial(r_1^2)}{\partial y_1} & \frac{\partial(r_2^2)}{\partial y_1} & \cdots & \frac{\partial(r_N^2)}{\partial y_1} \\ \frac{\partial(r_1^2)}{\partial y_2} & \frac{\partial(r_2^2)}{\partial y_2} & \cdots & \frac{\partial(r_N^2)}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(r_1^2)}{\partial y_N} & \frac{\partial(r_2^2)}{\partial y_N} & \cdots & \frac{\partial(r_N^2)}{\partial y_N} \end{bmatrix} \tag{3.13}$$

$$= \mathbf{x}\begin{bmatrix} 2r_1\frac{\partial r_1}{\partial y_1} & \cdots & 2r_N\frac{\partial r_N}{\partial y_1} \\ \vdots & \ddots & \vdots \\ 2r_1\frac{\partial r_1}{\partial y_N} & \cdots & 2r_N\frac{\partial r_N}{\partial y_N} \end{bmatrix} \tag{3.14}$$

$$= 2\mathbf{x}\begin{bmatrix} r_1\frac{\partial r_1}{\partial y_1} & \cdots & r_N\frac{\partial r_N}{\partial y_1} \\ \vdots & \ddots & \vdots \\ r_1\frac{\partial r_1}{\partial y_N} & \cdots & r_N\frac{\partial r_N}{\partial y_N} \end{bmatrix} \tag{3.15}$$

The size of $\frac{\partial(\mathbf{r}^2)}{\partial \mathbf{w}}$ is then $3 \times N$ from the product of $\mathbf{x}_{3\times N}$ with the jacobian of size $N \times N$. Consequently, its expectation ends up having a size of 3×1 since it is nothing but a temporal mean over N samples. The jacobian in equation 3.15 can be concisely expressed as $\begin{bmatrix} r_1\frac{\partial r_1}{\partial \mathbf{y}} & r_2\frac{\partial r_2}{\partial \mathbf{y}} & \cdots & r_N\frac{\partial r_N}{\partial \mathbf{y}} \end{bmatrix}$ where each column is the product of the derivative $\frac{\partial r_1}{\partial \mathbf{y}}$ and the scalar r_k and is of size $N \times 1$. Deriving $r_k = \mathbf{y}T_k\mathbf{y}^T$ w.r.t $\mathbf{y}$ using the product rule of differentiation,

$$\begin{aligned} \frac{\partial r_k}{\partial \mathbf{y}} &= \mathbf{y}\frac{\partial}{\partial \mathbf{y}}(T_k\mathbf{y}^T) + \mathbf{y}T_k\frac{\partial}{\partial \mathbf{y}}(\mathbf{y}^T) \\ &= \mathbf{y}\frac{\partial}{\partial \mathbf{y}}(\mathbf{y}T_k^T) + \mathbf{y}T_k \\ &= \mathbf{y}T_k^T + \mathbf{y}T_k = \mathbf{y}(T_k^T + T_k) \end{aligned} \tag{3.16}$$

where $\frac{\partial}{\partial \mathbf{y}}(T_k\mathbf{y}^T) = T_k^T$ comes from the fact that the differential of $T_k\mathbf{y}^T$, a vector, will remain the same even when it is transposed and the derivative is computed element-wise. This result is owing to the fact that T_k is not symmetric. If it were symmetric, then the result would have been $2\mathbf{y}T_k$.

For conciseness, we will represent the sum $T_k + T_k^T$ as $\mathbf{T}_k$. Finally to be consistent with our convention, using the same argument of the differential being immutable under transpositions, the row vector $\frac{\partial r_k}{\partial \mathbf{y}}$ can be transposed into a column vector and the matrix $\frac{\partial \mathbf{r}}{\partial \mathbf{y}}$ can be built as

$$\frac{\partial \mathbf{r}}{\partial \mathbf{y}} = \begin{bmatrix} r_1\mathbf{T}_1\mathbf{y}^T & \cdots & r_N\mathbf{T}_N\mathbf{y}^T \end{bmatrix} \tag{3.17}$$

giving $R'(\mathbf{w})$ in equation 3.11 as

$$R^{'}(\mathbf{w}) = -2\mathbf{x}E\left\{\begin{bmatrix} r_1\mathbf{T}_1\mathbf{y}^T & \cdots & r_N\mathbf{T}_N\mathbf{y}^T \end{bmatrix}\right\}$$

which can be further simplified to

$$R'(\mathbf{w}) = -2\mathbf{x}\begin{bmatrix} \mathbf{T}_1\mathbf{y}^T & \cdots & \mathbf{T}_N\mathbf{y}^T \end{bmatrix}\mathbf{r}^T/N \tag{3.18}$$

since the expectation is a temporal mean, the element-wise multiplication with r_k can be replaced by multiplication with the vector $\mathbf{r}^T$ which also simplifies the computation.

Next, to simplify the calculation of the second derivative of $R(\mathbf{w})$, we perform column-wise matrix multiplication in equation 3.18, omitting the scalar multiplication and division, to obtain

$$\begin{aligned} R'(\mathbf{w}) &= -\mathbf{x}\begin{bmatrix} \mathbf{T}_1 r_1\mathbf{y}^T + & \cdots & +\mathbf{T}_N r_N\mathbf{y}^T \end{bmatrix} && (3.19)\\ &= -\mathbf{x}\sum_{k=1}^{N}\mathbf{T}_k r_k\mathbf{y}^T && (3.20) \end{aligned}$$

And since differentiation and summation are interchangeable based on the sum rule, $R''(\mathbf{w})$ can be obtained by

$$\begin{aligned} R''(\mathbf{w}) &= -\mathbf{x}\sum_{k=1}^{N}\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{w}} && (3.21)\\ &= -\mathbf{x}\sum_{k=1}^{N}\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \mathbf{w}} && (3.22)\\ &= -\mathbf{x}\left(\sum_{k=1}^{\mathbf{N}}\frac{\partial(\mathbf{T_k r_k y^T})}{\partial \mathbf{y}}\right)\mathbf{x}^T && (3.23) \end{aligned}$$

The derivative of $\mathbf{T}_k r_k\mathbf{y}^T$ w.r.t $\mathbf{y}$ is then obtained by the product rule of differentiation.

$$\begin{aligned} \frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}} &= \mathbf{T}_k\frac{\partial r_k}{\partial \mathbf{y}}\mathbf{y}^T + \mathbf{T}_k r_k && (3.24)\\ &= \mathbf{T}_k\left(\frac{\partial r_k}{\partial \mathbf{y}}\mathbf{y}^T + r_k\right) && (3.25) \end{aligned}$$

which is of size $N \times N$. Consequently, the size of $R''(\mathbf{w})$ turns out to be 3×3 since the sum of $\frac{\partial(\mathbf{T}_k r_k\mathbf{y}^T)}{\partial \mathbf{y}}$ over N samples is also of size $N \times N$.

The autocorrelation function and its derivatives can now be incorporated as a priori information into existing BSS methods to formulate semi-BSE methods, the formulation of which is presented in the following chapters.

3.3/ RECAPITULATION

This chapter laid down the ground work for the coming chapters in this part. The video databases needed for rPPG measurement analysis were introduced and the corresponding experimental setup was established. We discussed the formulation of ICA and the intuition behind it so that it can improved upon in a semi-blind source extraction algorithm. We also presented autocorrelation as a periodic measurement, which can be used to affect the aforementioned improvement. With the prerequisites dealt with, we can now dive into the details of the algorithms that were developed in this thesis, which are presented in the following chapters.

4

Multi-objective Optimization using Autocorrelation and Independent Component Analysis

Semi-blind Source Separation methods are aimed to improve upon Blind Source Separation methods by incorporating information about the desired signal in the optimization function of the BSS method in question. The significance of periodicity and its presence in a large class of biomedical signals has already been discussed and a formulation of autocorrelation as a periodicity metric has been established in the previous chapter. This periodicity metric can be combined with a BSS method, specifically, the ICA objective function in the case of traditional rPPG measurement, thereby converting the BSS ICA problem to a *semi*-Blind Source Extraction (BSE) problem, since we are not entirely blind anymore with regards to the type of signal that needs to be extracted.

In this chapter, we propose to reformulate the objective function of ICA to make it a better posed problem by making two augmentations. First, we require only one component, i.e., the rPPG pulse from the mixture of the temporal traces. As a result, the problem of component *separation* can be modified into that of component *extraction*. This requirement is not uncommon and is manifested in various applications. For instance, the On-Off simulation scheme of fMRI experiments [34].

Second, we know that the blood volume pulse embedded in the RGB temporal traces is by definition periodic (or at least quasi-periodic). Consequently, we use the periodicity of the rPPG signal as an a priori information to help extract the most periodic component. To this end, we use autocorrelation as the measure of periodicity for guiding the ICA separation algorithm. To compensate for the negative values, the mean of squared autocorrelation is used as the periodicity measure. The rPPG pulse extraction is accomplished by using a multi-objective optimization approach to maximize both mean squared autocorrelation and negentropy [24], a measure of non-gaussianity fit for remote photoplethysmography.

One possibility is to use autocorrelation as a constraint to nudge the algorithm towards selecting components having periodicity higher than a given threshold. However, formulation of such a constraint is critically dependent on the threshold which is quite complex to select, especially in stochastic scenarios such as that of rPPG measurements [1]. A

[1] This argument is, in fact, the basis of augmenting our cICA algorithm with a chrominance constraint as will be discussed in chapter 5. The addition of an extra constraint simplifies the overall problem and compensates for the requirement of threshold selection.

better solution would be to use autocorrelation as one of the objective function thereby maximizing autocorrelation along with negentropy formulated as a multi-objective optimization problem. The use of autocorrelation as a periodicity measure and formulation of the multi-objective optimization problem is presented next.

Multi-objective optimization problems are generally formed by relaxing the constraints in a constrained optimization problem and interpreting them as additional objective functions. In the following section we describe the formulation of the rPPG extraction module using Multi-objective optimization with Autocorrelation as a periodicity measure and ICA, henceforth referred to as MAICA.

We reiterate the basic premise of rPPG measurement and its formulation for ICA here. The rhythmic cardiac pulse, appearing as variations in skin color, is assumed to be linearly mixed into the temporal traces of color data from cameras. The problem of rPPG measurements is posed as a signal separation problem where the rhythmic cardiac pulse is assumed to be linearly mixed into the temporal traces of color data from cameras. If the time varying color traces channels are represented as $\mathbf{x} = (x_1, x_2, ..., x_n)^T$, which is an instantaneous linear mixture of the original m independent signals denoted as $\mathbf{c} = (c_1, c_2, ..., c_m)^T$. In the case of rPPG measurement, $n = 3$ for the commonly used RGB channel images. Then the process of mixing can be formulated as $\mathbf{x} = \mathbf{Ac}$, where the matrix $\mathbf{A}_{n\times m}$ represents the linear memoryless mixing of the channels. To extract the cardiac signal, the optimum weighting matrix $\mathbf{w}^*$ has to be estimated which separates out the undesired components due to noise, illumination, motion and other sources, i.e. $\mathbf{y} = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w} \in \mathbb{R}^3$. The goal of the MAICA algorithm is then to combine periodicity information in the form of autocorrelation with the objective function of ICA in order to improve the rPPG signal estimates. Specifically, the rPPG pulse extraction is accomplished by using a multi-objective optimization approach to maximize both mean squared autocorrelation and negentropy [24], a measure of non-gaussianity fit for remote photoplethysmography, which is the one of the most commonly used objective functions of traditional ICA.

4.1/ PROPOSED METHOD

In this section, we present the formulation of the multi-objective optimization algorithm, the constituent objective functions and linear scalarization which is a simple and intuitive manner of combining multiple obective functions.

4.1.1/ OBJECTIVE FUNCTIONS

Our two objective functions correspond to negentropy and autocorrelation respectively of the output $\mathbf{y} = \mathbf{w}^T\mathbf{x}$ where the ideal orthogonal row vector $\mathbf{w} \in \mathbb{R}^3$, obtained after optimization, extracts the desired component representing the cardiac signal from the RGB temporal traces $\mathbf{x}$.

$$\begin{aligned} &\text{Maximize} && J(\mathbf{y}),\ R(\mathbf{y}) \\ &\text{Subject to} && h(\mathbf{w}) = 0 \end{aligned} \tag{4.1}$$

where $R(\mathbf{y}) = R(\mathbf{w}^T\mathbf{x})$, which is eventually a function of $\mathbf{w}$, is the mean squared autocorrelation given by

$$R(\mathbf{y}) = E\{\mathbf{r}^2\} \tag{4.2}$$

with $\mathbf{r} = [r_1, r_2, \cdots, r_N]$ being the autocorrelation and r_k is given by equation 3.6. The negentropy, $J(\mathbf{y})$ is the generic contrast function for ICA defined by [26] as $H(\mathbf{y}_{gauss}) - H(\mathbf{y})$. $H(.)$ is the differential entropy and $\mathbf{y}_{gauss}$ is a random variable with a variance equal to that of the output signal $\mathbf{y}$. In practice, an approximation of negentropy is used for ease of computation and flexibility given by

$$J(\mathbf{y}) \approx \rho[E\{G(\mathbf{y})\} - E\{G(v)\}]^2 \tag{4.3}$$

where ρ is a positive constant, v is a Gaussian variable having zero mean and unit variance. G(.) can be any non-quadratic function as suggested by [24]. A good general purpose function is given by

$$G(y) = \frac{\log\cosh(ay)}{a} \tag{4.4}$$

with $1 \leq a \leq 2$. Finally, as suggested by [34] the constraint, $h(\mathbf{w}) = E\{\mathbf{y}^2\} - 1 = 0$ was introduced to make sure that objective functions $J(\mathbf{y})$ and $R(\mathbf{y})$ and the weighting vector $\mathbf{w}$ are bounded.

4.1.2/ Linear Scalarization

A simple way to incorporate the a priori information in the optimization problem is to scalarize the multi-objective optimization by forming a single-objective optimization such that the solutions to the scalarized problem are the set of feasible solutions commonly known as Pareto optimal solutions [51]. The linearly scalarized version of the multi-objective contrast function then becomes

$$\begin{aligned} &\text{Maximize} \quad \overline{J}(\mathbf{y}) + \overline{R}(\mathbf{y}) \\ &\text{Subject to} \quad h(\mathbf{w}) = 0 \end{aligned} \tag{4.5}$$

where

$$\overline{J}(\mathbf{y}) = \frac{J(\mathbf{y}) - J_{min}}{J_{max} - J_{min}} = s_1(J(\mathbf{y}) - J_{min}) \tag{4.6}$$

and

$$\overline{R}(\mathbf{y}) = \frac{R(\mathbf{y}) - R_{min}}{R_{max} - R_{min}} = s_2(R(\mathbf{y}) - R_{min}) \tag{4.7}$$

are the normalized versions of the respective objective functions in order to compensate for the disparities in scale and $s_1 = \frac{1}{J_{max} - J_{min}}$ and $s_2 = \frac{1}{R_{max} - R_{min}}$. Ideally, the boundary values of the objective functions correspond to their global maximum and minimum values. However, to emulate a live scenario as much as possible, all the processing was performed over a temporal window of 30 seconds. Using the actual boundary values in this case would not conform to the emulation.

Consequently, J_{max} and R_{max} were calculated by using a sinusoidal signal emulating an ideal blood volume pulse, $y_s = \sin(t)$ where t corresponds to the time coordinates of the current temporal window. This was done by taking the maximum values of the objective functions over the frequency range of human heart rates, $\mathcal{F} \in [.7, 3]$ Hz. The minimum values, J_{min} and R_{min}, were calculated in the same manner for a uniform random signal y_r for the temporal window t.

$$J_{max} = \max_{\mathcal{F}} J(y_s),\ R_{max} = \max_{\mathcal{F}} R(y_s) \tag{4.8}$$

$$J_{min} = \min_{\mathcal{F}} J(y_r),\ R_{min} = \min_{\mathcal{F}} R(y_r) \tag{4.9}$$

4.1.3/ LAGRANGE MULTIPLIERS AND MULTI-OBJECTIVE OPTIMIZATION

Finally, the entire problem was expressed as a Lagrange multiplier [11] formulation where a Newton-method like approach was used for optimization. The augmented Lagrangian method was used because of its wider applicability and robustness against the equality constraints owing to the penalty term that punishes violations to those constraints, making it more stable than the classical method of Lagrange multipliers. The Augmented Lagrangian for equation 4.5 is given as

$$\mathcal{L} = [\overline{J}(\mathbf{y}) + \overline{R}(\mathbf{y})] - \lambda h(\mathbf{w}) + \frac{1}{2}\gamma \|h(\mathbf{w})\|^2 \tag{4.10}$$

where λ is the lagrange multiplier for the constraint $h(\mathbf{w})$, $\overline{J}(\mathbf{y})$ and $\overline{R}(\mathbf{y})$ are given by equations 4.6 and 4.7 respectively. $\|.\|^2$ denotes the Euclidean norm and the term $\frac{1}{2}\gamma\|.\|^2$ is the penalty term that makes sure that the optimization problem is held at the condition of local convexity assumption: $\nabla^2_{ww}\mathcal{L} > 0$ which ensures continuity and the consequent existence of an optimal $\mathbf{w}$. To find the maximum of $\mathcal{L}$ in equation 4.10 a Newton-like learning method was used to iteratively adapt $\mathbf{w}$

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta(\mathcal{L}^{''}_{\mathbf{w}_k})^{-1}\mathcal{L}^{'}_{\mathbf{w}_k} \tag{4.11}$$

where k is the iteration index, η is the positive learning rate to avoid uncertainty in convergence and $\mathcal{L}^{'}_{\mathbf{w}_k}$ is the first derivative of $\mathcal{L}$ at step k w.r.t $\mathbf{w}$ given by

$$\mathcal{L}^{'}_{\mathbf{w}_k} = s_1\bar{\rho}E\{\mathbf{x}G^{'}_y(\mathbf{y})\} + s_2E\{R^{'}_{\mathbf{w}}(\mathbf{w})\} - \lambda E\{\mathbf{xy}\} \tag{4.12}$$

where the sign of $E\{G(\mathbf{y})\} - E\{G(v)\}$ gives the value of $\bar{\rho} = \pm\rho$, $G^{'}_y(\mathbf{y})$ and $R^{'}_{\mathbf{w}}(\mathbf{w})$ are the first derivatives of $G(\mathbf{y})$ and $R(\mathbf{w})$ w.r.t $\mathbf{y}$ and $\mathbf{w}$ respectively. The Hessian $\mathcal{L}^{''}_{\mathbf{w}_k}$ in equation 4.11, is calculated as

$$\mathcal{L}^{''}_{\mathbf{w}_k} = s_1\bar{\rho}\mathbf{R}_{\mathbf{xx}}E\{G^{''}_y(\mathbf{y})\} + s_2E\{R^{''}_{\mathbf{w}}(\mathbf{w})\} - \lambda\mathbf{R}_{\mathbf{xx}} \tag{4.13}$$

the inversion of which is not problematic because $\mathbf{R}_{\mathbf{xx}}$ being the covariance matrix of the whitened and centered signal $\mathbf{x}$ is an identity matrix. $G^{''}_y(\mathbf{y})$ and $R^{''}_{\mathbf{w}}(\mathbf{w})$ are second order derivatives and $\mathcal{L}^{''}_{\mathbf{w}_k}$ is of size 3×3. The first and second derivatives of $R(\mathbf{w})$ are not trivial and are presented in the appendix. The optimum multiplier λ^* is obtained iteratively based on a gradient-ascent method [34]:

$$\lambda_{k+1} \quad = \quad \lambda_k + \gamma h(\mathbf{w}_k) \tag{4.14}$$

Following the above equations, the optimization procedure converges to the optimum point defined by the doublet $(\mathbf{w}^*, \lambda^*)$ representing the tuned parameter and final weighting matrix $\mathbf{w}^*$ which is then used to obtain the final rPPG signal. Results and analysis of the performance of MAICA are discussed in section 4.3. Next, we present an augmentation to the MAICA method in the form of an adaptive step size that serves to improve the optimization process, specifically by improving convergence times.

4.1.4/ PARAMETER FREE MAICA

Most BSS algorithms use a fixed step-size parameter. However, use of a fixed step-size creates several problems. To maximize the objective function rapidly, the step-size should

be set to a large value when the objective function is small. To get a precise weighting matrix $\mathbf{w}$, on the contrary, the step-size should be set to a small value. In (4.11), the step-size selection is even more critical because it defines the relative importance between negentropy and autocorrelation in the update procedure. Several adaptive step-size have been proposed in the literature based, for instance, on an exact line search of μ [43] or solving the optimum step-size that minimizes the cost function with an iterative process based on a gradient method [20]. In this method, we use the adaptive step-size proposed by Nakajima *et al.* [48] where the step-size is set anti-proportional to the first gradient of the cost function. The adaptive step-size method are formulated as follows

$$\mu_1' = \frac{J(\mathbf{y})}{2\left\|J'_{\mathbf{w}_k(\mathbf{y})}\right\|^2} \quad \text{and} \quad \mu_2' = \frac{R(\mathbf{y})}{2\left\|R'_{\mathbf{w}_k(\mathbf{y})}\right\|^2} \tag{4.15}$$

where $\|.\|^2$ means the Frobenius norm. The entire problem is then solved using a first order iterative optimization algorithm procedure with the objective function

$$\mathcal{L} = \overline{J}(\mathbf{y}) + \overline{R}(\mathbf{y}) \tag{4.16}$$

instead of the Augmented Lagrangian method for the classical Multi-objective optimization formulation in equation 4.10. To find the maximum of $\mathcal{L}$ in Eq. 4.16, gradient-ascent iteratively adapts $\mathbf{w}$ with

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_1 J'_{\mathbf{w}_k}(\mathbf{y}) + \mu_2 R'_{\mathbf{w}_k}(\mathbf{y}) \tag{4.17}$$

where k is the iteration index, μ_1 and μ_2 are the step-size for negentropy and autocorrelation part of the objective function, $J'_{\mathbf{w}_k}(\mathbf{y})$ and $R'_{\mathbf{w}_k}(\mathbf{y})$ are the first derivatives of $\overline{J}(\mathbf{y})$ and $\overline{R}(\mathbf{y})$ at step k w.r.t $\mathbf{w}$.

Even if the complete procedure is almost entirely free from any manual parameter adjustment, we experimentally observe that a threshold on the value of μ_1 can be useful because the derivative of the negentropy is sometimes quite unstable. Next, the details of the system framework required for signal acquisition and the implementation details of the MAICA method and its parameter free variant are presented in section 4.2.

4.2/ SYSTEM FRAMEWORK

The workflow of the procedure as depicted in figure 4.1 is presented here. Temporal RGB traces, $\mathbf{x} = [x_1, x_2, x_3]^T$ where each x_m, $m \in [1...3]$, corresponds to a temporal trace of size N of each channel, were generated by frame-wise spatial averaging of the pixels (face-cropped or skin-segmented). To this end, face detection and tracking was first performed using the Viola-Jones and the Kanade-Lucas-Tomasi implementations provided by the computer vision toolbox of MATLAB. Corner detection in the detected face was performed for tracking to crop the face based on facial landmarks. Skin detection as formulated by Conaire *et al.* [36] was then performed to select the candidate pixels which were then spatially averaged to obtain a triplet of RGB values per frame and concatenated to obtain the RGB temporal traces.

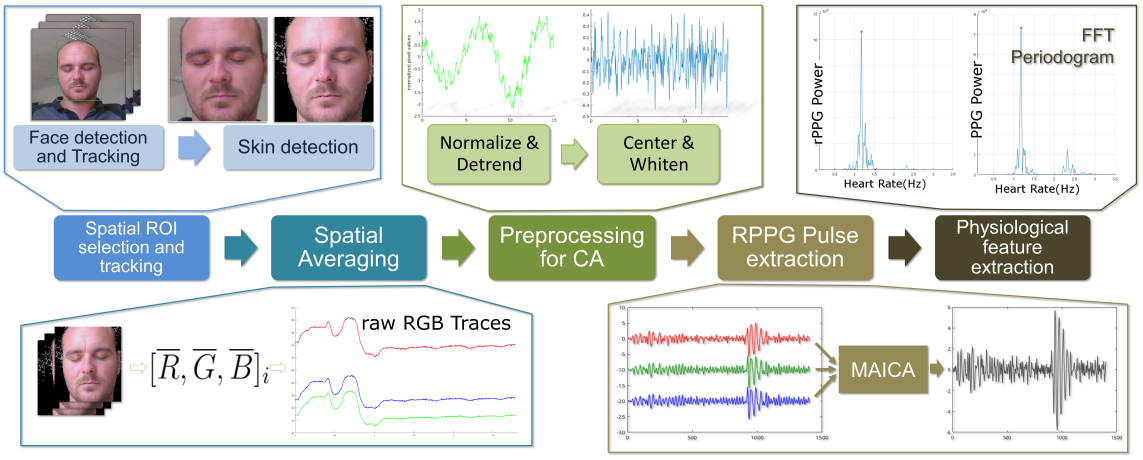


Figure 4.1: System framework for MAICA and its parameter free variant

These temporal RGB traces were then detrended using a smoothness priors approach, with the regularization parameter λ set to 500, proposed by Karjalainen and Pasi [29] to remove low frequency trends in the signal. Next, after normalization, two additional preprocessing steps are generally recommended for ICA to simplify calculations. First, *centering* was performed so that the obtained signal $\mathbf{y}$ in $\mathbf{y} = \mathbf{W}\mathbf{x}$ is zero-mean. Next, *whitening* was performed to ensure that the components were uncorrelated and their variances equal to unity. The traces were then passed to the rPPG extraction module where the MAICA algorithm was used to extract the rPPG signal.

After the rPPG signal was obtained, the per window heart rate was calculated from the highest peak of the FFT filtered within the acceptable range of heart rate $\mathcal{F} \in [.7, 3]$ Hz over a 30 second moving window using a step size of 0.5 second for our in-house datasets. Although, ICA is known to work much better with signals of longer duration, as mentioned earlier, all the processing was performed over a 30 second window, using the weighting matrix $\mathbf{w}_k$ obtained at window k as an initial estimate for calculation of $\mathbf{w}_{k+1}$ at the next window. This 30 second window size was chosen as a trade-off between speed and availability of enough data for convergence. On the other hand, a 15 second window was needed for the MMSE-HR dataset owing to the shorter length of many constituent videos. The window-wise heart rate estimations were then smoothed using a Kalman filter.

The Kalman filter helped to remove spurious outliers resulting from abrupt variations in illumination and/or motion. Since we only want to filter out spurious outliers, irrespective of the method being used, the filter was periodically reconfigured if the change in HR was beyond a threshold of 6 bpm. This ensured that the filter followed the measured HR values as closely as possible, while simultaneously avoiding outliers. These values were fixed for the analysis of all the videos in the three different datasets that we used. They were chosen to represent a margin of 1 BPM in the motion and measurement noise models and were verified by trial and error in their ability to discard spurious outliers.

We present the experiments and results next in section 4.3

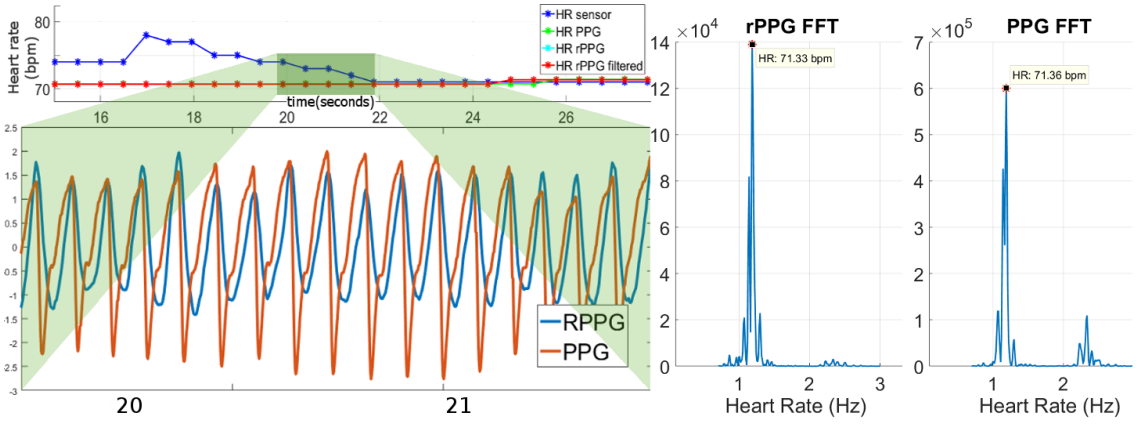


Figure 4.2: Extracted rPPG signal for video 10 of the SIMPLE dataset of the UBFC-RPPG database. HR_sensor is HR from the finger sensor. HR_rPPG_filtered is HR_rPPG after Kalman filtering

4.3/ RESULTS AND ANALYSIS

As was mentioned in section 3.1 in chapter 3, the UBFC-RPPG and MMSE-HR video databases were used to analyze the algorithms developed. The parameter free variant of MAICA was analyzed against the UBFC-RPPG database, to specifically assess the performance gain achieved. But first, it is worth looking at a typical rPPG signal and its FFT and some preliminary analysis to establish the feasibility of the proposed method.

Figure 4.2 shows the rPPG and PPG signals from a fairly simple video and their heart rate estimations, HR_{rPPG} vs HR_{PPG} respectively, using the Fast Fourier Transform (FFT). The other two quantities HR_{sensor} and $HR_{rPPG_{filtered}}$ depict the HR from the pulse oximeter and HR_{PPG} after Kalman filtering respectively. The high correlation between the rPPG and PPG signal is clearly visible in the figure, and so is the correlation between HR_{rPPG} and HR_{PPG}, which in fact are barely visible because of the overlapping values of HR_{sensor} and $HR_{rPPG_{filtered}}$, owing to the video being a simple one.

To gain an intuition on the advantage of incorporating periodicity information into the ICA source separation algorithm, a quick analysis of one of the difficult videos, where traditional ICA is not entirely successful, from our UBFC-RPPG database is presented here. The analysis was carried over sequential temporal windows of 30s and the corresponding weighting matrices that extract the rPPG signal of ICA in figure 4.3a are compared with those obtained by MAICA in figure 4.3c. It can be seen from figure 4.3d that the weighting matrices that simultaneously maximizes negentropy and periodicity indeed result in a heart rate closer to the ground truth heart rate. Moreover, the SNR obtained for the measured signal was improved from -10.89 for ICA to -3.39 for MAICA (also seen by the overall increase in SNR in table 4.1), illustrating the advantage of incorporating the periodicity information in the algorithm.

Having looked at the feasibility analysis for MAICA, its performance comparison against traditional ICA can be examined. Figure 4.4 shows the correlation comparisons between ICA [49] and MAICA for the two databases where HR_{PPG} and HR_{rPPG} are plotted against each other. The metrics PRECIS 2.5 and PRECIS 5 show the percentage of windows where $\delta = |HR_{rPPG} - HR_{PPG}| < 2.5$ and 5 beats per minute (bpm) respectively. MAE corresponds to the average mean absolute error between HR_{rPPG} and HR_{PPG} in

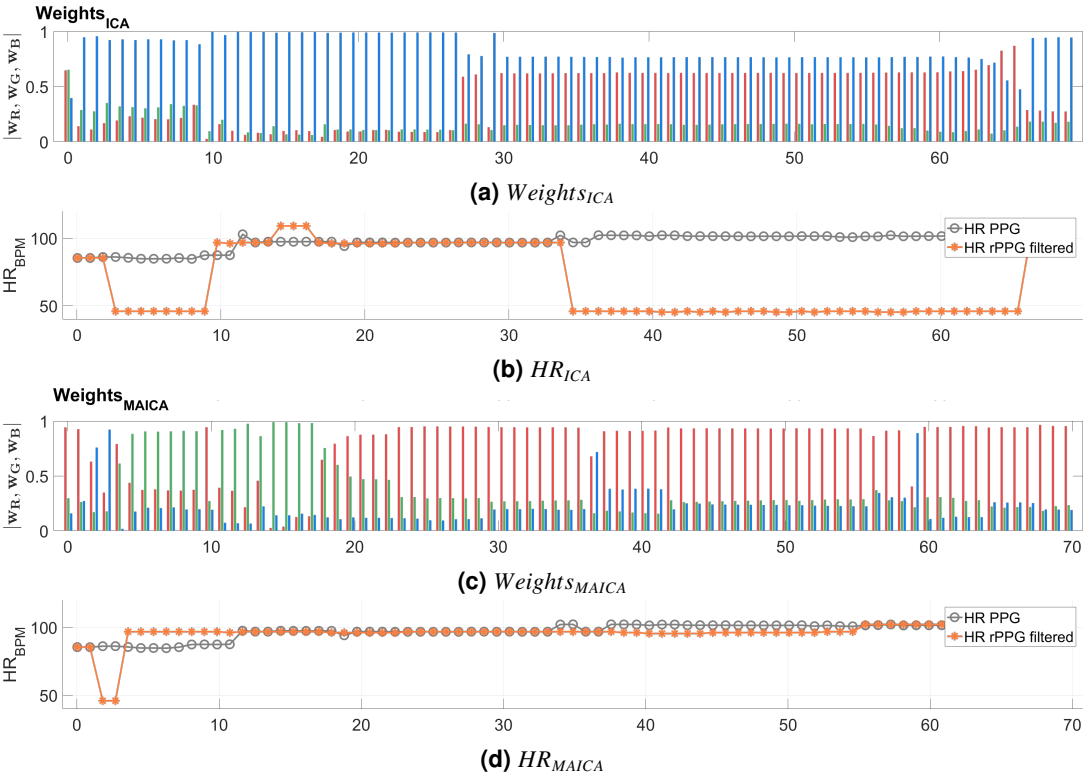


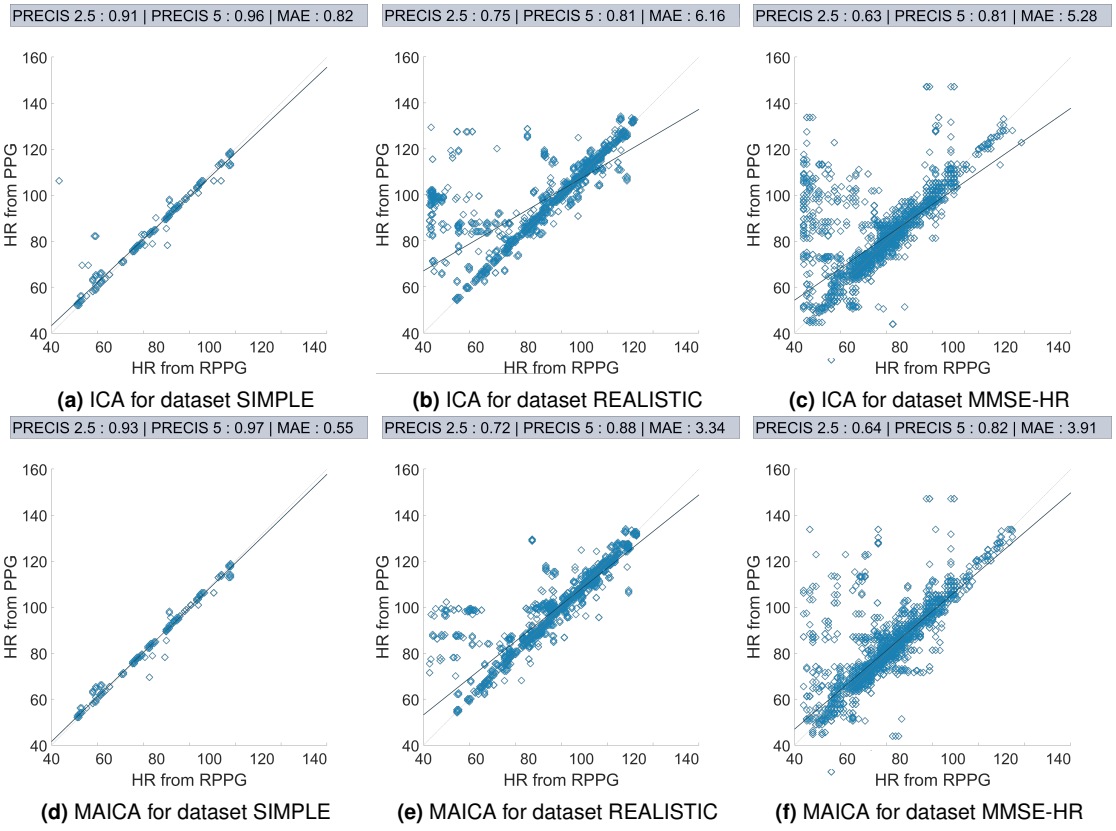
Figure 4.3: Window-wise weight analysis showing ICA weights and HR in (a) and (b) vs MAICA weights and HR in (c) and (d). For the exact same data, different values of $\mathbf{w} = |w_R, w_G, w_B|$ corresponding to the RGB channels can extract an accurate rPPG signal. Absolute values of the weights are shown for concise display.

bpm calculated over all the windows for all videos.

As the name suggests, the analysis of the SIMPLE dataset, which can be thought of emulating a patient at rest, was not challenging and both ICA and MAICA performed well. However, MAICA did result in fewer outliers than ICA. The REALISTIC and the MMSE-HR datasets were understandably more challenging. However, MAICA demonstrated a better correlation between HR_{PPG} and HR_{rPPG}. It is also evident that the lesser number of outliers were a direct cause of better convergence towards the correct rPPG signal. The MAICA method also provided stronger peaks in the FFT periodogram along with better HR estimates. It also provides a better performance than the Self-Adaptive-Matrix Completion method introduced in [81] whose MAE is 7.6 bpm as compared to our 5.2 bpm. However, this comparison is not very clear since we do not use the same ROI selection method as them. It is also noteworthy that the amount of candidate pixels that are used to generate our rPPG signal is larger in comparison to their's owing to the use of skin detection, which might result in lesser noise and an eventually better estimation. Furthermore, as already been mentioned, incorporation of a sophisticated ROI selection method is equally possible and might even improve the estimations further.

4.3.1/ Performance analysis of MAICA

Table 4.1 shows the accuracy comparisons between ICA and MAICA and other state of the art methods, viz., PCA [53], Green [42], CHROM [58], POS [83], and G-R [39]. The



(a) ICA for dataset SIMPLE **(b)** ICA for dataset REALISTIC **(c)** ICA for dataset MMSE-HR

(d) MAICA for dataset SIMPLE **(e)** MAICA for dataset REALISTIC **(f)** MAICA for dataset MMSE-HR

Figure 4.4: Correlation Analysis for rPPG comparing ICA and MAICA for the three datasets

metrics used apart from MAE are signal-to-noise ratio (SNR) and Pearson's correlation coefficient (r) between heart rate calculated using the rPPG signal, HR_{rPPG} and the heart rate calculated using the ground truth PPG waveform, HR_{PPG}. The windowed method is computationally more taxing, owing to the smaller window length, but is more realistic. The SNR (dB) was calculated as the ratio of the power of the main pulsatile component of the PPG to that of the background noise to accommodate the wide dynamic range of the signals. However, it is to be noted that the MMSE-HR database does not provide the ground truth waveforms, thereby obliging the use of the main pulsatile component of the RPPG instead of the PPG for the SNR calculation. As a result, the SNR values for the MMSE-HR database are not really relevant and are omitted in table 4.1. It is worth mentioning here that although the MAICA is not faster than ICA since the autocorrelation objective function has to be additionally calculated, its novelty lies in the combination of the periodicity information with the notion of independence used by the ICA algorithm.

In literature, there are several works which uses ICA for rPPG extraction. However, the core algorithm for it remains the same. The ICA implementation used for our analysis has been adapted from FastICA [26]. Furthermore, the analysis of all the methods was performed using exactly the same pre and post processing steps like normalization, filtering and smoothing. The exact metrics of ICA from related state of the art methods such as [49], [54] and [64] could not be used because they all use their own private databases which were inaccessible to the public. However, as mentioned earlier, the core algorithm of ICA remains the same making the metrics in Table 5.1 applicable. Furthermore, comparison with smart ROI selection methods such as [76], [63], [70] and [73] was not deemed relevant in order to limit the comparison amongst source separation methods,

	UBFC-RPPG						MMSE-HR	
	SIMPLE			REALISTIC				
	MAE	SNR	r	MAE	SNR	r	MAE	r
MAICA	**0.55**	2.98	**0.99**	**3.34**	**-0.26**	**0.89**	**3.91**	**0.86**
ICA	0.82	2.73	0.98	6.16	-1.23	0.76	5.28	0.70
PCA	2.04	-1.43	0.97	9.65	-3.45	0.67	9.15	0.49
Green	9.86	-1.61	0.29	7.73	-2.78	0.68	10.65	0.47
CHROM	0.72	**3.04**	**0.99**	3.81	-0.93	0.87	5.59	0.83
POS	0.67	2.57	**0.99**	4.73	-1.60	0.80	5.77	0.82
G-R	0.67	1.97	**0.99**	9.79	-3.10	0.65	8.56	0.58

Table 4.1: Performance comparisons between the various methods using Mean Absolute Error (MAE), Signal-to-Noise Ratio (SNR) and Pearson's correlation coefficient (r)

and the fact that our method can easily be incorporated into an ROI selection framework.

Regarding the performance metrics, MAICA and ICA both yield good rPPG signals for the SIMPLE dataset. However, MAICA did remove the few outliers that are present in the case of ICA. This is reflected in the low MAE and high SNR values as shown in table 4.1. This can be attributed to the fact that the subjects were generally relaxing, mostly with their eyes closed, which resulted in minimal motion artifacts. On the other hand, the REALISTIC dataset was more challenging since the subjects were actually working on the computer and were only requested to keep their hand still for the PPG sensor. Similarly, the MMSE-HR database was challenging owing to it being an emotion elicitation database. There were many instances where the subjects laughed out loud and exhibited considerable movements. This resulted in the usual problems arising from movement of the subjects. Consequently, the presence of outliers was more pronounced for both the REALISTIC dataset and the MMSE-HR database, which MAICA was able to reduce. This can be seen in table 4.1, where the pearson correlation coefficient was also closer to unity which is also evident from figure 4.4 where the fitting line was closer to the 45° line as compared to ICA.

Finally, in figure 4.5 we present the window-wise heart rate comparisons between ICA and MAICA, notably for videos where ICA is not entirely successful. The combination of the autocorrelation function with the ICA negentropy function in the premise of multi-objective optimization enables the extraction of a more accurate rPPG signal closer to the ground truth. Videos from the UBFC-RPPG SIMPLE dataset are not presented in this figure since there was not much difference between the performance between ICA and MAICA owing to the simplicity of these videos. With regards to the MMSE-HR dataset, it actually proved to be useful in highlighting the versatility of our method in presence of motion artifacts arising from facial expressions, as is visible in the snapshots in the figure. It is to be noted that almost all the videos of the MMSE-HR dataset manifested changing facial expressions and moving faces.

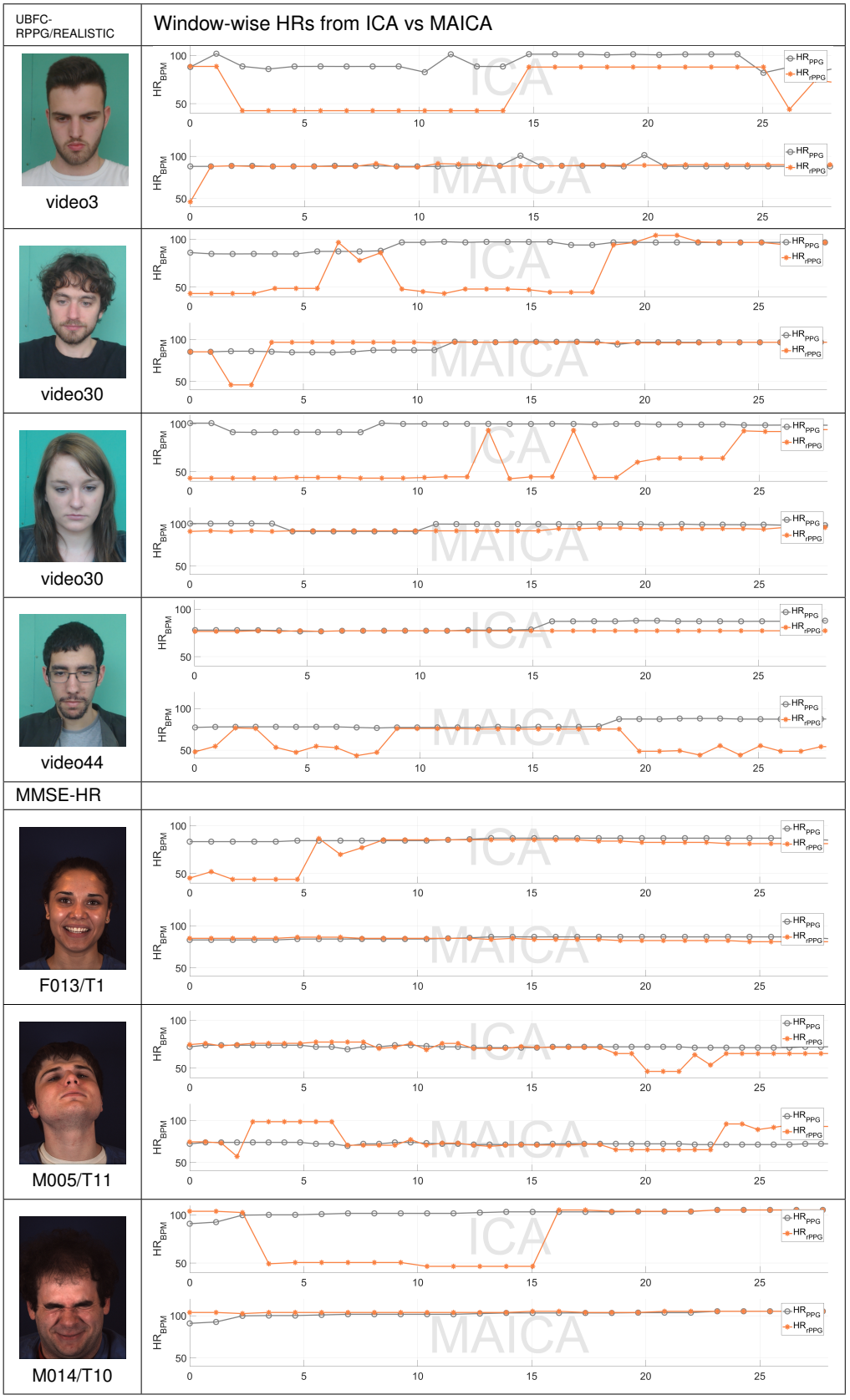


Figure 4.5: ICA vs MAICA for certain videos from the UBFC-RPPG\REALISTIC and MMSE-HR datasets. The FFT periodogram was used to perform the HR measurements over temporal windows of 30s and 15s respectively for the two datasets owing to the shorter lengths of videos in the latter dataset.

	MAE	SNR	r
MAICA	**3.18**	**0.06**	0.90
MAICAfix	3.34	-0.26	0.89
ICA	6.02	-1.11	0.79
PCA	9.65	-3.45	0.67
Green	7.73	-2.78	0.68
CHROM	3.81	-0.93	0.87
POS	4.73	-1.60	0.80
G-R	9.79	-3.10	0.65

Table 4.2: Performance comparisons of parameter-free MAICA and other methods using Mean Absolute Error (MAE), Signal-to-Noise Ratio (SNR) and Pearson's correlation coefficient (r)

4.3.2/ Performance analysis of parameter-free MAICA

We present here the performance comparison of the parameter free version of MAICA with adaptive step size. Since the major part of the method is same as the original MAICA, we present here a concise analysis only performed with the UBFC-RPPG database, especially to verify the feasibility and performance improvement possible with the an adaptive instead of a fixed step size for the optimization algorithm. Table 4.2 shows the accuracy comparisons between MAICA with adaptive step-size, with a fixed step-size (referred as MAICAfix here), ICA[49] and other state of the art methods, viz., PCA [53], Green [42], CHROM [58], POS [83], and G-R [39]. The metrics used are Mean Absolute Error (MAE), signal-to-noise ratio (SNR) and Pearson's correlation coefficient (r) between heart rate calculated using the rPPG signal and the heart rate calculated using the ground truth PPG waveform [76]. The SNR (dB) was calculated as the ratio of the power of the main pulsatile component of the PPG to that of the background noise to accommodate the wide dynamic range of the signals.

Not ignoring the fact that the parameter-free MAICA gives the highest SNR of all the methods, a value of 0.06 is still low. Even so, it is interesting to observe that the MAE values are acceptable for most state-of-the-art methods. The CHROM method [58] is undoubtedly the most reliable rPPG techniques from the literature because it systematically outperforms other methods. Finally, the results obtained with the adaptive step-size are only slightly better than the ones with a fixed step-size, but the proposed method has the undeniable advantage of not having to empirically fix the optimal value of the threshold. In this respect, this advantage makes the parameter-free MAICA method a notch better than the regular MAICA and the cICA method which is presented in the next chapter, one of limitations of which is the selection of appropriate thresholds.

4.4/ Summary

In this chapter we presented a novel semi blind source extraction method, MAICA, for the application of rPPG measurements using multi-objective optimization with mean squared autocorrelation and negentropy as the objective functions. The method provides better results than other state of the art methods while removing the extra step for choosing the best component. The periodogram of the extracted signals was also consistently closer

to that of the PPG. Our method can also be combined with other methods like smart ROI selection [63] to further obtain better rPPG estimations. The two methods, regular MAICA and parameter-free MAICA presented in this chapter and the corresponding results obtained have been published and validated in the Biomedical Signal Processing and Control journal [a], and in the IEEE BioHealth and Engineering conference [i].

As has already been discussed, another alternative is to incorporate the periodicity information as a constraint in the optimization scheme instead of using a separated objective function. Furthermore, incorporation of constraints based on the physical properties of the skin, which affects the scattered light thereof in a unique manner, can lead to interesting results. We next examine this alternate solution of incorporating periodicity and chrominance information as a constraint instead of an objective function in chapter 5.

5

Constrained Independent Component Analysis

In the previous chapter, we looked at the MAICA algorithm, a Semi-blind Source Extraction (semi-BSE) method, that was used to incorporate this a priori information in the form of autocorrelation combined with the objective function of ICA in a multi-objective optimization scheme. It is typical of a multi-objective optimization algorithm to return a set of optimas corresponding to a given requirement. This issue was treated by using the simple technique of linear scalarization in the previous chapter.

It would be interesting to approach the problem of a priori information incorporation from a different angle, viz. using the a priori information as a constraint rather than an objective function. This alternative undoubtedly is not free from its own complications, such as the choice of a threshold for the constraint, which is critical. We alleviate this issue by incorporating an extra constraint based on the physical properties of the skin, in order that the solution space be easier to converge onto, thereby mitigating, in part, the issue of threshold tuning. This improves the convergence of the optimizer as compared to using only one for the constraints, enabling the estimation of a more accurate rPPG signal. Consequently, we use the constrained ICA framework to combine constraints based on periodicity and chrominance with the objective function of ICA to make it a better posed problem.

The technique of Constrained ICA can be used to incorporate these additional requirements and a priori information in the form of constraints. One approach of cICA, termed as ICA with reference, comprises of using additional knowledge related to the sources and desired signals constraints in the form of a reference signal [32]. However, employing a sufficiently accurate reference signal is prone to the evident complication of choosing its right frequency. This can be done in two possible ways. One alternative is to repeatedly compare the extracted rPPG signal to reference signals of different frequencies [80], as discussed in chapter 2. The other alternative is to update the frequency of the reference signal continuously, in effect making it a parameter to optimize. This increases the complexity of the problem and reduces the probability of convergence. A PPG signal is a very apt reference for rPPG extraction whose synthesis depends critically on the required frequency, even more so than the actual shape of the signal.

To avoid this limitation, we use autocorrelation as the a priori information for guiding the cICA separation algorithm which then chooses the most periodic component representing the blood volume pulse. To further aid the convergence, we apply chrominance-based constraints based on the standardized skin tone as used by De Haan *et al.* [58].

The use of two constraints increases the probability of convergence towards the best rPPG signal. In this chapter we develop a new semi-blind source extraction method, using constraints based on periodicity and physical properties of the skin. Next, we present the proposed method and explain in detail the two constraints used.

5.1/ PROPOSED METHOD

Let us recall the basic formulation of ICA where the required signal vector $\mathbf{y}$ of size N is extracted from the RGB temporal traces matrix $\mathbf{x}$ of size $3 \times N$ using a weighting matrix $\mathbf{w}$ of size 3×1 according to $\mathbf{y} = \mathbf{w}^T\mathbf{x}$. As mentioned earlier, we aim to perform component extraction instead of separation, which is reflected in the change in the size of the weighting matrix from 3×3 in basic ICA to 3×1 for cICA. The cICA algorithm aims to optimize for the best weighting matrix $\mathbf{w}^*$, which maximizes the objective function while satisfying the imposed constraints.

Typical biomedical signals like ECG and PPG signals are known to be periodic or semi-periodic. This implicit property of periodicity of biomedical signals can be exploited to guide the component extraction process in converging to the component with the highest periodicity as has been already established in the previous chapter with the MAICA algorithm. Accordingly, we use autocorrelation as one of the constraints to nudge the algorithm towards selecting components having periodicity higher than a given threshold. Furthermore, we also incorporate constraints based on the physical properties of the skin, which directly effect the light reflected therefrom, and consequently the color or *chrominance* information therein. This constraint is developed using the CHROM method which is undoubtedly one of the most reliable techniques in literature [58]. These two constraints are next described in detail.

5.1.1/ CHROMINANCE BASED CONSTRAINT

Although autocorrelation does help the optimizer to converge to a weighting matrix that extracts the correct component for simple videos, in a more realistic scenario it is prone to having not so well defined maxima. Additionally, in fitness scenarios with repetitive movements, the assumption that the most periodic component being the rPPG signal is perturbed by the periodic motion component of the fitness activity. This calls for the use of another constraint, which is not fundamentally affected by periodic components, to aid the convergence for which the CHROM constraint [58] is a suitable candidate.

To confirm this requirement, and to correlate the autocorrelation with the weights, the mean squared autocorrelation was plotted against all the possible orthonormal weights, (w_R, w_G, w_B) whose components represent the contribution of each of the RGB channels in forming the rPPG signal. All these possible weight vectors in $\mathbb{R}^3$ are spanned by the standard basis $w_1 = (1, 0, 0), w_2 = (0, 1, 0), w_3 = (0, 0, 1)$, i.e.,

$$(w_R, w_G, w_B) = w_1 + w_2 + w_3 \tag{5.1}$$

A temporal section of the RGB traces, 30 seconds long, $\mathbf{x}_t$, was used to perform this weights analysis. To plot the autocorrelation as a function of the vector space (w_R, w_G, w_B), each vector w was scaled by the corresponding mean squared autocorrelation of $\mathbf{y}_t = w\mathbf{x}_t$ to obtain the plot in figure 5.1. This plot can be thought of as a deformation of the

unit sphere, owing to the orthonormality of the weghting vectors, by the mean squared autocorrelation of $\mathbf{y}_f$. The ideal weighting vector giving the maximum autocorrelation is depicted as $\mathbf{w}_{\mathbf{autocorr}}$ and the CHROM weighing vector is depicted as $\mathbf{w}_{\mathbf{chrom}}$.

It can be seen from figure 5.1 that the mean squared autocorrelation is symmetric with respect to $\mathbf{w}$, i.e., there will always be a dual of a given weighting vector which would give the same autocorrelation. This is not necessarily a disadvantage since the optimizer will converge irrespective of the initial direction chosen. However, it is also visible that the plot is not very *peaky* i.e. the maximum autocorrelation is localized to an area of smaller slope, which might make the convergence slower towards the end. It is also noteworthy that the weighting matrix corresponding to the maximum autocorrelation, $\mathbf{w}_{\mathbf{autocorr}}$ and the CHROM weighting matrix, $\mathbf{w}_{\mathbf{chrom}}$ depicted as vectors point around the same vicinity. In other words, the rPPG signal extracted using the CHROM method also exhibits high autocorrelation.

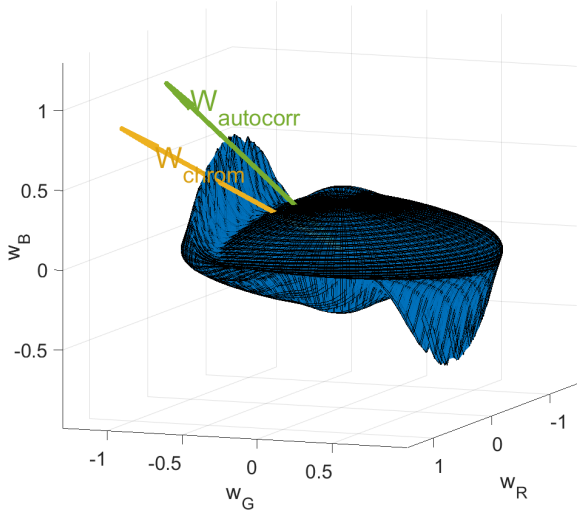


Figure 5.1: Mean-squared Autocorrelation, $E\{\mathbf{r}^2\}$ vs the weighting matrix, $\mathbf{w}$

On the other hand, chrominance-based methods tend to be restrictive in choosing the weighting matrix based on their linear formulation. According to [58], a chrominance signal which incorporates the maximum photoplethysmographic information is obtained using a standardized skin tone resulting in an algorithm that can work correctly regardless the color of the illuminant. The CHROM signal is given by

$$S = X_f - \alpha Y_f \tag{5.2}$$

where $X_f = 3R_f - 2G_f$ and $Y_f = 1.5R_f + G_f - 1.5B_f$ are the projections of the RGB traces on to standardized skin tone space and $\mathbf{x} = [R_f, G_f, B_f]^T$ is the bandpass filtered temporal RGB trace of size $3 \times N$. α is the ratio between the standard deviations of X_f and Y_f giving

$$S = 3(1 - \frac{\alpha}{2})R_f - 2(1 + \frac{\alpha}{2})G_f + 3\frac{\alpha}{2}B_f = \mathbf{w}_{\mathbf{chrom}}\mathbf{x} \tag{5.3}$$

where $\mathbf{w}_{\mathbf{chrom}}$ is the CHROM weighting matrix. The goal of the proposed rPPG extraction algorithm is to converge towards the weighting matrix that simultaneously gives a component of high periodicity and is within the vicinity of the weights $\mathbf{w}_{\mathbf{chrom}}$ up to a certain threshold. An analysis of the effect of the combination of these two constraints is presented in the next section.

5.1.2/ COMBINATION OF PERIODICITY AND CHROMINANCE BASED CONSTRAINTS

To assess the effect of the combination of periodicity and chrominance constraints on the cICA algorithm, and its utility in fitness scenarios, three videos were recorded on a fitness bike. They are categorized as LIGHT, MODERATE and INTENSIVE based on the speed of motion and intensity of training. Figure 5.2 shows a snapshot from the three videos which exhibit a prominent periodic motion along with accuracy comparisons between recovered rPPG from CHROM, cICA using only the periodicity constraint and cICA using both the periodicity and the chrominance constraint. The CHROM method was able to extract the correct rPPG signal for the LIGHT video (figure 5.2d) whereas cICA with only the periodicity constraint converged to the component representing the strong periodic motion (figure 5.2g). As expected, the combination of the CHROM and periodicity constraints resulted in convergence to the correct rPPG signal (figure 5.2j).

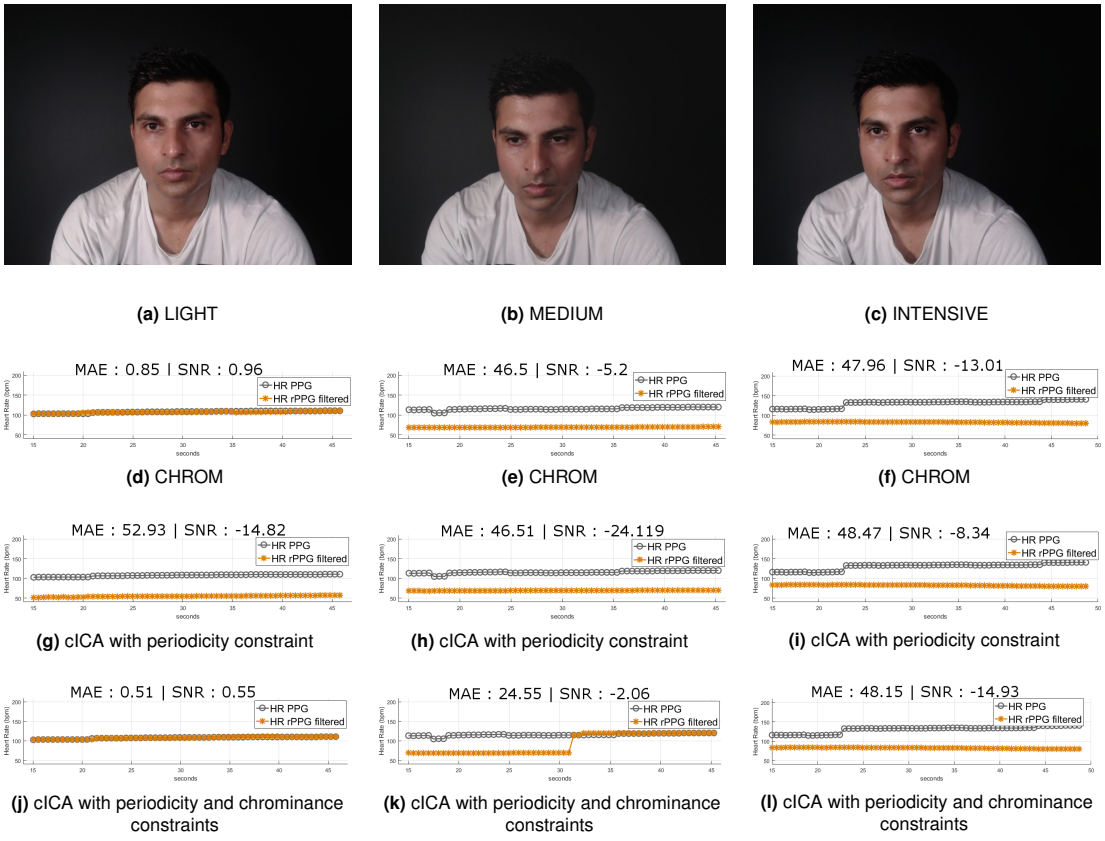


Figure 5.2: Utility of the combination of periodicity and chrominance constraints in a fitness scenario

However, the CHROM method was not able to extract the correct component for the MEDIUM and INTENSIVE videos where the variations due to motion overwhelm the rPPG variations (figure 5.2e). Interestingly, where both CHROM and cICA with just the periodicity constraint failed separately, their combination resulted in a partial convergence to the rPPG signal as is visible in figure 5.2k. This can be attributed to the constriction of the solution space resulting in the optimizer to converge to the correct rPPG signal.

Finally, for the INTENSIVE video, the motion component was much stronger and none

of the methods succeeded in extracting the rPPG signal (figures 5.2f, 5.2i and 5.2l). This calls for the use of a motion compensation scheme to mitigate such high intensity motions.

Consequently, these two constraints, the autocorrelation being a bit too lenient and the chrominance based constraint being too restrictive, in choosing the best weighting matrix, can be combined to guide the optimizer in choosing a weighting matrix with optimum flexibility. The combination is also advantageous for fitness scenarios with limited periodic motion, without any motion compensation. This implementation showed improved results compared to both ICA and CHROM methods the analysis of which is presented in chapter 5.3. The use of these two constraints in our framework is described in the next subsection.

5.1.3/ CONSTRAINED ICA

As seen earlier in chapter 4 the generic contrast function of ICA as defined by [26], the negentropy function given by $J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$ is used as the main objective function, where $H(.)$ and $\mathbf{y}_{gauss}$ have the same meaning as mentioned in section 4.1 of chapter 4. The approximation of negentropy as given in the FastICA method [26] is stated here for convenience

$$J(\mathbf{y}) \approx \rho[E\{G(\mathbf{y})\} - E\{G(v)\}]^2 \tag{5.4}$$

where as before, ρ is a positive constant, v is a zero mean, unit variance Gaussian and G(.) is a non-quadratic function

$$G(\mathbf{y}) = \frac{\log\cosh(a\mathbf{y})}{a} \tag{5.5}$$

with $1 < a < 2$.

Constrained ICA aims to alleviate the issues of ICA with the help of Lagrange multiplier methods. Lagrange multiplier methods [11] are a tool for performing constrained optimization problems following the general form

$$\text{minimize } f(\mathbf{X}), \text{ subject to } g(\mathbf{X}) \leq 0,\ h(\mathbf{X}) = 0 \tag{5.6}$$

where $f(\mathbf{X})$ is the objective function, $g(\mathbf{X})$ is a set of inequality constraints and $h(\mathbf{X})$ is a set of equality constraints.

The objective of obtaining the weighting matrix to give the optimum cardiac pulse using cICA can be fulfilled with the help of the inequality constraint

$$g(\mathbf{w}) = \epsilon(\mathbf{w}) - \zeta \leq 0 \tag{5.7}$$

where $\mathbf{w}$ represents a single demixing weight vector of size equal to the number of input channels and $\epsilon(\mathbf{w})$ represents the set of constraints to be satisfied. The optimum $\mathbf{w}$ then extracts the optimum cardiac pulse using $\mathbf{y} = \mathbf{w}^{\mathrm{T}}\mathbf{x}$. Using the average of squared autocorrelation as a constraint gives $g(\mathbf{w})$ as

$$g_1(\mathbf{w}) = \zeta_1 - E\{\mathbf{r}^2\} \leq 0 \tag{5.8}$$

where ζ_1 denotes the threshold for the lower bound of the optimum autocorrelation. This constraint guides the optimizer towards choosing the weighting matrix that results in a signal of high periodicity, with the minimum expectation of its mean squared autocorrelation as ζ_1. Next, the CHROM constraint is defined as

$$g_2(\mathbf{w}) = \|\mathbf{w} - \mathbf{w}_{\mathbf{chrom}}\| - \zeta_2 \leq 0 \tag{5.9}$$

where $\mathbf{w_{chrom}}$ is the CHROM weighting matrix from equation 5.3, and ζ_2 is the threshold for the upper bound for discrepancy between the optimum and the CHROM weighting vectors. This constraint guides the optimizer to converge towards the CHROM weighting matrix $\mathbf{w_{chrom}}$.

5.1.4/ cICA Optimization Algorithm

The general cICA problem can be defined as [34]

$$\begin{aligned} \text{Maximize :} \quad J(\mathbf{y}) &= \rho[E\{G(\mathbf{w}^{\mathrm{T}}\mathbf{x})\} - E\{G(v)\}]^2, \\ \text{Subject to :} \quad g_i(\mathbf{w}) &\leq 0,\ h(\mathbf{w}) = E\{\mathbf{y^2}\} - 1 = 0 \end{aligned} \tag{5.10}$$

where $J(\mathbf{y})$ is the one-unit contrast function as defined in equation 4.3, $g_i(\mathbf{w})$ is the set of inequality constraints to be satisfied from equations 5.8 and 5.9, and $h(\mathbf{w})$ constrains the output $\mathbf{y}$ to have unit variance.

The augmented Lagrangian formulation as adapted from [34] was used primarily because of its robustness owing to the use of penalty parameters to maintain the convexity assumption [11].

$$\begin{aligned} \mathcal{L}(\mathbf{w}, \mu, \lambda) &= J(\mathbf{y}) - \frac{1}{2\gamma_i}\left[\left\{[max\{0, \bar{g}_i(\mathbf{w})\}]^2 - \mu_i^2\right\}\right] \\ &\quad -\lambda h(\mathbf{w}) + \frac{1}{2}\beta\,\|h(\mathbf{w})\|^2 \end{aligned} \tag{5.11}$$

where $\bar{g}_i(\mathbf{w}) = \mu_i + \gamma_i g_i(\mathbf{w})$, μ_i and λ are the lagrange multipliers corresponding to $g_i(\mathbf{w})$ and $h(\mathbf{w})$ respectively. $\|.\|$ denotes the Euclidean norm and the terms $\frac{1}{2}\gamma_i\,\|.\|^2$ and $\frac{1}{2}\beta\,\|.\|^2$ are the penalty terms that makes sure that the optimization problem is held at the condition of local convexity assumption: $\nabla^2_{xx}\mathcal{L} > 0$, γ_i and β being the constraint-wise penalty parameters.

The first derivative of $\mathcal{L}$ w.r.t $\mathbf{w}$ required for the optimization given by

$$\mathcal{L}'_w = \bar{\rho}E\{\mathbf{x}G'_y(\mathbf{y})\} - \frac{1}{2}\mu E\{g'_i(\mathbf{w})\} - \lambda E\{\mathbf{xy}\} \tag{5.12}$$

where $\bar{\rho} = \pm\rho$ depending on the sign of $E\{G(\mathbf{y})\} - E\{G(v)\}$, $G'_y(\mathbf{y})$ and $g'_i(\mathbf{w})$ are the first derivatives of $G(\mathbf{y})$ and $g_i(\mathbf{w})$ w.r.t $\mathbf{y}$ and $\mathbf{w}$ respectively. The Hessian $\mathcal{L}''_{\mathbf{w}_k}$ is calculated as

$$\mathcal{L}''_{\mathbf{w}_k} = \bar{\rho}\mathbf{R_{xx}}E\{G''_y(\mathbf{y})\} - \frac{1}{2}\mu E\{g''_i(\mathbf{w})\} - \lambda \tag{5.13}$$

the inversion of which is not problematic because $\mathbf{R_{xx}}$ being the covariance matrix of the whitened and centered signal $\mathbf{x}$ is an identity matrix. $G''_{y^2}(\mathbf{y})$ and $g''_i(\mathbf{w})$ are second order derivatives and $\mathcal{L}''_{\mathbf{w}_k}$ is of size $m \times m$ which are described in chapter 3 in detail.

The expectation in the equations were calculated by using all the samples of the input signal $\mathbf{x}$. The first and second derivatives were then fed into the `fmincon` function of MATLAB using the interior point algorithm [25] to obtain the final weighting matrix $\mathbf{w}^*$ which was then used to obtain the final rPPG signal. The advantage of the interior point algorithm over Newton's method is that at any given iteration, it can choose between a Newton step or a conjugate gradient step, based on whether the Newton step is possible or not. Although, the Lagrangian is globally convex in our case, the `fmincon` function was used since it provided the convenience of plugging in the derivatives as well as tools for analyzing iterations and step-wise function and derivative values.

5.2/ SYSTEM FRAMEWORK

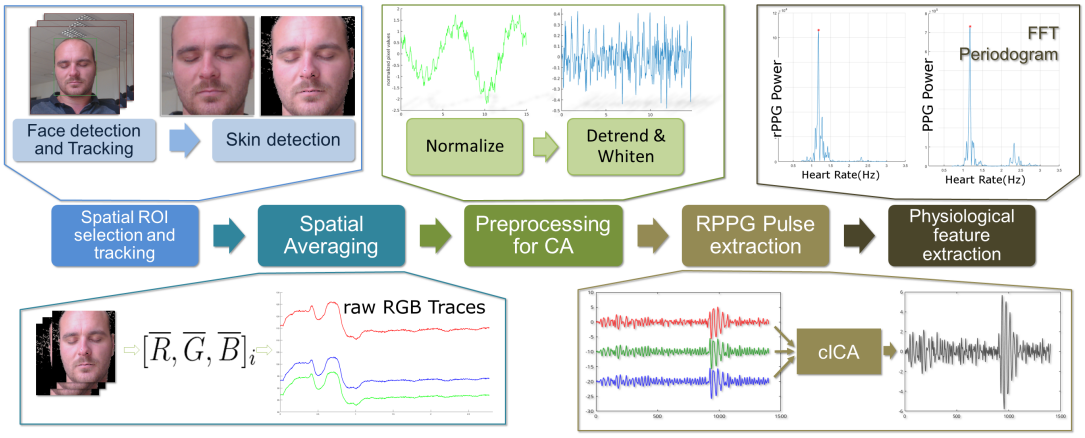


Figure 5.3: Flowchart of the proposed method

Figure 5.3 depicts the entire workflow of the procedure. The workflow follows our rPPG signal estimation framework where the MAICA method was replaced by the cICA method. The temporal RGB traces $\mathbf{x} = [x_1\ x_2\ x_3]^T$ were obtained and preprocessed using detrending and normalization in the same manner as described in section 4.2 of chapter 4. The window-wise heart rate from the estimated rPPG signal was also calculated just as was done with the MAICA method, using FFT filtered between $\mathcal{F} \in [.7, 3]$ Hz just as it was done with the MAICA method.

5.3/ RESULTS AND DISCUSSION

The UBFC-RPPG and MMSE-HR databases were used for the analysis of cICA, as for MAICA. Since the specifics regarding the databases have already been detailed in chapter 3, we can directly discuss the performance of cICA.

Table 5.1 shows the performance comparisons between ICA, cICA, MAICA and other state of the art methods, viz., PCA [53], Green [42], CHROM [58], POS [83], and G-R [39]. The metrics MAE, SNR and r are as described in chapter 4. As with the MAICA method, the ICA implementation used for the comparison was adapted from FastICA [26]. It can be observed that the cICA algorithm gives similar performance to the MAICA algorithm, albeit a slightly higher MAE for the MMSE-HR algorithm. The addition of the chrominance constraint combats to some extent against periodic disturbances (as discussed in section 5.1.2), but cannot cope as much with the spurious large movements present in several videos of the MMSE-HR database, which can attribute the slightly lower MAE than MAICA.

The cICA algorithm does perform well in comparison to other state of the art methods. In comparison to MAICA, cICA has the advantage of the ability to incorporate more constraints without increase in complexity as might happen with MAICA, based on the a priori information to be incorporated. The computation times of cICA and MAICA are similar, longer than ICA, owing to the calculation of the autocorrelation and chrominance values.

Table 5.1: Performance metrics

	UBFC-RPPG						MMSE-HR	
	SIMPLE			REALISTIC				
	MAE	SNR	r	MAE	SNR	r	MAE	r
cICA	0.62	2.01	**0.99**	**3.14**	-0.75	**0.91**	4.69	0.79
MAICA	**0.55**	**2.98**	**0.99**	3.34	**-0.26**	0.89	**3.91**	**0.86**
ICA	0.67	2.70	0.98	6.02	-1.11	0.79	5.84	0.67
PCA	2.04	-1.43	0.97	9.65	-3.45	0.67	9.15	0.49
Green	9.86	-1.61	0.29	7.73	-2.78	0.68	10.65	0.47
CHROM	0.72	3.04	**0.99**	3.81	-0.93	0.87	5.59	0.83
POS	0.67	2.57	**0.99**	4.73	-1.60	0.80	5.77	0.82
G-R	0.67	1.97	**0.99**	9.79	-3.10	0.65	8.56	0.58

It was also worth assessing the resilience of the cICA algorithm against changes in parameters such as image resolution and window length. This analysis was performed in two parts. The MMSE-HR database was chosen to assess the effect of changes in image resolution since it provides images of resolution 1040x1392 which offer the possibility to asses scaled down versions of the frames, to 35% and 75% respectively, in favor of the UBFC-RPPG database which provides images of a lower resolution, viz. 640x480 pixels. On the other hand, due to the shorter duration of videos in the MMSE-HR database, the UBFC-RPPG REALISTIC dataset was chosen to assess the effect of changes in the window length, the assessment being done against window lengths of 10 and 20 seconds, respectively.

Table 5.2 lists the average results of this assessment over the three datasets. The metrics in the last row correspond to the original results. It was observed that there was a slight decline in the performance with respect to the window length of 10 seconds, which is expected, owing to the lack of enough data for the ICA objective function to establish independence. Correspondingly, frames of lower resolution reflect loss of spatial information. Even though this loss is in itself not too deteriorating for the signal quality, the fact that it was coupled with the 15 second window length for the MMSE-HR database, explains the slightly higher MAE values for frames scaled down to 35% and 70%.

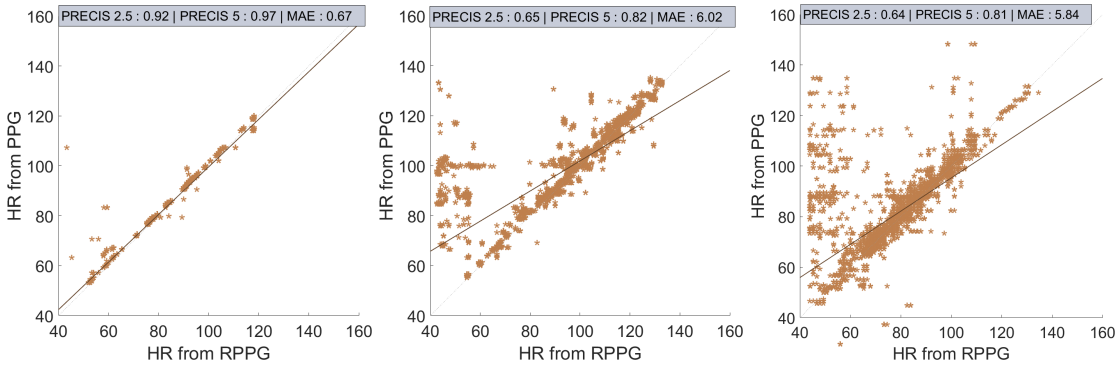
Finally, since cICA is essentially an optimization algorithm where the weights are randomly initialized, it was worth assessing its consistency over multiple runs. Figure 5.5 shows the box plot comparing the MAEs with cICA for the UBFC-SIMPLE dataset without much movement under ambient light, the UBFC-REALISTIC dataset with subjects working on a computer under ambient light, and the MMSE-HR dataset with subjects exhibiting facial expressions under indoor lighting. These tests on the datasets were performed 20 times. It is visible from the box plot that the cICA method performs consistently resulting in MAEs in the range $[0.53, 1.96]$ bpm, $[2.49, 4.1]$ bpm and $[2.63, 5.86]$ bpm for the UBFC-SIMPLE, UBFC-REALISTIC and MMSE-HR datasets respectively.

The global correlation analysis using window-wise calculations between HRs from PPG versus RPPG obtained from all the videos in each dataset for the skin-segmented pixel data for one particular run is presented in Figure 5.4. It is worth mentioning that the MAE values in figure 5.4 and table 5.1 differ from those in figure 5.5, which are averaged

Table 5.2: Effect of scale and window length

Temporal window length	UBFC-RPPG REALISTIC		
	MAE	SNR	r
10s	6.19	-1.34	0.72
20s	4.09	-1.18	0.83
30s	3.14	-0.78	0.91

Scale	MMSE-HR	
	MAE	r
35%	7.06	0.57
75%	6.68	0.61
100%	4.69	0.79



(a) ICA for UBFC-RPPG/SIMPLE **(b)** ICA for UBFC-RPPG/REALISTIC **(c)** ICA for MMSE-HR

(d) cICA for UBFC-RPPG/SIMPLE **(e)** cICA for UBFC-RPPG/REALISTIC **(f)** cICA for MMSE-HR

Figure 5.4: Correlation analysis of ICA vs cICA

over 20 executions, but are obviously in range. Moreover, differences in range of 10^{-2} bpm are inconsequential.

As was with MAICA, the analysis of the SIMPLE dataset was quite easy for both ICA and cICA. However, cICA did remove the few outliers that are present in the case of ICA. This is reflected in the low MAE and high SNR values as shown in table 5.1. This can be attributed to the fact that the subjects were generally relaxing, mostly with their eyes closed, which resulted in minimal motion artifacts.

On the other hand, as with MAICA, the REALISTIC dataset was slightly more challenging owing to the fact that the subjects were emulating the activity of using the computer, while the MMSE-HR database was challenging owing to the movements exhibited by the subjects as already mentioned in the discussion of chapter 4.

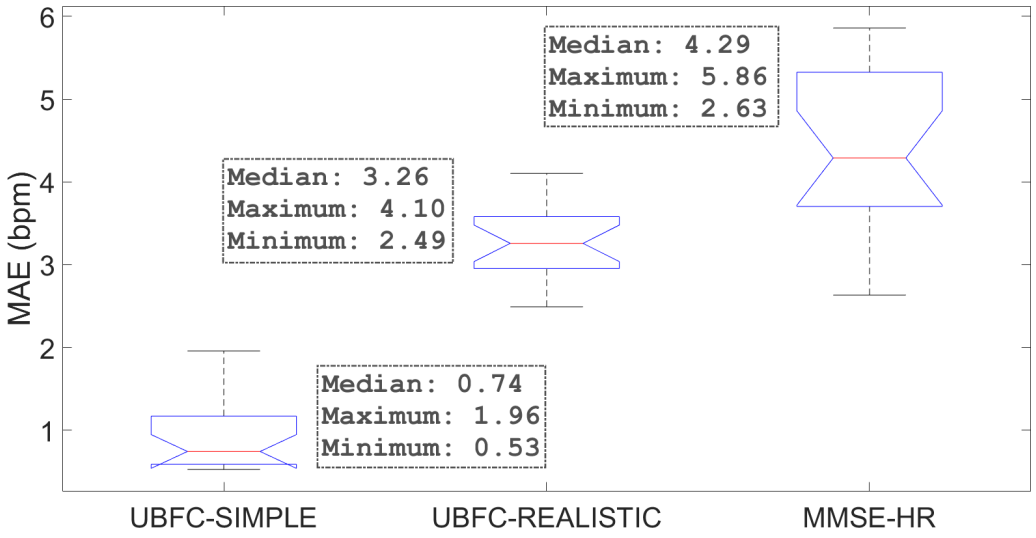


Figure 5.5: Box plot of MAE from cICA over 20 observations for the two databases

5.4/ SUMMARY

In this chapter we presented a novel semi blind source separation method for the application of rPPG measurements using autocorrelation and chrominance based constraints to guide the ICA separation process. The cICA using autocorrelation and chrominance constraints provides better result than simple ICA while removing the extra step for choosing the best component. The periodogram of the extracted signals was also consistently closer to that of the PPG.

The inclusion of the chrominance constraint can also aid for rPPG measurement in scenarios comprising of periodic movements. Since in this case, the autocorrelation constraint is likely to be contaminated by the signal corresponding to the periodic movement, the optimizer can favor the signal satisfying both, the autocorrelation and the CHROM constraints. Furthermore, for improving accuracy, better face and skin detectors and trackers can be investigated. Also, even though the CHROM constraint helps the convergence to the correct rPPG signal for limited movements, it does fail when they are more pronounced in speed and intensity. The method can thus benefit with motion compensation which itself is another interesting subject for research. Another limitation of cICA is the requirement of appropriate thresholds for the constraints in the optimization procedure. A scheme along the lines of the adaptive step-size of the parameter-free MAICA method can be used for optimal threshold estimation can be used to improve the cICA method further.

Although cICA and MAICA succeed in incorporating a priori periodicity information as required, are essentially optimization based methods which have complications often due to convergence issues. A methodology that avoids the expensive optimization scheme, at least partly, can be beneficial. Finally, the research undertaken and the results obtained in this chapter have been published and validated in the Biomedical Online Engineering journal [b] and in the IEEE E-Health and Bio-engineering conference [ii].

In the following chapter, we propose a new semi-blind source extraction method as an improvement to *the optimization* based methods MAICA and cICA discussed so far.

6

Periodic Variance Maximization

As already discussed, extraction of underlying source signals from multichannel physiological signal mixtures has been generally performed using Independent Component Analysis (ICA) in the context of Blind Source Separation (BSS) in problems such as extraction of electrocardiogram (ECG) signals, separation of fetal ECG, heart rate estimation using remote photoplethysmography (rPPG) and speech analysis. The independent components mixed into the multichannel sensor data are typically separated by maximizing independence, using metrics such as non-gaussianity, kurtosis, or mutual information [24]. However, BSS methods fail to profit from the quasi-periodic information which is a very common property among many physiological signals. Furthermore, the periodicity constraint can also mitigate against small motion and illumination disturbances.

In chapters 4 and 5 we formulated MAICA and cICA, two optimization based algorithms which exploit this periodicity criterion as a priori information to successfully extract the quasiperiodic physiological signal sources. This exploitation of quasiperiodicity can benefit in scenarios such as remote, and possibly long-term, monitoring of geriatric patients and infants, patients with severe physical trauma, computer users and drivers, in telemedicine, with more and more applications coming to light recently. This criterion is also exploited in extracting fetal ECG signals from maternal ECG signal recordings [50] and can be adapted to extraction of other afore-mentioned physiological signals. It is worth noting that the exploitation of periodicity is not enough in scenarios where the perturbing signals are themselves periodic, e.g. heart rate measurement in fitness scenarios. In such a case, additional constraints would be needed to extract the desired physiological signals.

However, since these algorithms are essentially optimization based, they are prone to being inevitably indeterministic and their complexity analysis susceptible to being cumbersome. As already seen in the previous chapters, additional regression analysis was needed to verify the robustness of these methods. Not disregarding the effectiveness of the methods already developed in the previous chapters, a different avenue for estimation of the rPPG signal was sought after and is presented in this chapter.

To this end, the iterative subspace decomposition procedure of [50] is enhanced to extract the underlying quasi-periodic signal of an unknown period embedded into the signal recordings. This new algorithm, aptly coined as Periodic Variance Maximization (PVM), is applied to remote photoplethysmography to extract the cardiac signal embedded in the RGB temporal traces. The PVM algorithm aims to find the unknown period of the desired signal, by combining two approaches. First, the iterative subspace decomposition procedure, that estimates a periodicity maximizing basis for a given frequency, and second, a

global optimization algorithm of tabu search to find the frequency with the highest global periodicity over the search space. The proposed method can be used to extract any desired quasi-periodic signal of an unknown period from a mixture of signals, for any type of physiological signal measurement scenarios where periodic motion is not involved.

We validate our methodology with the application to remote photoplethysmography (rPPG) and analyse its performance against two public databases, the UBFC-RPPG [76] database and the MMSE-HR [84] database. The rest of the chapter is organized as follows. In section 6.1 we present an overview of periodicity maximization algorithms and state of the art rPPG methods. The algorithm is described in section 6.2 followed by performance analysis in section 6.3.

6.1/ Related work on periodic signal extraction

Sameni *et al.* have exploited the periodic nature of ECG signals to separate multichannel fetal and maternal ECG recordings using periodic component analysis. They achieve this separation by maximizing a measure of periodicity [41]. They further expanded on their work by proposing a generalized deflation framework to separate the target signal based on not only periodicity, but any other quantifiable properties, such as SNR, stationarity, and spectral contrast, from noisy multichannel recordings [50]. However, in their work, the period of the desired signal to be separated is fixed and known, calculated from the maternal ECG.

Tsalaile *et al.* provide an improvement to this method, by allowing the sequential blind source extraction of quasi-periodic signals having *time-varying* periods, by diagonalizing time-lagged autocorrelation matrices at time-varying lags [45]. Despite the ability of their method to handle periods varying over time, these periods are still known, and are a part of the a priori information. On the contrary, we propose to extract the quasi-periodic signal of an *unknown* period from the signal mixtures.

In the field of speech analysis, Saul *et al.* have used an eigenvalue method to analyze and enhance weak periodic signals which is insensitive to phase thanks to the use of Hilbert transforms. They perform the periodicity measurements using efficient sinusoidal fits to extract the fundamental frequencies, albeit with extensive auditory preprocessing, habitual to the domain of speech analysis which are computationally intensive and may hurt its feasibility for real-time applications.

In this chapter, we propose a method to extract periodic or quasi-periodic signals of unknown period embedded into multi-channel temporal signal recordings. The proposed technique is applied to remote photoplethysmography in order to extract the cardiac signal embedded in the RGB temporal traces. The quasi-periodic cardiac signal is mixed in the light reflected by the tissue with other signals such as changes in incident light or motion induced shadow casting variations. This mixed signal is then captured by the camera. The proposed PVM algorithm aims to find the quasi-periodic cardiac signal from the mixture. Moreover, contrary to the known problem of fetal ECG extraction [50], the period is unknown and consequently it is an interesting problem for us.

Recently, research on remote photoplethysmoraphy measurements has been on the rise where different classes of methods have been proposed to extract the quasi-periodic cardiac signal embedded in RGB temporal traces built from sequential video frames.

However, periodicity information has not been exploited in the cotext of rPPG measurement. To the best of our knowledge, the combination of principal component analysis with periodicity maximization has not been combined for extracting a quasi-periodic signal of unknown period. The Periodicity Variance Maximization algorithm is presented in the next section.

6.2/ PROPOSED METHOD

The proposed method aims to extract the quasi-periodic signal embedded in the recorded signals which is marked by high periodicity, or in other words *high periodic variance*, corresponding to its fundamental period τ^*. For a centered temporal signal, $\mathbf{y} = \mathbf{y}(t)$, we define high periodic variance at period τ as the property that corresponds to a high variance $\sum \mathbf{y}\mathbf{y}^T$, as well as a high *lagged* variance $\sum \mathbf{y}\mathbf{y}_\tau$, where $\mathbf{y}_\tau = \mathbf{y}(t+\tau)$. Typically, a periodic or a quasi-periodic signal exhibits high periodic variance at τ corresponding to its fundamental frequency $f = 1/\tau$. Typical biomedical signals such as electrocardiography, electromyography, neural and photoplethysmography signals are quintessentially quasi-periodic and more often than not, periodic.

The method takes as input temporal traces $\mathbf{x} \in \mathbb{R}^M$ of length N and M channels, centered by subtracting the channel-wise temporal means, to extract the most periodic signal of an unknown period, using $\mathbf{y} = \mathbf{w}^T\mathbf{x}$. Here, $\mathbf{w} \in \mathbb{R}^M$ is the optimum weighting vector which gives the desired signal $\mathbf{y}$ with the highest periodic variance, obviously corresponding to the fundamental period of the desired signal. To this end, we use Generalized EigenValue Decomposition (GEVD) on the pair of the lagged covariance and covariance matrices of $\mathbf{x}$ defined as

$$P_\mathbf{x} = \mathbf{x}\mathbf{x}_\tau^T, \quad C_\mathbf{x} = \mathbf{x}\mathbf{x}^T \tag{6.1}$$

where $\mathbf{x}_\tau$ are the temporal traces centered and lagged by τ seconds. Estimating the unknown fundamental period is not an uncommon problem in a broad range of applications. An optimization scheme can be used to estimate the unknown fundamental period τ^* of the desired signal by maximizing a periodicity metric similar to the one defined in [50] given by

$$\mathcal{P}(\tau, \mathbf{w}) \doteq \frac{E_t\{\mathbf{y}\mathbf{y}_\tau\}}{E_t\{\mathbf{y}^2\}} = \frac{\mathbf{w}^T P_\mathbf{x} \mathbf{w}}{\mathbf{w}^T C_\mathbf{x} \mathbf{w}} \tag{6.2}$$

where $E_t\{\cdot\}$ represents the temporal expectation.

Intuitively, the periodicity metric $\mathcal{P}$ represents the extent of periodic information in the signals since it is the ratio of the lagged covariance matrix to the covariance matrix in a different basis. If the constituent signals were completely periodic, $P_\mathbf{x}$ and $C_\mathbf{x}$ would be equivalent giving $\mathcal{P} = 1$. Also, to ensure that the generalized eigenvalues are real, $P_\mathbf{x}$ needs to be symmetrized using $P_\mathbf{x} = (P_\mathbf{x} + P_\mathbf{x}^T)/2$. This symmetrized $P_\mathbf{x}$ matrix combines the two way variances between the channels, for instance, variances between the channels (R, G) and (G, R). This symmetrization represents the overall lagged covariance among the channels to be maximized, to make it as similar as possible to $C_\mathbf{x}$.

It is to be noted that $C_\mathbf{x}$ is resymmetrized only to be coherent in scale with $P_\mathbf{x}$, since being a covariance matrix, it is already symmetric. After symmetrization, GEVD is performed on the pair $(P_\mathbf{x}, C_\mathbf{x})$ to estimate the matrices W and D such that

$$W^T P_\mathbf{x} W = D, \quad W^T C_\mathbf{x} W = I \tag{6.3}$$

where the diagonal matrix D contains the generalized eigenvalues corresponding to the eigenvectors $W = [\mathbf{w}_1, \cdots, \mathbf{w}_N]$, the eigenvalues being real and sorted in the ascending order along the diagonal.

For a set of temporal signals $\mathbf{x} \in \mathbb{R}^M$, these generalized eigenvectors have three important properties. First, they correspond to a change of basis onto which the projection $W^T\mathbf{x}$ of the original signals $\mathbf{x}$ are entirely uncorrelated. This is evident from the fact that the covariance matrix in this new space given by $(W^T\mathbf{x})(\mathbf{x}^T W)$ from equation 6.3 is diagonal, ensuring maximum variance of the data. Second, they also diagonalize the lagged covariance matrix $P_\mathbf{x}$, making the projections of $\mathbf{x}$ and $\mathbf{x}_\tau$ on this new basis entirely uncorrelated as well, which is possible only when the lagged signal $\mathbf{x}_\tau$ is extremely similar to the original signal $\mathbf{x}$ at that given lag. In fact, the magnitude of the generalized eigenvalues actually reflects the amount of similarity between the original and the lagged signals. And finally, the first eigenvector $\mathbf{w}_1$ corresponding to the largest generalized eigenvalue that maximizes the ratio defined in equation 6.2, also referred to as the Rayleigh Quotient [17]. Consequently, the projection of the original signals $\mathbf{x}$ to the new basis represented by $\mathbf{w}_1$ captures the maximum periodic content. Owing to these properties of GEVD, it can be used in an iterative algorithm to extract the components exhibiting high periodicity and simultaneously containing the maximum information.

6.2.1/ Iterative Periodic Variance Maximization

We can now proceed with the formulation of an optimization scheme that iteratively maximizes the periodicity of the projected signals over the range of frequencies corresponding to the specific application. For instance, in the context of rPPG measurements, this range corresponds to the human heart rate. Evidently, the optimizer needs to estimate the optimum value of the pair $(\tau, \mathbf{w})$ which maximizes the periodicity metric $\mathcal{P}$ equation 6.2.

The most periodic signal can then be obtained by estimating the weighting matrix $\mathbf{w}^*$ that maximizes $\mathcal{P}$ over the frequency search space $\mathcal{F} \in [f_{min}, f_{max}]$. Of course, owing to the time domain formulation of our problem, the optimization needs to be performed over the period τ, with a step of $1/F_s$ seconds, over the temporal search space $\tau \in [\tau_{min}, \tau_{max}]$ corresponding to the frequency range $[f_{max}, f_{min}]$. Here F_s is the sampling rate of the recorded signals. An implicit advantage to this formulation in the time domain is the restricted search space corresponding to discrete integer time lags of $[\tau_{min}F_s, \tau_{max}F_s]$ which facilitates the use of this method in a live scenario, provided that the sampling rate is not exceedingly high.

For instance, in the application of rPPG measurement, the frame rate is typically between 20 to 30 frames/second for conventional cameras. Coupled with the limited range of human heart rates $\mathcal{F} \in [40, 200]\ bpm$, or $[.3, 1.5]$ second, this makes the live implementation of this method undoubtedly feasible. Indeed, the search space depends on F_s, facilitating finer searches at higher frame rates. This two-variable optimization of the pair $(\tau, \mathbf{w})$ can be simplified into the following two steps.

6.2.1.1/ Estimate the basis that maximizes periodicity

The first core step comprises of estimating the weighting matrix $\mathbf{w}^*$, which is nothing but a change of basis, that maximizes the periodicity metric in equation 6.2 for a given time

Algorithm 1 Optimum Basis Estimation

procedure OBE($\mathbf{x}, \tau$) ▷ Estimate the best $\mathbf{w}$ for τ
  for $k = 1 : K$ **do**
    Calculate $P_{\mathbf{x}}$ and $C_{\mathbf{x}}$ ▷ Using equation 6.1
    $W \leftarrow$ GEVD($P_{\mathbf{x}}$,$C_{\mathbf{x}}$)
    $\mathbf{s} \leftarrow W^T\mathbf{x}$ ▷ Project onto new basis
    $[\mathbf{s}_1 ... \mathbf{s}_R] \leftarrow G([\mathbf{s}_1 ... \mathbf{s}_R])$ ▷ Smooth R channels
    $\mathbf{x} \leftarrow W^{-1}\mathbf{s}$ ▷ Back project to original space
  end for
  $W = [\mathbf{w}_1 \cdots \mathbf{w}_M]$ ▷ M channels
  $\mathbf{y} = \mathbf{w}_1^T\mathbf{x}$ ▷ Most periodic component
  $\mathbf{w}^* = \mathbf{w}_1$
  return $\mathbf{w}^*$
end procedure

period τ. This sub-method was adapted from [50] and is listed in algorithm 1. It starts with a GEVD step that orders the components by their periodicity, based on the magnitude of the generalized eigenvalues. The decomposed signals are then projected onto the new subspace represented by W. Next, denoising is performed on the first R signals by using a wavelet denoiser, using the parameters similar to [41]. And lastly, the mixture of the projected and denoised signals are back-projected onto the original space. With regards to the denoising step, it could have been applied directly to the input signal $\mathbf{x}$, but the advantage of applying it after the linear decomposition is that we can benefit from the improved signal quality of the first few extracted components.

The above steps are repeated for K iterations. The output of this core step is the periodicity maximizing basis $\mathbf{w}_1$ which is the generalized eigenvector corresponding to the highest generalized eigenvalue of the pair $(P_{\mathbf{x}}, C_{\mathbf{x}})$. Furthermore, for $K > 1$, the temporal signals are filtered to enhance periodic components. The impact of different values of K and R on the eventual estimation of the most periodic signal is presented in section 6.3 with the application of rPPG measurement.

Finally, we can represent the output of this step concisely as a function $\mathcal{W}(\tau)$ and consequently rewrite the periodicity metric equation 6.2 solely as a function of τ:

$$\mathcal{P}(\tau) = \frac{\mathcal{W}(\tau)^T P_{\mathbf{x}} \mathcal{W}(\tau)}{\mathcal{W}(\tau)^T C_{\mathbf{x}} \mathcal{W}(\tau)}. \tag{6.4}$$

This objective function can now be maximized by using an appropriate optimization scheme the details of which are presented in the next subsection.

6.2.1.2/ Optimize the periodicity metric $\mathcal{P}(\tau)$

To select an appropriate optimization scheme, the objective function $\mathcal{P}(\tau)$ in equation 6.4 warrants some examination. First, it is evident that for temporal signals of a specific sampling rate, it is a continuous function over the corresponding temporal search space. The complexity, however, lies in the calculation of the derivative $\mathcal{W}'(\tau)$ owing to the presence of the GEVD step. In the general sense, differentiation of an eigendecomposition represents the change in the eigenvalues with respect to change in the original data. Although,

such a differentiation is mathematically possible, it is admittedly non-trivial [14] [38] and even more complicated with respect to the period τ, which points towards the use of a derivate free optimization approach.

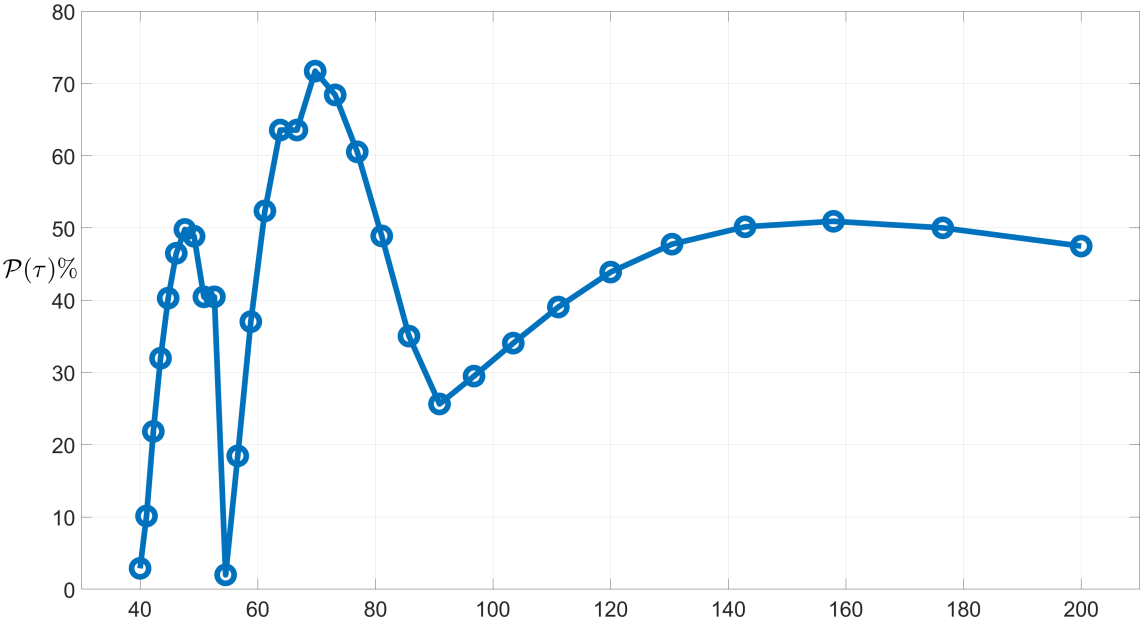


Figure 6.1: Typical distribution of $\mathcal{P}(\tau)$

Another feature of interest was the shape of the objective function. To assess this, application to the problem of rPPG signal estimation was chosen. Using videos from UBFC-RPPG and MMSE-HR databases the objective function values of $\mathcal{P}(\tau)$ for several videos from UBFC-RPPG and MMSE-HR databases were examined. Figure 6.1 shows the variation of $\mathcal{P}$ with respect to τ, calculated over a window of 15 seconds for a typical video, taken from the MMSE-HR database. It can be observed that the variation of $\mathcal{P}(\tau)$ is prone to multiple local maximas. This is expected behavior owing to the product $\mathbf{x}\mathbf{x}_\tau^T$ in the calculation of $\mathcal{P}(\tau)$. Specifically, a signal with a real period of τ^* seconds will exhibit a relatively higher value for $\mathcal{P}(\tau)|_{\tau=\tau^*}$. Accompanied by this global maximum are local maximas exhibited at $\tau = n\tau^*, n \in \mathbb{Z}$, albeit lower than $\mathcal{P}|_{\tau=\tau^*}$, and with magnitudes inversely proportional to n. This appearance of multiple local maximas, with decreasing magnitudes, is similar to the plot of the autocorrelation function. Additionally, the advantage of a limited search space, suggests the use of a global optimization scheme for best results. Algorithms that have the ability to escape local optimas serve as an appropriate solution.

Consequently, the tabu search algorithm [16] was used to perform the global optimization of $\mathcal{P}(\tau)$. Glover describes tabu search as a *meta-heuristic* superimposed on another heuristic. As the name suggests, *tabu* search strives to escape local optimas by disallowing already examined function values, nominating them as *tabu* moves. In this respect, it can also be considered a "weak inhibitive" search, where in our case, the best τ is searched intelligently over the limited search space to obtain the optimum, instead of doing a brute force search. However, even though the inhibitive search results in reduced computation times, a brute force search isn't too computationally taxing in our case, owing to the limited search space. But the advantage of tabu search lies in its ability to be overlain on a more sophisticated optimization scheme if needed. This in turn renders the

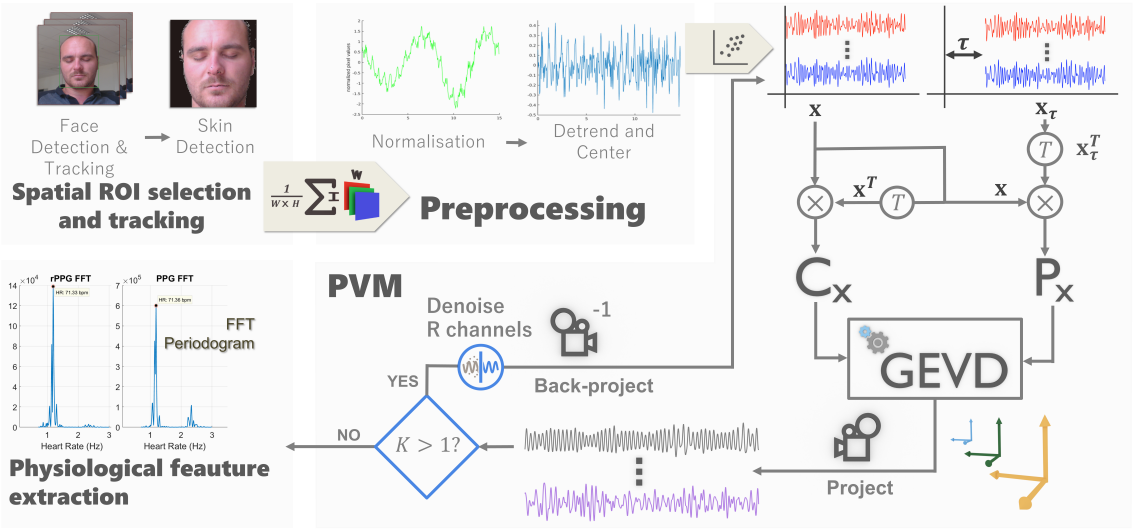


Figure 6.2: System framework and Periodic variance maximization overview

PVM method more flexible. For instance, in applications where the search space might be larger, a more advanced optimization scheme can be easily used instead of the weak inhibitive search used in the application of rPPG measurement.

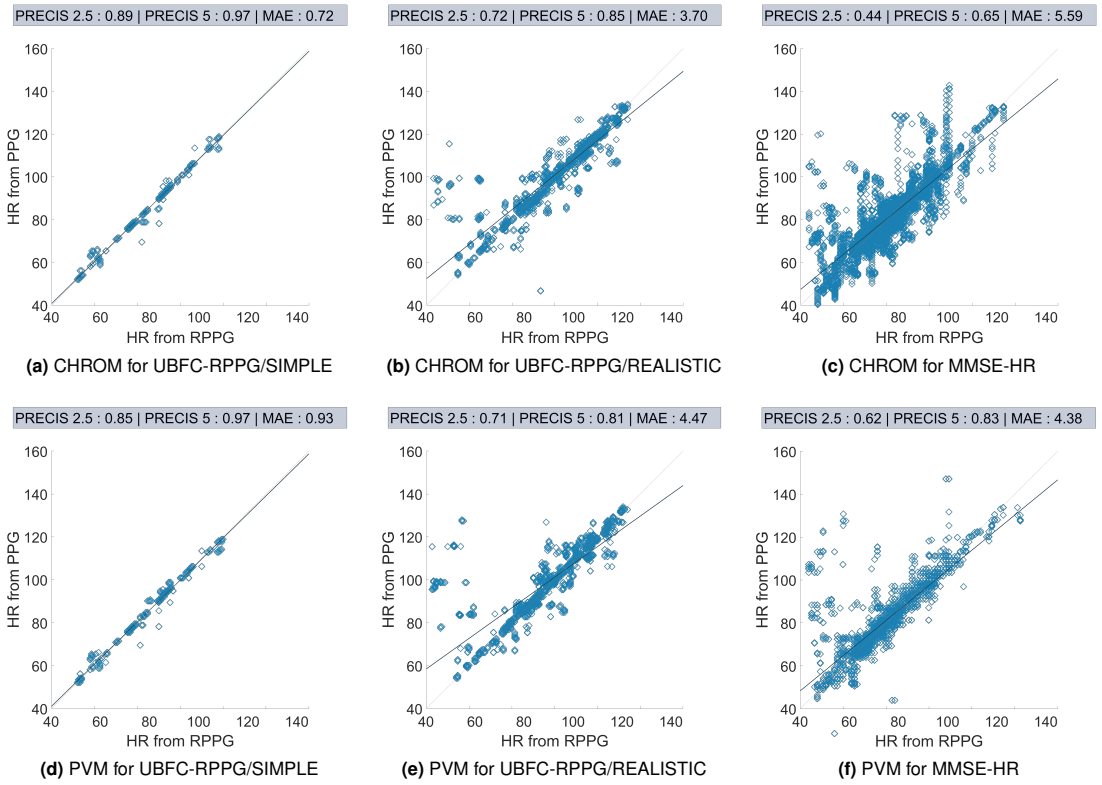
6.2.2/ Periodic Variance Maximization applied to rPPG signal estimation

We present here the applicability of the PVM algorithm implemented in the context of rPPG signal estimation using sequential video data from conventional cameras. The workflow of the procedure as depicted in figure 6.2 is presented here whose acquisition and preprocessing steps are similar to those for the cICA and MAICA methods. The temporal RGB traces, $\mathbf{x} = [x_1, x_2, x_3]^T$ were also generated in the same manner as for the analysis of cICA and MAICA the details of which have already been presented in chapter 4, and were detrended and normalized as before. These traces were then passed to the PVM algorithm which performed GEVD on the lagged covariance and covariance matrices $P_{\mathbf{x}}$ and $C_{\mathbf{x}}$ after which the first channel was denoised, giving the final rPPG signal. The rPPG signal was then post-processed using the same Kalman filtering scheme and the heart rate calculated using the same FFT scheme as for MAICA and cICA. We present the results of the experiments in the next section.

6.3/ Results and Discussion

Similar to the analysis of cICA and MAICA, the PVM algorithm was validated using the UBFC-RPPG and MMSE-HR [84] databases the details of which can be found in chapter 3. The performance analysis of the PVM method is presented here.

Figure 6.3 presents the overall correlation analysis between the PVM method and the CHROM method which in our experience is one of the most consistent and computationally efficient methods for rPPG estimation. The performance of the PVM method is similar



(a) CHROM for UBFC-RPPG/SIMPLE **(b)** CHROM for UBFC-RPPG/REALISTIC **(c)** CHROM for MMSE-HR

(d) PVM for UBFC-RPPG/SIMPLE **(e)** PVM for UBFC-RPPG/REALISTIC **(f)** PVM for MMSE-HR

Figure 6.3: Correlation comparison plots for CHROM vs PVM. The metrics PRECIS 2.5 and PRECIS 5 represent the percentage of windows where $\delta = |HR_{rPPG} - HR_{PPG}| < 2.5$ and 5 bpm respectively.

to that of the CHROM method, albeit with a slightly higher average MAE. However, it can be clearly seen that PVM exhibits a higher correlation between PPG and rPPG heart rates, with the values more grouped towards the ideal correlation line, especially for the MMSE-HR dataset in figure 6.3f.

Table 6.1 shows the performance comparisons of PVM, cICA and MAICA with other state of the art methods, viz., ICA [24], PCA [53], Green [42], CHROM [58], Plane Orthogonal to Skin (POS) [83], and G-R [39], the ICA implementation being adapted from FastICA [26] as mentioned before. As earlier, only the MAE values for the MMSE-HR dataset are presented since the SNR values are inconsequential owing to the absence of the groundtruth PPG signal.

One possible hypothesis for this inconsistency in the MAEs can be as follows. Since the REALISTIC dataset was recorded under ambient light, the effect of the fluorescent light source in certain videos was more pronounced, most probably due to lesser ambient light. The 50 Hz flicker of fluorescent lights is a well known problem in video recordings where each video frame is exposed at different light pulses. The discrepancy between our 30 Hz frame rate and the 50 Hz flicker might result in a perturbed periodic signal which overshadows the cardiac signal in certain cases and is selected by PVM.

The CHROM method is able to overcome this problem because of the projection of RGB signals onto a different subspace where only the reflections specific to the skin properties are enhanced. On the other hand, all the videos in the MMSE-HR database were recorded with a more sophisticated lighting setup, thus avoiding such an issue and

Table 6.1: Performance metrics for PVM

	UBFC-RPPG						MMSE-HR	
	SIMPLE			REALISTIC				
	MAE	SNR	r	MAE	SNR	r	MAE	r
PVM	0.93	2.01	**0.99**	4.47	**-0.22**	0.82	4.38	0.82
cICA	0.62	2.01	**0.99**	**3.14**	-0.75	**0.91**	4.69	0.79
MAICA	**0.55**	2.98	**0.99**	3.34	-0.26	0.89	**3.91**	**0.86**
ICA	0.67	2.70	0.98	6.02	-1.11	0.79	5.84	0.67
PCA	2.04	-1.43	0.97	9.65	-3.45	0.67	9.15	0.49
Green	9.86	-1.61	0.29	7.73	-2.78	0.68	10.65	0.47
CHROM	0.72	**3.04**	0.99	3.70	-0.32	0.87	5.59	0.83
POS	0.67	2.57	0.99	4.73	-1.60	0.80	5.77	0.82
G-R	0.67	1.97	0.99	9.79	-3.10	0.65	8.56	0.58

resulting in a better performance for the PVM method. However, the estimation and elimination of this background flicker is in itself not a difficult issue and can be solved with an appropriate filtering scheme, which incidentally is one of the issues we aim to address in our future work.

It can also be observed from the metrics that the cICA and MAICA methods provide lower MAEs and higher SNRs than the PVM method. Since PVM primarily performs maximization of periodicity and correlation, it is more susceptible to the effect of the 50Hz flicker light as compared to cICA, which optimizes over the entire search space with the help of the chrominance constraint, and to MAICA, which chooses from multiple candidate solutions thanks to the multi-objective optimization scheme. However, the precise reason because of which these methods are slightly more accurate, which is the exhaustive optimization over the entire search space, is the cause of their longer computational times. In this respect, the PVM finds the most periodic signal, without the expensive optimization, and thus is fast enough to be used in a real-time environment, since for each frame, only tabu search (over a small search space) and GEVD calculation is needed. As is in many computational paradigms, there is a trade-off here between speed and accuracy, which of course can be improved by a more sophisticated filtering scheme to remove the background filter.

The MMSE-HR database was challenging owing to it being an emotion elicitation database and the already discussed issue of subjects exhibiting large motion since they were required to express those specific emotions. Interestingly, the PVM method outperformed the CHROM method by a larger margin for this database. This highlights the relative robustness of the PVM method against motion variances as well. The better performance of the PVM method can also be attributed to the sophisticated lighting setup of the MMSE-HR database, as is evident from the narrower spread and the closeness of the fitting line the 45° line for PVM as compared to CHROM.

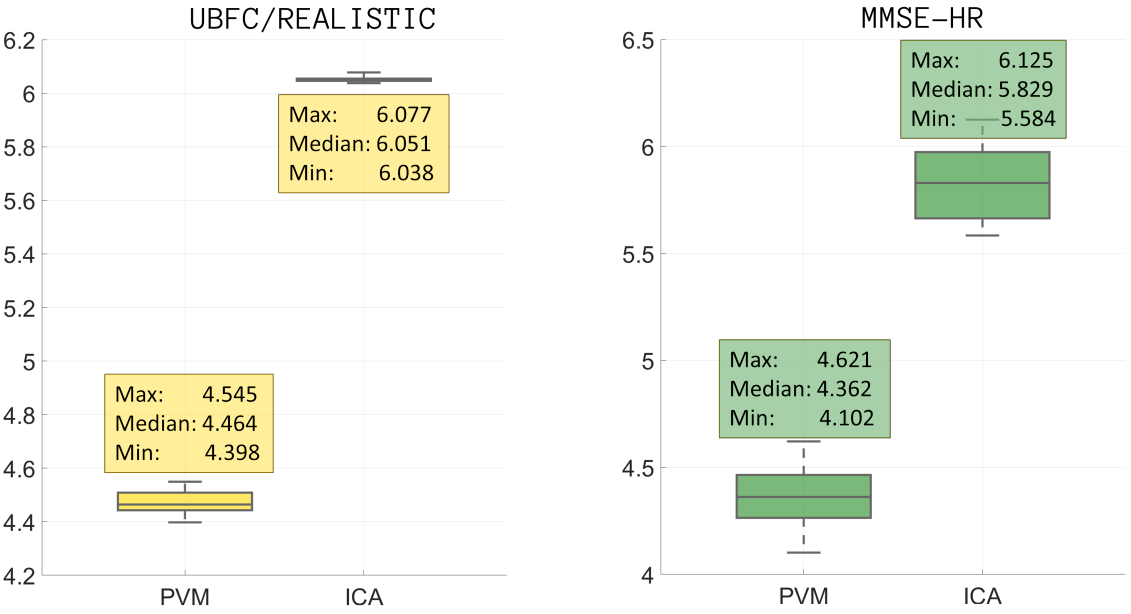


Figure 6.4: Box plot of MAE from the PVM and ICA algorithms over 20 observations for the two databases

Effect of the parameters K and R on the overall result The value of K depends on the number of dimensions of the desired subspace, which is 2 in our case. However, for our application, a value of $K = 1$ gives optimal results, which again makes this algorithm suitable for use in live scenarios. Similarly, the performance of the method with respect to $R \in [1 : M]$ where $M = 3$ is the number of channels, was also performed. It was equally observed the algorithm was able obtain the most periodic component even by denoising just one channel thanks to the efficiency of the GEVD step to successfully sort components in order of decreasing periodicity.

Finally, to assess the consistency of the overall algorithm, and because of the use of the tabu search global optimization, it was worth assessing the performance of the method over multiple runs. Figure 6.4 shows the box plot comparing the MAEs with the PVM and ICA algorithms for the two challenging datasets: the REALISTIC dataset from the UBFC-RPPG database, with subjects working on a computer under ambient light, and the MMSE-HR dataset with subjects exhibiting facial expressions under indoor lighting. These tests on the datasets were performed 20 times. The PVM method has a consistent performance, giving MAEs in the range $[4.39, 4.55]$ bpm and $[4.10, 4.62]$ bpm for the UBFC-RPPG/REALISTIC and MMSE-HR datasets respectively. The consistency analysis for the other baseline methods was not required because only the PVM and ICA methods are optimization based. The box plots for the SIMPLE dataset are not shown since they exhibit MAEs of less than 1 bpm for both PVM and ICA methods and their comparison was deemed inconsequential. It is also worth mentioning that the MAE values in figure 6.3 and table 6.1 differ from those in figure 6.4, which are averaged over 20 executions, but are obviously in range. Moreover, differences in range of 10^{-2} bpm are inconsequential.

6.4/ SUMMARY

A slightly different approach as compared to traditional optimization (for the principal objective function) was presented in this chapter. The Periodic Variance Maximization algorithm extracts the most periodic signal of an unknown period from a mixture of temporal recordings, with a drastically reduced search space, thereby opening up the possibility for it to be used in real-time applications. Moreover, the PVM algorithm is fairly generic and can be used in any problem domain where periodicity maximization, or maximization of properties such as SNR or spectral contrast is called for, in which case it would maximize the specific property. It can also be combined with smart-ROI selection methods to exploit the spatial and/or facial features to enhance the resulting signal. Its vulnerability, however lies with scenarios having periodic motion, e.g. fitness based rPPG signal estimation. Motion compensation schemes, which itself is another subject of research, can be beneficial in such cases. Future work comprises of addressing the issue of perturbation arising from background light flicker.The work presented in this chapter and the corresponding results obtained have been published and validated in the Computer Vision for Physiological Measurement workshop, in the Computer Vision and Pattern Recognition conference [iii]. The extension of the PVM algorithm to the multi-linear case using tensor analysis to estimate rPPG signal strength across the skin region is discussed in the next chapter.

7

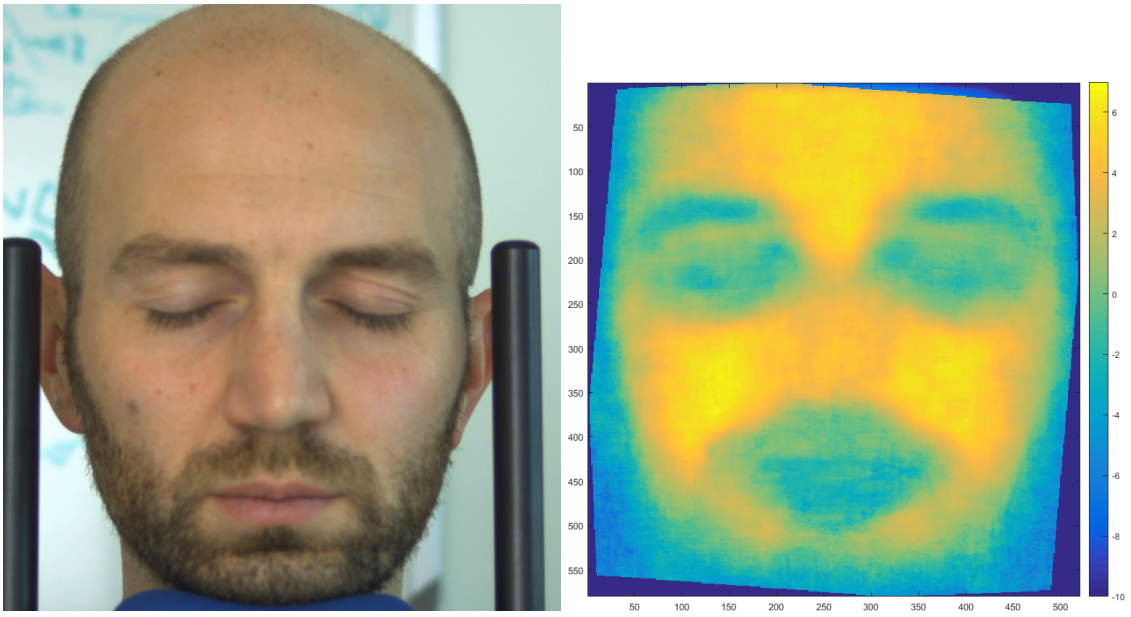
Generalized Eigenvalue Decomposition in higher dimensions

In the previous chapter we presented the Periodic Variance Maximization (PVM) algorithm that extracts the most periodic component from temporal mixtures of RGB channels obtained by frame-wise spatial averaging. We also discussed the issue of information loss with spatial averaging and the requirement of extending the PVM algorithm to a higher dimensional case where the essential spatial information is kept intact. This spatial information is significant because there might be interesting relationships between signals arising from different positions such as changes in the signal strength at different positions due to spatial variations in the blood perfusion.

To validate the idea of the spatial variation of relative strength of the rPPG amplitude, an initial experiment was performed where 10 videos were recorded using a setup so as the face remains completely stationary. Videos were recorded at 30 frames/second, at a resolution of 1024×768 pixels, and the rPPG signal was obtained from face cropped regions of the video for each pixel using the CHROM method [58] after resizing to 580×520 to reduce quantization noise. After that, pixel-wise heart rates were calculated using FFT filtered between the normal human heart rate range of $[.7, 3]$ Hz. Finally, pixel-wise SNRs were calculated and averaged over all the frames across all the videos. Figure 7.1 shows a sample video frame depicting the setup on the left and the average SNR map on the right. It can be observed from the SNR map that the relative amplitude of the rPPG signal varies across the face, and is stronger around the forehead and cheek regions. In fact, this knowledge is not entirely new, and has been utilized in various smart-roi based rPPG extraction methods as discussed in chapter 2 [81], [63], [70].

In the context of rPPG measurement, the methods that have been developed up until now in this thesis, and for that fact, most others in literature, average the skin pixels, thereby essentially loosing these spatial relationships. Moreover, for other signal sources such as ECG and neural signals, keeping all the sources is indispensable for an accurate analysis of the problem at hand. However, a related point to consider is that one of the well-known uses of spatial averaging, as is seen in the commonly used operation of *smoothing*, is the reduction in quantization noise. This fact is vital and needs to be taken into consideration while implementing any algorithm that deals with spatial information.

To develop such a spatially aware algorithm a relevant application is crucial. Apropos of this requirement, the application of Spatial rPPG Distribution Estimation (SrPDE)



(a) Setup for stationary face videos **(b)** The SNR map

Figure 7.1: Relative strength of the rPPG signal across the face obtained using the CHROM method.

is an interesting challenge with its abundance of spatial information, especially found in facial videos. The results of SrPDE can reveal underlying information about the relative spatial amplitude across the skin region, or perhaps even the PPG transit time, which represents the phase differences in the rPPG signal across various spatial points on the skin. A straightforward but naive solution to this problem would be to repeat the PVM algorithm, or in that case, any algorithm that works on averaged signals, for every pixel. But apart from being uninventive, this sequential treatment of the problem is prone to being computationally expensive, redundant and highly unscalable. Although parallelization and machine level optimizations can assuage this problem to a certain extent, the improvements obtained by such techniques on a fundamentally redundant method shall be limited.

A more interesting and advantageous alternative might be to design an algorithm that can perform a suitable core method simultaneously on a set of values (pixels), while giving the same results as when the method was performed sequentially on each value (pixel). Such a method would even prove advantageous in applications apart from SrPDE, especially dealing with temporal data, where a core algorithm needs to be repeated over a spatial distribution. For conserving homogeneity and coherence with the thesis, we shall develop the theory behind such an algorithm with regards to the application of SrPDE, keeping in mind, however, that the algorithm should be adaptable to a different application without much effort. The core idea behind such an algorithm is explored next.

Let us consider the covariance C between the RGB channels calculated over a temporal window of N samples, for each pixel of the video frame of size $W \times W$. Figure 7.2 depicts this matrix of size $3W \times 3W$ made up of a block matrix of size $W \times W$, where each 3×3 block represents the pixel-wise covariance. Consider a simple operation, such as a linear combination $aC^{k,l}_{(1,3)} + bC^{k,l}_{(2,3)}$ between the $(1,3)$ and $(2,3)$ elements of each block. Here $k, l \in [1, W]$ represent the row and column indices of each block in C. One manner of

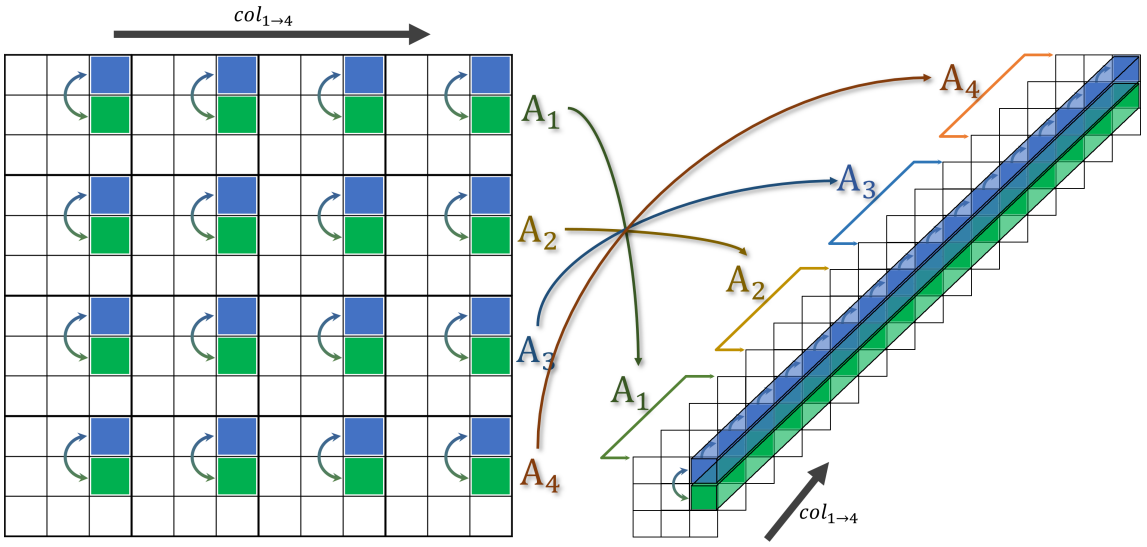


Figure 7.2: Transformation of sequential block-wise operation to simultaneous vectorized operation on a 3^{rd} order tensor. Block-wise operation between to scalar elements is transformed into an operation between two vectors (blue and green cuboids).

performing this operation simultaneously is to perform the vector operation $a\mathbf{c}_{(1,3)} + b\mathbf{c}_{(2,3)}$ where $\mathbf{c}_{(i,j)}$ represents the vector formed by the elements $(i, j)_w$ from the tensor obtained after transforming the matrix C, and where $i, j \in [1, 3]$ and $w = (i-1)W + j$ such that $w \in [1, W^2]$, as shown in figure 7.2 for $W = 4$. In this transformation, each 3×3 block of C is concatenated in the 3^{rd} dimension, row by row to obtain a tensor of size $3 \times 3 \times W^2$. This leads to the possibility of performing *implicit parallelization* which is found in modern high level languages, such as MATLAB [95], in the form of vectorized operations. Vectorized operations are themselves an independent subject of research in optimization for speeding up mathematical operations on arrays using low-level custom instructions [85].

Using such a tensor representation, simple operations can be easily performed over a large number of block matrices, which in essence are computationally more efficient than sequential operations on each block. The challenge, however, lies in iterative algorithms which might affect each block separately, but can be solved using appropriate manipulations (which we shall discuss later in sections 7.2.1 and 7.2.2). In this chapter we describe the development of such a higher dimensional analogue for the core algorithm of the PVM method, the Generalized Eigenvalue Decompisition (GEVD). Specifically, we extend the QZ algorithm [46] which is used for GEVD calculation of a pair of matrices, to two higher dimensional variants, starting with the Block QZ algorithm and culminating in its optimized successor the Tensor QZ algorithm. We start with formally defining the application of SrPDE in the next section, followed by the idea of block diagonalization and its analogy to simultaneous diagonalization of flattened block matrix of size $W \times W$, composed of 3×3 blocks[1]. Then, we review the original QZ algorithm for completeness and relevance to our higher dimensional case, a detailed treatment of which can be referred to in the original work [46]. Next we present the Block QZ algorithm, its advantages and pitfalls, and finally the Tensor QZ algorithm which is essentially a transformed version of the Block QZ algorithm.

[1]We stick to the development based on SrPDE, where the covariance between RGB channels results in 3×3 blocks per pixel. However, for a different application, the number of channels m must be used instead of 3 and the algorithm steps adapted accordingly.

7.1/ Spatial rPPG Distribution Estimation

For the application of Spatial rPPG Distribution Estimation (SrPDE), as mentioned earlier, pixel by pixel sequential application of the PVM algorithm, or in that case, any rPPG signal estimation algorithm, is a naive solution. It is worth noting here that for this specific application, an analytic method such as CHROM [58] can be quite easily scaled up for pixel-wise estimation. However, the CHROM method is essentially tailored towards rPPG measurement, whereas the PVM method is more generic in nature and can be easily adapted to other types of BSE problems. On the other hand, BSS and semi-BSE methods such as ICA, cICA and MAICA, have the complications of being optimization based and the required calculation of autocorrelation and chrominance functions and their derivatives, the combination of which makes these methods ill-suited for a higher dimensional transformation. Based on this arguments a higher dimensional analogue of the GEVD algorithm proves to be the most appropriate choice for this application.

With respect to the problem of SrPDE, for simplifying the computations we restrict the calculations to square video frames, which is quite feasible since we do crop each frame to the face, giving tensors of size $W \times W \times 3 \times 3$. Also, these covariance and periodic covariance tensors shall be calculated using the fundamental period obtained from the averaged signal initially. Nonetheless, it would indeed be interesting to compare the SrPDEs over different periods and build the final Spatial rPPG Distribution (SrPD), much like it was done using tabu search in chapter 6. This seems the most obvious course of future work, though in this chapter, we develop the core SrPDE algorithm.

To construct the input data, we start with the pixel-wise formulation of the PVM algorithm. The method extracts the quasi-periodic signal representing the cardiac pulse corresponding to a fundamental period τ, by constructing covariance and periodic covariance matrices $C_{\mathbf{x}}$ and $P_{\mathbf{x}}$, the details of which have already been discussed in chapter 6. For an RGB temporal signal of $\mathbf{x} \in \mathbb{R}^{3\times N}$ with N discrete samples, $C_{\mathbf{x}}$ and $P_{\mathbf{x}}$ are of size 3×3, which when estimated over each pixel result in a tensors $\mathcal{P}_{\mathbf{x}}$ and $\mathcal{C}_{\mathbf{x}}$ of size $W \times H \times 3 \times 3$ where $W \times H$ is the size of the video frame. For signal extraction scenarios other than the rPPG signal estimation, the tensor sizes will obviously be different, especially the number of channels might be more than 3. However, as mentioned earlier, the theory and algorithms developed in this chapter can be easily adapted to different channel sizes.

Before examining block diagonalization, it is worth mentioning here that there was an initial attempt at performing SrPDE using generic higher dimensional analysis using tensors. One of the algorithms attempted, but which was unsuited for the problem at hand was multilinear PCA (MPCA) [40]. It was chosen because of the fact that the PVM method can be thought of as a modified periodic PCA. This is not far from the truth, since PVM strives to find a subspace that maximizes the covariance as well as the periodic covariance of the underlying source data. In essence, a straightforward extension would be to perform multilinear PVM adapted from the algorithm of MPCA.

However, the issue with MPCA and Higher Order SVD (HOSVD) [44] for the case of SrPDE was due to the fact that these techniques are aimed to discover underlying relationships between the source variables just like PCA, except for in higher dimensions. As a result of this requirement, there is no possibility to treat a specific dimension of the source data, for instance the temporal dimension in the case of biomedical signals, as the desired one for calculation of the covariance. In other words, searching for highest

variance (and periodic variance) only in the temporal dimension is not feasible using MPCA or HOSVD, since it tries to maximize the covariance across all the dimensions. As a result, although a very interesting and seminal work in its own, MPCA could not be used for the problem of spatial rPPG distribution estimation and a different alternative was sought after.

In the next section, we dive into the details of formulating a strategy that effectively performs GEVD for the pair of lagged covariance and covariance matrices over each pixel, without performing it sequentially, pixel by pixel.

7.1.1/ Block Diagonalization

A viable option is to flatten the tensors $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$ into matrices, and rearrange them in a manner that emulates implicit parallelism. This formulation should effectively perform the diagonalization of the pixel-wise pair of 3×3 matrices, $P_\mathbf{x}^{(k,l,t)}$ and $C_\mathbf{x}^{(k,l,t)}$, where $k, l \in [1, 3]$ and $t \in [1, W^2]$ as before. Such a rearrangement of the tensors into matrices leads to the possibility of performing *block diagonalization*, which is not a new concept in itself, and has been shown to be equivalent to calculating the eigenvalues and eigenvectors [9]. It has indeed been used to solve the eigenvalue problem for certain applications [12], [15], [28] and [47] which, in fact, led to the idea of using it in SrPDE. It comprises of extending the general eigenvalue problem to calculation of block eigenvalues and eigenvectors [8] which are defined later in this section. Block diagonalization is defined only for matrices in $\mathbb{R}^2$ which leads us to the requirement of flattening the tensors $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$.

To elaborate, let us consider the $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$ tensors of size $W \times W \times 3 \times 3$, representing the spatial lagged covariances and covariances per pixel, calculated over N samples, corresponding to video frames of size $W \times W$. The lagged covariance and covariance matrices per pixel occupy the third and fourth dimension of $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$. Flattening these tensors will lead to the source matrices $\mathbb{P}_\mathbf{x}^\mathbf{s}$ and $\mathbb{C}_\mathbf{x}^\mathbf{s}$ of size $3W \times 3W$ as depicted on the left of figure 7.3. $\mathbb{P}_\mathbf{x}^\mathbf{s}$ and $\mathbb{C}_\mathbf{x}^\mathbf{s}$ are reshaped versions of the tensors $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$, where the 3×3 lagged covariance and covariance matrices per pixel, $P_\mathbf{x}$ and $C_\mathbf{x}$, are flattened onto the first and second dimensions resulting in block matrices of size $W \times W$ made up of 3×3 blocks. These block matrices are then rearranged into 3×3 block matrices, $\mathbb{P}_\mathbf{x}$ and $\mathbb{C}_\mathbf{x}$, with each block of size $W \times W$. This rearrangement is at the core of enabling the use of block diagonalization. Such an *interlaced* matrix is depicted on the right of figure 7.3. The block diagonalization of these interlaced block matrices, would then be equivalent to diagonalizing each pixel-wise pair of 3×3 blocks, which itself is equivalent to performing generalized eigenvalue decomposition on those pixel-wise pairs. The obvious advantage of this is the elimination of redundant steps performed per pixel.

We present here the details of the interlacing process. Let $\mathbb{P}_\mathbf{x}^\mathbf{s}$ and $\mathbb{C}_\mathbf{x}^\mathbf{s}$ represent the source matrices obtained by flattening $\mathcal{P}_\mathbf{x}$ and $C_\mathbf{x}$ as depicted on the left of figure 7.3. The interlacing operation $\mathcal{I}$ is depicted in figure 7.3. Any arbitrary element in the original matrix on the left can be depicted by $\mathbf{X}_{i,j}^{k,l}$, where the subscripts $i, j \in [1, 3]$ represent the indices in the 3×3 block and the superscripts $k, l \in [1, W]$ represent position of the 3×3 in the $W \times W$ block matrix. After translation the element $\mathbf{X}_{i,j}^{k,l}$ occupies a new position $\mathbf{X}_{k,l}^{i,j}$ of the now 3×3 block matrix with each block of size $W \times W$. That is, after interlacing, any block i, j of size $W \times W$ in the interlaced matrix on the right is occupied by the $(i, j)^{th}$ elements of the 3×3 block corresponding to all the pixels. For example, every top left

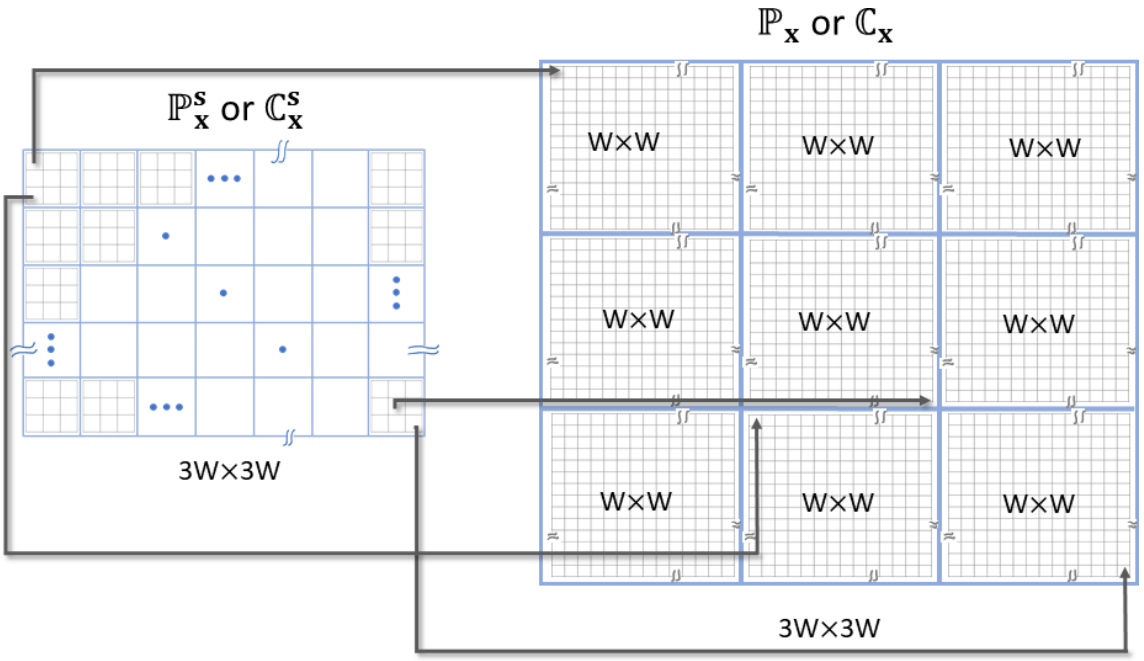


Figure 7.3: Interlacing the $W \times W$ matrix of 3×3 blocks (left) to a 3×3 matrix of $W \times W$ blocks (right)

element of each 3×3 block shall be grouped into the top left $W \times W$ block after interlacing.

$$\mathcal{I}(\mathbf{X}^{k,l}_{i,j}) = \mathbf{X}^{i,j}_{k,l} \tag{7.1}$$

The effect of this interlacing is that its block diagonalization produces zeros in all the blocks except the diagonal blocks. If this diagonalization follows the principles of eigenvalue decomposition using, for instance, unitary transformations, the block diagonal matrix should represent the block eigenvalues [9]. Block diagonalization and block eigenvalue decomposition has not seen much exposure in biomedical signals analysis, although the underlying theory is well established and dates a few decades. The basic idea behind the block eigenvalue problem is presented here [8],

Definition 4: Block Eigenvalue

A matrix X of order n is a block eigenvalue of order n of a matrix A of order mn if there exists a block vector V of full rank, such that $AV = VX$. V is a block eigenvector of A.

where the matrix A is block partitioned into $m \times m$ blocks of $n \times n$ matrices and the block vector V is of size $mn \times n$ [27]. Block diagonalization seeks to obtain a transformation under which matrix A is similar to a matrix D of the form

$$D = \begin{bmatrix} A_1 & 0_n \\ 0_n & A_2 \end{bmatrix} = S^{-1}AS, \tag{7.2}$$

where 0_n denotes the zero matrix of order n, and $m = 2$ in the above example. This definition can be naturally applied to the eigenvalue problem where D turns out to be the matrix of block eigenvalues of order n and S turns out to be the matrix of block eigenvectors of order $mn \times n$, V_i where $i \in [1, m]$.

We take the opportunity here to mention that use of Matrix Sector Functions (MSF) [13] for block diagonalization was initially investigated. MSF is an interesting fast algorithm that has one of its applications as the block diagonalization of square matrices. This algorithm was initially tested, but was deemed unsuitable due to it operating on a single matrix and its inability to control the size of the resultant block eigenvalues. The first issue required the conversion of the GEVD problem of two matrices A and B given by $Av = \lambda Bv$, where λ is an eigenvalue and v is the corresponding eigenvector, to the eigen decomposition of the single matrix $C = B^{-1}A$ which may be problematic. Added to the second issue of not being able to fix the block eigenvalue sizes, this algorithm could not be used for the application of SrPDE, although block diagonalization on synthetic matrices was quick and accurate.

Finally, owing to the lack of readily usable algorithms for block *generalized* eigenvalue decomposition it was decided to go back to the roots and directly extend the ubiquitously used QZ algorithm [46] for generalized eigenvalue decomposition. The rest of this section is organized as follows. We first present an overview of the original QZ algorithm for completeness and as the basis for the extended block and tensor based algorithms. Owing to its generic nature, these algorithms shall be easily adaptable to applications that need spatial coherence in conjunction with temporal signal mixtures, in addition to the application of SrPDE. Then we detail the formulation of the Block QZ algorithm with Block Householder reflections. Finally, these two block based algorithms are then optimized to their tensor-based equivalents. This optimized version was required owing to the iterative step in the QZ algorithm the details of which are presented in the next section.

7.1.2/ THE QZ ALGORITHM

The QZ algorithm was first introduced by Moser *et al.* to solve the generalized matrix eigenvalue problem $Ax = \lambda Bx$ with general square matrices A and B [46]. When B is not singular, this problem reduces to the eigenvalue decomposition problem $B^{-1}Av = \lambda v$. However, when B is non-invertible, the traditional eigenvalue decomposition algorithm cannot be used. Even when B is non-singular, there might be other issues such as the problem being ill-posed with a large difference in the the magnitude of the eigenvalues or e.g. when A and B have a common null space which reduces $\det(A - \lambda B)$ to zero. The QZ algorithm was born out of this necessity to be able to find the generalized eigenvalues of even near-singular matrices. It gets its name from the idea that there are unitary matrices Q and Z so that QAZ and QBZ are both upper triangular and that the corresponding eigenvalue problems $QAZy = \lambda QBZy$ and $Av = \lambda Bv$ are unitary equivalent. Owing to this equivalence, the eigenvalues for both the problems are identical and the eigenvectors fulfill the equation $v = Zy$.

At its core, the algorithm employs Householder transformations [3], which are unitary transformations that can be used to transform a matrix to its upper triangular form and are described in section 7.1.2.2. The algorithm performs the Householder transformations in a strategic manner so as to simultaneously triangularize a given pair of matrices, after which the generalized eigenvalues can be extracted. A summary of the QZ algorithm is presented here, followed by a detailed explanation and a pictorial aid in tables 7.1 and 7.2 in order to better understand the higher dimensional analogues.

The QZ algorithm comprises of four steps, which use Householder transformations to place zeros at strategic locations in the matrix. To summarize, the first stage comprises

of reducing the matrix A to upper *Hessenberg form*, while simultaneously reducing B to upper triangular form. A Hessenberg matrix is similar to an upper triangular matrix except that it has non-zero elements on its sub-diagonals. The second stage is a generalization of the implicit double shift QR algorithm [5] [7] which, incidentally, is used to calculate the eigenvalues of real and complex matrices. This reduces A to a quasi-triangular form, with no two consecutive sub-diagonal elements equal to zero, while maintaining the upper-triangularity of B. The third stage then takes care of reducing the quasi-triangular matrix A to triangular, thereby allowing the extraction of eigenvalues. Finally, the last stage comprises of extracting the eigenvectors from the triangular matrices and projecting them back on to the original coordinate space. Our block and tensor QZ algorithms essentially extrapolate these four stages to higher dimensions by operating on blocks of matrices and tensors at a time instead of a single matrix and vectors. Owing to the abundance of libraries for matrix operations and vectorization, such a formulation provides a computational advantage as compared to pixel-wise redundant GEVD. These algorithms are described in sections 7.2.1 and 7.2.2, after a review on Householder transformations, followed by its higher dimensional variants in section 7.1.2.2.

7.1.2.1/ ALGORITHM DETAILS

The QZ algorithm gets its name from the standard notation of orthogonal matrices Q, Z being the orthogonal matrices that introduce zeros in specific locations in the matrix. A Householder transformation $\mathscr{H}_r(k)$ transforms a vector such that there remain $r-1$ zeros after its k^{th} element. These vectors can be either columns or rows. When the vector is a column, the transformations Q_m are always pre-multiplications applied to rows k through $k+r-1$. When the vector is a row, the transformations Z_m are post-multiplications applied to columns k through $k+r-1$. These transformations are always applied on the pair of matrices. The algorithm does not actually provide the eigenvalues λ_i but gives α_i and β_i, the diagonal elements of QAZ and QBZ, Q and Z being the product of all the orthogonal transformations Q_m and Z_m respectively. The eigenvectors can then be calculated using the product of the accumulated Z_m. Table 7.1 summarizes the steps of the QZ algorithm. We present the salient parts of the QZ algorithm here so that the discussion of its higher dimensional analogues is coherent, followed by the detailed description of the Householder transformation in section 7.1.2.2.

Before looking at the algorithm in detail, it is worthwhile to describe the notations used in tables 7.1 and 7.2. The graphics in these tables serve as an overview of the state of the matrices before and after these operations. Note that the matrices of order $n = 6$ are used in graphics for relevance to the global comprehension. The elements that need to be reduced are highlighted by using a black square around them. The rows or columns that are being manipulated are indicated by arrows. For all the operations, the pre or post multiplications affected by Q and Z respectively, are applied to the pair of matrices. On the other hand, the depiction of iterative steps are admittedly difficult and would benefit from a better visual aid such as a moving picture. To make do, the iterative steps are depicted as recursing arrows pointing in the direction of increasing loop index. For an exhaustive explanation of each step, the reader is advised to refer to the original paper [46].

Stage 1 The first stage of the algorithm is quite straightforward and comprises of using Householder reflections to reduce B to triangular and A to upper Hessenberg form.

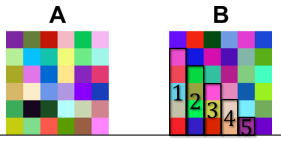
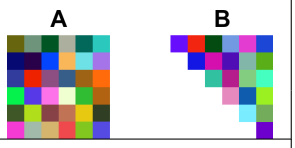
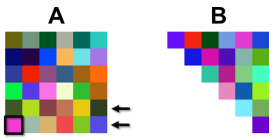
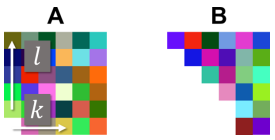
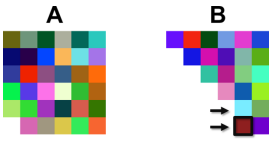
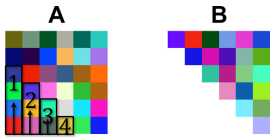
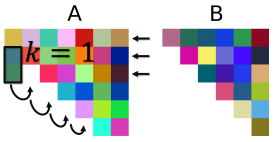
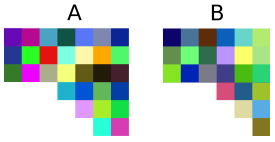
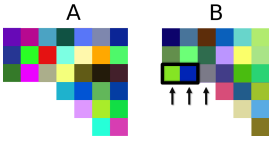
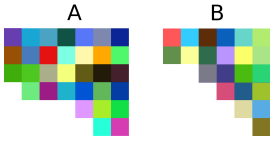
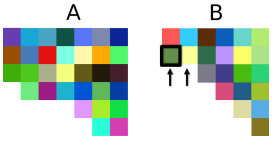
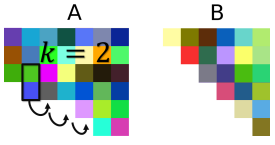
	Before	After
Reduce A to upper Hessenberg and B to triangular form		
1. For $k = 1, 2, \cdots, n-1$, (i) choose $Q_k \in \mathscr{H}_{n-k+1}(k)$ to nullify $b_{k+1,k}, b_{k+2,k}, \cdots, b_{n,k}$ (ii) $B \leftarrow Q_k B$, A $\leftarrow Q_k A$		
2. For $k = 1, 2, \cdots, n-2$, (i) For $l = n-1, n-2, \cdots, k+1$,		
(a) choose $Q_{kl} \in \mathscr{H}_2(l)$ to nullify $a_{l+1,k}$		
(b) $A \leftarrow Q_{kl} A, B \leftarrow Q_{kl} B$		
(c) choose $Z_{kl} \in \mathscr{H}_2(l)$ to nullify $b_{l+1,k}$		
(d) $B \leftarrow B Z_{kl}, A \leftarrow A Z_{kl}$		
For $k = 1, 2, \cdots, n-2$,		
Iterative strategy to reduce A to quasi-triangular, keeping B triangular		
1. Compute a_{10}, a_{20}, and a_{30}, using 7.5		
2. For $k = 1, 2, \cdots, n-2$,		
(i) determine $Q_k \in \mathscr{H}_3(k)$ to nullify $a_{k+1,k-1}$ and $a_{k+2,k-1}$		
(ii) determine $Z_k' \in \mathscr{H}_3(k)$ to nullify $b_{k+2,k+1}$ and $b_{k+2,k}$		
(iii) determine $Z_k'' \in \mathscr{H}_2(k)$ to nullify $b_{k+1,k}$		

Table 7.1: A summary of QZ algorithm to obtain generalized eigenvalues and eigenvectors of a square matrix of order n . The graphics on the right depict the state of the matrices before and after a given operation. Stages 1 and 2.

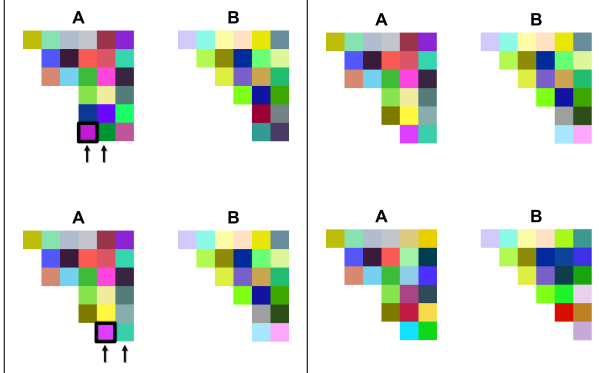
3. Determine $Q_{n-1} \in \mathscr{H}_2(n-1)$ to nullify $a_{n,n-2}$ 4. Determine $Z_{n-1} \in \mathscr{H}_2(n-1)$ to nullify $b_{n,n-1}$	
Reduce quasi-triangular A to triangular, keeping B triangular and extract the eigenvalues	
1. For every 2×2 block above the non-zero sub-diagonal elements (i) Compute the eigenvalue λ from the characteristic equation after applying the appropriate origin shift (ii) Choose $Z \in \mathscr{H}_2$ to nullify the first element of the larger of the two rows of $A - \lambda B$ (ii) Choose $Q \in \mathscr{H}_2$ to annihilate the $(2,1)$ element of the smaller of the two matrices of AZ and λBZ	
Compute the generalized eigenvectors of the reduced problem by back-substitution, as in the "$hqr2$" method [7]	

Table 7.2: Overview of the QZ Algorithm, continued. Stages 3 and 4

The triangularization of B is done by successive applications of Householder reflections $\mathscr{H}_{n-k+1}(k)$ for each column $k \in [1, n-1]$. Next, the modified A is then reduced to upper Hessenberg form iteratively nullifying elements below the diagonal denoted by $A_{(l,k)}$ where $l \in [n-1, k+1]$ for each column $k \in [1, n-2]$. Each elimination is performed by a row operations using Q which introduces a rogue element in the already triangular B. This rogue element is nullified by using an appropriate column operation (see step 2. of the first part in table 7.1). At the end of this process, A is converted to upper Hessenberg form and B is converted to upper triangular form.

Stage 2 The second stage comprises of executing an iterative scheme that reduces A to quasi upper-triangular form, in which no consecutive subdiagonal elements are non-zero, while keeping B triangular. To do so, the implicit shift method for the QR algorithm [5] is generalized to the QZ algorithm by using unitary equivalences. To justify the use of implicit shifts, it would be prudent to discuss the explicit QZ step, which approaches the reduction of A to upper triangular form by assuming that B is non-singular and examining the standard QR algorithm for $C = AB^{-1}$ which uses implicit shifts [2]. In brief,

1. The transformation Q is determined such that QC is upper triangular.

[2] A shift σ applied to a matrix C is obtained as $C - \sigma I$. The similarity to the eigenvalue problem is not coincidental here.

2. Then a Z is determined such that QAZ is upper Hessenberg and QBZ is upper triangular.
3. A and B are updated as QAZ and QBZ respectively.

This outline is then developed into a series of unitary transformations operating on A and B instead of AB^{-1} which fulfills the goal of transforming A and B to upper Hessenberg and upper triangular forms respectively. However this leads to the problem of potential instability [46], which in turn calls for the requirement of the implicit double shift step, which is employed to improve convergence of the QR algorithm and handle the instability of the explicit QZ step.

The main idea behind the *implicit shift* method generalized to the QZ algorithm using unitary equivalences is as follows. Supposing that A is upper Hessenberg and B is upper triangular, if Q and Z are unitary matrices such that QAZ and QBZ are upper triangular, then Q is determined entirely by its first row by the theorem in [6, p. 352], owing to the fact that AB^{-1} and $QAB^{-1}Q^*$ are both upper Hessenberg, where Q^* is the conjugate transpose of Q.

Thus, the iteration is built around two objectives. First, find the first row of Q, which is nothing but the first row that would be obtained from a double shifted QR applied to AB^{-1}. Second, determine Q and Z such that QAZ is upper Hessenberg, QBZ is upper triangular, maintaining that Q has the correct first row. With regards to the first objective, the calculation of the first two columns of AB^{-1} is easy since A is upper Hessenberg and B is upper triangular. Hence, Q can be determined by the first two columns of AB^{-1} in conjunction with the shifts. A small provision needs to be made when determining the Q_1 for the first column. Just as in the regular implicit shift QR algorithm, it is handy to think of Q_1 as the Householder transformation that nullifies two of the three non-zero elements in the fictitious zeroth column of A. Such a Q introduces unwanted elements at positions $(2,1)$, $(3,1)$ and $(3,2)$ in Q_1B (see first row of images on the right of table 7.1).

Then, for the second part, since we cannot pre-multiply so as to keep the first row intact, the unwanted elements in Q_1B can be reduced using two Z's, a Z_1' in $\mathscr{H}_3(1)$ which nullifies the $(3,1)$ and $(3,2)$ elements, and a Z_1'' in $\mathscr{H}_2(1)$ which nullifies the resulting $(2,1)$ element. Taking $Z_1 = Z_1'Z_1''$, we now have Q_1BZ_1 as upper triangular. This is depicted in the last two rows of table 7.1. Finally, when this Z_1 is applied to Q_1A, it results in two unwanted elements in the first column under its subdiagonal (last row of table 7.1, far right). This marks the restart of the loop where the two elements under the subdiagonal of the first column of Q_1AZ_1 (recall that we started with the fictitious zeroth column) need to be nullified using a Q_2, continuing the process and eventually *chasing* the unwanted non-zero elements towards the lower, right-hand corners. The iteration stops when no consecutive subdiagonal elements are non-zero. This just means that the matrix A is in a quasi-triangular form, where any non-zero element located on the subdiagonal will always have zeros on either side of it on the subdiagonal.

The fictitious zeroth column of A is determined analogously to the implicit double shift algorithm by taking shifts σ_1 and σ_2 to be the two zeros of the 2×2 problem

$$det(\bar{A} - \sigma\bar{B}) \tag{7.3}$$

where

$$\bar{A} = \begin{bmatrix} a_{n-1,n-1} & a_{n-1,n} \\ a_{n,n-1} & a_{n,n} \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} b_{n-1,n-1} & b_{n-1,n} \\ 0 & b_{n,n} \end{bmatrix} \tag{7.4}$$

σ_1 and σ_2 are not calculated using the quadratic polynomial to avoid the loss of non-negligible, off-diagonal elements in the shift calculation. Instead, the ratio of the three non-zero elements of the first column of $(AB^{-1} - \sigma_1 I)(AB^{-1} - \sigma_2 I)$ are directly calculated using the technique detailed in "hqr_2" [5], from formulas that involve only the differences of diagonal elements. These formulas are listed as follows for $q = n - 1$ for an order n square matrix.

$$\begin{aligned} a10 &= \left[\left(\frac{a_{qq}}{b_{qq}} - \frac{a_{11}}{b_{11}}\right)\left(\frac{a_{nn}}{b_{nn}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{qn}}{b_{nn}}\right)\left(\frac{a_{nq}}{b_{qq}}\right) + \left(\frac{a_{nq}}{b_{qq}}\right)\left(\frac{b_{qn}}{b_{nn}}\right)\left(\frac{a_{11}}{b_{11}}\right)\right] \cdot \left(\frac{b_{11}}{a_{21}}\right) \\ &\quad + \frac{a_{12}}{b_{22}} - \left(\frac{a_{11}}{b_{11}}\right)\left(\frac{b_{12}}{b_{22}}\right), \\ a20 &= \left(\frac{a_{22}}{b_{22}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{21}}{b_{11}}\right)\left(\frac{b_{12}}{b_{22}}\right) - \left(\frac{a_{qq}}{b_{qq}} - \frac{a_{11}}{b_{11}}\right) - \left(\frac{a_{nn}}{b_{nn}} - \frac{a_{11}}{b_{11}}\right) \\ &\quad + \left(\frac{a_{nq}}{b_{qq}}\right)\left(\frac{b_{qn}}{b_{nn}}\right), \\ a30 &= \frac{a_{32}}{b_{22}} \end{aligned} \tag{7.5}$$

Stage 3 The next stage of the algorithm further reduces the A to triangular form by tackling the 1×1 and 2×2 sub-problems, owing to the absence of any two consecutive non-zero elements in the subdiagonal of A. The procedure starts by taking λ as the eigenvalue of the problem $E = A - \lambda B$, makes an appropriate origin shift and computes the eigenvalue from the characteristic equation

$$\lambda = \mu + p + sgn(p) \cdot \sqrt{r} \tag{7.6}$$

where $\mu = \frac{a_{11}}{b_{11}}$, $p = .5(\frac{(a_{22} - \mu b_{22})}{b_{22}} - \frac{b_{12}a_{21}}{b_{11}b_{22}})$, $q = \frac{a_{21}(a_{12} - \mu b_{12})}{b_{11}b_{22}}$ and $r = p^2 + q$. Using this value of λ, Z is chosen to nullify the larger of the two rows of $A - \lambda B$ and Q is chosen to nullify the $(2, 1)$ element of the smaller of the two matrices of AZ and λBZ, ensuring that the computed $(2, 1)$ elements of QAZ and QBZ are negligible.

Stage 4 Finally, the calculation of the generalized eigenvectors is trivial and is performed by a back-substition process, extended from the method hqr_2 [5]. The eigenvectors of the original problem can then be obtained by applying the accumulated Zs. This application can be done on the fly, at the same time when the Zs are applied to the input matrices during the earlier steps of the algorithm.

We present the higher dimensional extension of the QZ algorithm in section 7.2.1. But before we can dive into the Block and Tensor QZ algorithms, we need to look at Householder transformations and their extension to the Block and Tensor case in section 7.1.2.2.

7.1.2.2/ HOUSEHOLDER TRANSFORMATIONS

Crucial to the QZ algorithm are Householder transformations [3] which are orthogonal transformations that introduce zeros at specific positions in a matrix. To extend the QZ algorithm to the block and tensor cases, it is imperative that the Householder transformations be extended accordingly.

Since applying unitary transformations to a matrix conserves its eigenvalues, it is advantageous to reduce the matrices to a triangular form, thereby making it possible to

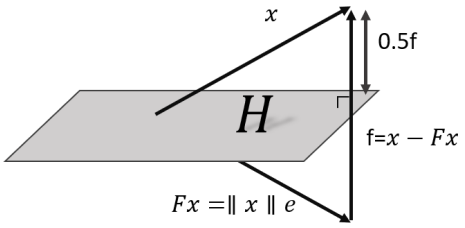


Figure 7.4: Householder reflection. F is the reflection across the hyperplane H that bisects the angle between x and Fx. $F = I - 2vv^T$ where $v = \frac{x-Fx}{\|x-Fx\|}$

compute the eigenvalues and eigenvectors, as seen in [7]. Unitary operators can be thought of as a complex generalization of rotations, since they preserve the inner product between vectors. Householder transformations, also called *reflections*, are orthogonal transformations that can be used to reduce a matrix to its upper triangular form. Such a reflection F can be chosen such that $A^{(k)} = FA$ has zeros in column k below the main diagonal.

Let $x = A - 1 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \in \mathbb{R}^3$ be the first column of $A \in \mathbb{R}^{3\times3}$ that needs to be reduced. Then, we seek the reflection F such that Fx is parallel to some unit vector e_1, i.e.,

$$Fx = \alpha e_1 = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} \tag{7.7}$$

where $e_1 = [1\,0\,0]^T$ and α is some constant. α can be obtained from considering two facts. First, that F is a reflection which preserves the length of x, which gives

$$\alpha = \pm \|x\| \tag{7.8}$$

Second, in order that the desired elements be nullified, F is supposed to *reflect* across some hyperplane H. This nullifying reflection can be achieved by choosing the sign of α to be the opposite of x_1. In other words,

$$\alpha = -sgn(x_1)\,\|x\| = s\,\|x\| \tag{7.9}$$

where $s = -sgn(x)$. It can further be seen from figure 7.4 that the reflection thus calculated also bisects the vector f where

$$f = x - \|x\|\,e = x - Fx = \begin{bmatrix} x_1 - \alpha \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 - s\,\|x\| \\ x_2 \\ x_3 \end{bmatrix} \tag{7.10}$$

In summary, the transformation F needed to zero all but the first element of x is the reflection across the hyperplane H orthogonal to $x - \|x\|\,e$. Let v be the normalized form of $x - Fx$ given by $v = \frac{x-Fx}{\|x-Fx\|}$. Then F is given by

$$F = I - 2vv^T \tag{7.11}$$

Equation 7.11 is derived here in order to better explain the algorithms for the block and tensor versions of the Householder transformation. Since x and $\|x\|\,e$ are of the same

length, they form the sides of an isosceles triangle. Consequently, the hyperplane H bisects the vector f orthogonally which means that $x - 0.5f$ is orthogonal to f, i.e.

$$(x - 0.5f) \cdot f^T = 0 \tag{7.12}$$

$$\Rightarrow \quad \mathbf{0.5} = \frac{xf^T}{\|f\|^2} \tag{7.13}$$

Now, using equation 7.13, the reflected $\|x\| e = x - f$, can be expressed as

$$\|x\| e = x - 2(\mathbf{0.5})f \tag{7.14}$$

$$= x - 2\frac{xf^T}{\|f\|^2} \cdot f \tag{7.15}$$

$$= x(I - 2\frac{ff^T}{\|f\|^2}) \tag{7.16}$$

Then, if we use $v = \frac{f}{\|f\|}$ as the normalized version of f, we obtain the householder reflection mentioned in equation 7.11 as

$$\|x\| e = Fx = x(I - 2vv^T) \tag{7.17}$$

The second term $\frac{2ff^T}{\|f\|^2}$ in equation 7.16, is worth a look. Recall from equation 7.10 that $f = \begin{bmatrix} u_1 & x_2 & x_3 \end{bmatrix}^T$, where $u_1 = x_1 - s\|x\|$. Keeping in mind that $s^2 = (-sgn(x))^2 = 1$ we have,

$$\begin{aligned} \|f\|^2 &= (x_1 - s\|x\|)^2 + {x_2}^2 + {x_3}^2 \\ &= 2\|x\|^2 - 2s\|x\| x_1 \end{aligned} \tag{7.18}$$

giving

$$\frac{2}{\|f\|^2} = \frac{1}{\|x\|(1 \cdot \|x\| - sx_1)} \tag{7.19}$$

$$\begin{aligned} &= \frac{1}{\|x\|((-s)(-s) \cdot \|x\| - sx_1)} \\ &= \frac{1}{-s\|x\|(x_1 - s\|x\|)} \\ &= \frac{-s}{\|x\| u_1} \end{aligned} \tag{7.20}$$

using the fact that $1 = s^2 = (-s)(-s)$, giving $\frac{1}{-s} = -s$.

Finally, the Householder reflection to transform the vector x can be obtained by calculating the scalar $\gamma = \frac{2}{\|f\|^2}$ by using equation 7.20, with $u_1 = x_1 - sgn(x_1)\|x\|$ and $f = x_1 - \|x\|$, then obtaining the matrix product ff^T and finally applying it as $x = x - \gamma(ff^T)x$.

All the formulation above was done considering the first column of a matrix in $\mathbb{R}^3$, in order to nullify all but the first element. However, the principle for reducing any number of elements of any arbitrary column remains the same. The general Householder reflection, $\mathscr{H}_r(k)$, for any column h that reduces the $r-1$ elements below row number k to zero, for an arbitrary sized matrix is listed in algorithm 2. The convention of reducing $r-1$ elements is taken to indicate that the element at row k is also part of the operation as in [46]. For instance, to reduce 2 elements starting at row k, a transformation in $\mathscr{H}_3(k)$ corresponds to

Algorithm 2 Householder transformation

1: **procedure** HOUSEHOLDER(A,k,r,h)
2: $\eta = \left\| A_{(k:k+r-1,h)} \right\|$
3: $s = -sgn(A_{(k,h)})$
4: $u_1 = A_{(k,h)} - s \cdot \eta$
5: $f = \begin{bmatrix} u_1 & A_{(k+1:k+r-1,h)} \end{bmatrix}^T$
6: $\gamma = \frac{2}{\|f\|^2}$
7: $A_{(k:k+r-1,:)} = A_{(k:k+r-1,:)} - \gamma \cdot (ff^T) \cdot (A_{(k:k+r-1,:)})$
8: **end procedure**

indices $k : k+r-1 = k : k+2$ that reduces $r-1 = 2$ elements. Note that the last operation is exactly the same as pre-multiplying A by Q that eliminates the desired elements.

With the overall picture of the QZ algorithm and Householder transformations, we can now proceed to detail the higher dimensional version of the QZ algorithm which is presented in the next section.

7.2/ PROPOSED HIGHER DIMENSIONAL VERSIONS OF THE QZ ALGORITHM

The Block QZ and Block Householder algorithms are aimed to eventually implement implicitly parallel computations, and perform GEVD on the interlaced matrix, which in turn would result in the block diagonalization of the pixel-wise 3×3 source matrices. The formulation of these Block QZ and Block Householder algorithms follows.

7.2.1/ BLOCK QZ ALGORITHM

Just as is the case with the basic QZ algorithm, Householder transformations are also at the core of the Block QZ algorithm. The Block QZ algorithm extracts the block generalized eigenvalues of the interlaced periodic covariance and covariance matrices, $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{C}_{\mathbf{x}}$, constructed by flattening the tensors $\mathcal{P}_{\mathbf{x}}$ and $\mathcal{C}_{\mathbf{x}}$ as visualized in section 7.1.1. Of course, the basic Householder transformations need to be equally extended to handle the interlaced block matrices $\mathbb{P}_{\mathbf{x}}$ and $\mathbb{C}_{\mathbf{x}}$ in a manner such that the reductions performed on these matrices be equivalent to reductions performed on the source matrices $\mathbb{P}^{\mathbf{s}}_{\mathbf{x}}$ and $\mathbb{C}^{\mathbf{s}}_{\mathbf{x}}$ representing pixel-wise periodic covariance and covariance matrices.

7.2.1.1/ BLOCK HOUSEHOLDER TRANSFORMATIONS

The extension of Householder transformations to the block case is fairly straightforward in essence. Apart from step 7 in algorithm 2 which is a matrix product, all the other steps are operations on scalars, which can be easily adapted to element-wise operations. The matrix product itself can be performed by using Hadamard products which we detail later in this section.

Consider the case of reducing the last two elements of the first column of each 3×3

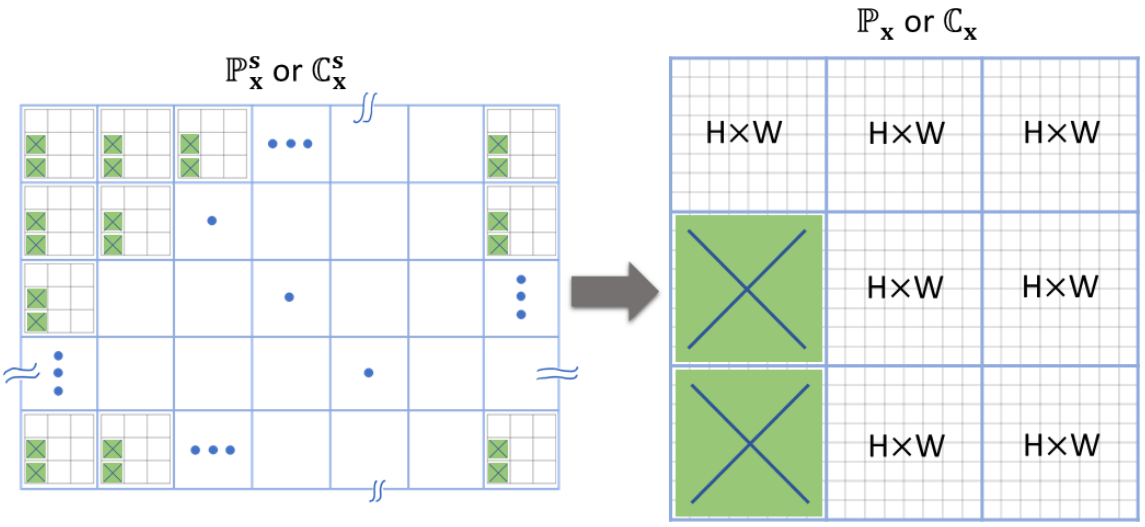


Figure 7.5: The Block Householder transformation. The elimination of the last two elements in the first column of each 3×3 block (left) is equivalent to eliminating the last two $W \times W$ blocks in the first block column (right).

block of the pixel-wise covariance matrix $\mathbb{C}^{\mathbf{s}}_{\mathbf{x}}$ in figure 7.5. Recall that to reduce 2 elements, we take $r = 3$ to incorporate the first element with respect to which the elimination is performed. As a result, we have $k = 1, r = 3$ and $h = 1$ for the procedure HOUSEHOLDER in algorithm 2, repeated over each 3×3 block of $\mathbb{C}^{\mathbf{s}}_{\mathbf{x}}$. The equivalent block operation on the interlaced matrix $\mathbb{C}_{\mathbf{x}}$ would be to reduce the two leftmost $W \times W$ blocks corresponding to R rows and C columns, where $R \in [W+1, W+2, \cdots, 3W]$ and $C \in [1, 2, \cdots, W]$ as seen in figure 7.5. As far as the algorithm is concerned, operations on scalar elements are now transformed to operations on blocks. This means that the index position $k = [1, 2, 3]$ is going to start at intervals of W to mark the beginning of a block, instead of being sequential. In fact, it is sequential, but with respect to blocks: it is *block sequential*. Consequently, k transformed to the block index position $k_\nabla = (k-1)W + 1$. The number of elements to be reduced r is transformed to $r_\nabla = rW$ representing the range of rows to be reduced, and the scalar h representing the column is transformed to a vector $h_\nabla = [h-1, h]W$ representing a set of columns. Together, r_∇ and h_∇ represent the block to be reduced, with k_∇ representing the starting index of the block.

Consequently, the scalars η, s, u_1 and γ in algorithm 2 are transformed to their block analogue matrices $\eta_\nabla, s_\nabla, u_{1_\nabla}$ and γ_∇ of size $W \times W$ each and the vector f is transformed to its block vector analogue f_∇ of size $3W \times W$. All the arithmetic operations are extended from scalar operations to to element-wise operations between the block analogue matrices. The matrix product operations required for calculating η_∇ in step 2 and the product $\gamma_\nabla \cdot (f_\nabla f_\nabla^T).(A_{k_\nabla:k_\nabla+r_\nabla-1,:})$ are non-trivial and need elaboration.

The calculation of the block norm η_∇ is just a simple element-wise extension given by

$$\eta_\nabla = \sqrt{\sum_{k=1}^{3} A_{k_\nabla:k_\nabla+r_\nabla-1,h_\nabla} \odot A_{k_\nabla:k_\nabla+r_\nabla-1,h_\nabla}} \tag{7.21}$$

where $A_{k_\nabla:k_\nabla+r_\nabla-1,c}$ represents the block vector from rows $(k-1)W+1 : (k+r-1)W$ and columns $(h-1)W+1 : hW$, h is the column block number, and r is the number of $W \times W$ blocks to be reduced and $\odot$ is the Hadamard product.

Finally, to compute the product $\gamma_\nabla \cdot (f_\nabla f_B^T).(A_{k_\nabla:k_\nabla+r_\nabla-1,:})$, we first reduce the product $\gamma \cdot (ff^T) \cdot (A_{(k:k+r-1,:)})$ from matrix multiplication to sum of Hadamard products. This can then be extended for each element for the block case. Note that the normal matrix product here is not suitable since it will loose the spatial coherence between the source 3×3 blocks, i.e. the product will not be the same as if it was performed sequentially for each 3×3 block. Let $k = 1, r = 3$ and $h = 1$ and let $A = [a_{(k,l)}]$ where $k, l \in [1, 3]$ be the 3×3 matrix.Then the vector $f = \begin{bmatrix} u_1 & a_{21} & a_{31} \end{bmatrix}^T$, which gives

$$\gamma \cdot (ff^T) \cdot (A_{(k:k+r-1,:)}) = \gamma \begin{bmatrix} u_1 \\ a_{21} \\ a_{31} \end{bmatrix} \begin{bmatrix} u_1 & a_{21} & a_{31} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{7.22}$$

$$= \begin{bmatrix} \gamma u_1 u_1 & \gamma u_1 a_{21} & \gamma u_1 a_{31} \\ \gamma a_{21} u_1 & \gamma a_{21} a_{21} & \gamma a_{21} a_{31} \\ \gamma a_{31} u_1 & \gamma a_{31} a_{21} & \gamma a_{31} a_{31} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{7.23}$$

The product of the two matrices in equation 7.23 can then be represented as the sum across the third dimension of the folded third order tensor as shown in figure 7.6.

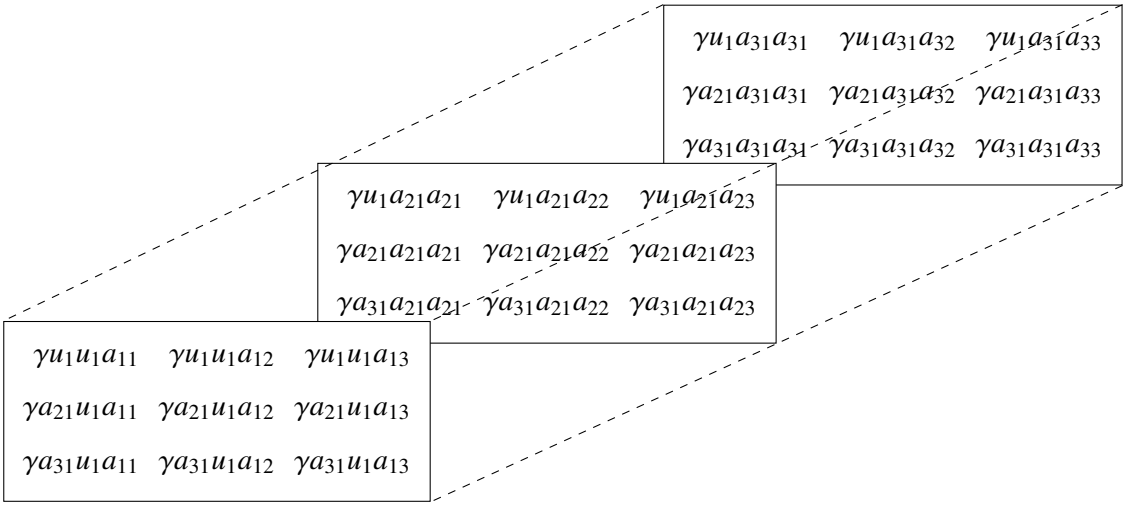


Figure 7.6: Matrix product using Hadamard product

Finally, the matrix product $\gamma \cdot (ff^T) \cdot (A_{(k:k+r-1,:)})$ is obtained by summing the tensor thus obtained in the third dimension. This conversion of the matrix product into sum of Hadamard products is advantageous when the scalars need to be extended to arbitrary sized matrices. Since the Hadamard product is invariant to the matrix size, each element of the tensor representation can be a matrix block of size $W \times W$ and the product $\gamma \cdot (ff^T) \cdot (A_{(k:k+r-1,:)})$ can be simultaneously calculated for all the pixels in a given block vector.

The block Householder algorithm works perfectly, reducing any desired $W \times W$ block on the lower left interlaced $3W \times 3W$ matrix. Coincidently, the first stage of the QZ algorithm, which is just a series of Householder transformations on different rows and columns, can be extended to apply to interlaced block matrices $\mathbb{B}_\nabla$ and $\mathbb{A}_\nabla$ to reduce them to upper triangular and upper Hessenberg forms respectively. The first stage works successfully as expected and scales well as the size of the matrix increases. Preliminary tests with 90×90 and 300×300 reduced computation times by 10 and 16 times respectively as compared

to sequential triangularization for each 3×3 block which took 0.58 and 5.8 seconds, while the block version took 0.05 and 0.35 seconds $W = 90$ and $W = 300$, respectively.

The problem, however, arises in the second stage of the QZ algorithm which is iterative. For the source matrices, the number of iterations required to reduce each pair of 3×3 matrices to quasi-triangular and triangular forms can be arbitrary. This is due to the fact that the iterative strategy depicted in the second part of table 7.1 needs to be repeated until no consecutive subdiagonals of A, in the scalar case, are non-zero. For the block case, this means that certain elements in the $W \times W$ blocks will be reduced to zero earlier than others. In which case, the algorithm should avoid any operations on those elements. However, given the structure of our block matrices, such selection of elements, although not impossible, is not intuitive, even inelegant. In effect, binary masks of size $W \times W$ were used to select the elements to be operated upon. Although this solution works and provides accurate results, it does have issues. For instance, the algorithm needs to run for as many iterations as it takes to reduce the 3×3 block that requires the maximum number of iterations, which of course adds that many more mask multiplications per step of each Block Householder transformations. Moreover, the formulation of matrix product using Hadamard product seems complicated, even though it works accurately and calls for elucidation. A simpler alternative should be possible.

This infeasibility of the Block Householder transformations for the iterative strategy of the QZ algorithm prompted us to approach the problem from a different angle. Specifically, a higher dimensional representation which would allow us to operate on all the elements simultaneously, while providing the choice to select and deselect specific elements without requirement of complex formulations. A tensor representation of the 3×3 matrices corresponding to each pixel is one such representation that fulfills our criteria. Next, we present the Tensor QZ algorithm in section 7.2.2, starting with the formulation of the Tensor Householder transformation in section 7.2.2.1, which is a crucial part in the development of the former, just like its scalar counterparts.

7.2.2/ Tensor QZ Algorithm

While Block Householder transformations deal with flattened interlaced matrices (figure 7.3), Tensor Householder transformations operate on 3rd order tensors. The Tensor Householder transformation for treating the tensor representations of the non-interlaced matrices follows.

7.2.2.1/ Tensor Householder Transformations

The source tensors of size $W \times W \times 3 \times 3$, with each block of size 3×3 corresponding to a pixel are rearranged into a tensor of size $3 \times 3 \times W^2$ as shown in figure 7.7. Recall from figure 7.2 that each 3×3 block is represented by $a_{(k,l)_w}$, where $k, l \in [1, 3]$ and $w \in [1, W^2]$. This can be thought of as being analogous to vectorizing a matrix. The advantage of this structure is that blocks corresponding to different pixels can be referenced by a single index in the third dimension as opposed to a range of indices in the interlaced block representation. This simplifies the selection of specific blocks to operate upon as compared to the flattened block matrices without much complication. This is because vector representations can be better optimized at the machine level as is seen in the vectorized operations of MATLAB [95]. For other languages, existing data structures such as

linked lists can be employed in the implementation to select and deselect required blocks (nodes) without additional overhead. Finally, this vectorized tensor has the advantage of simplifying the last step of Householder algorithm, which was understandably complex for the Block version. The MATLAB library MTIMESX [55], which is optimized for fast matrix multiplication, also supports multi-dimensional products and coincidentally provides improved computation times for the tensor product for the last step of the Householder algorithm. The details of the Tensor Householder transformation are as follows.

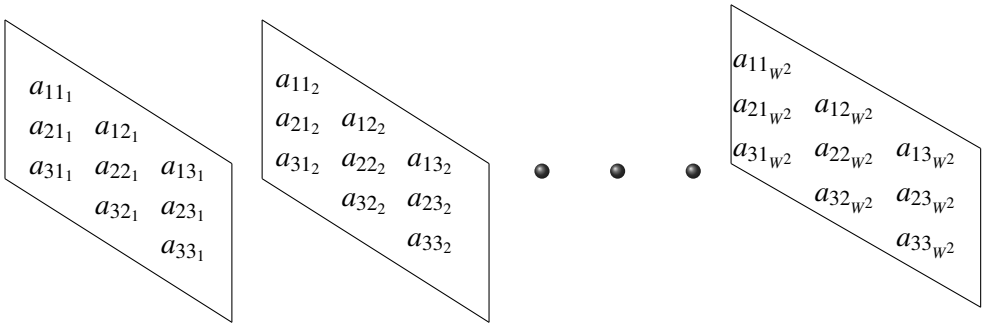


Figure 7.7: The Vectorized representation of the $W \times W \times 3 \times 3$ tensor $C_{\mathbf{x}}$. The w^{th} block $A_w = [a_{k,l}]$ where $k \in [1,3], l \in [1,3]$, represents the covariance or periodic covariance of pixel $(w/W) + (w \, mod \, W)$, where mod is the modulo operation giving the remainder of w/W.

Compared to the block version, the tensor based Householder algorithm does not require any conversion between the indices of the elements to be operated in question. Specifically, the scalars in algorithm 2, viz. k, r and c corresponding to the row, the number of elements to be nullified and column respectively, remain unchanged. This is because the index of the 3×3 block to be operated upon is along the third dimension and thus does not affect the block wise operations.

Consequently, as envisioned in figure 7.2 at the beginning of this chapter, the scalars η, s, u_1 and γ become vectors $\eta_\diamond, s_\diamond, u_{1_\diamond}$ and $\gamma_\diamond$; the 3×1 vector f becomes the matrix, $f_\diamond$. The final step of calculating the product $\gamma_\diamond \cdot (f_\diamond f_\diamond^T) \cdot \mathcal{A}_{(k:k+r-1,:,:)}$ is performed using the MTIMESX library [55] leveraging low level calls and multi-threading. The operation itself is a product of block wise Hadamard products to the tensor $\mathcal{A}_{(k:k+r-1,:,:)}$. We emphasize that use of such an optimized library is possible because of the simpler structure which is not the case with the Block Householder algorithm, where use of such a library was not possible.

With the tensor Householder algorithm in place, the realization of the Tensor QZ algorithm is quite straightforward. The four stages of the original QZ algorithm shown in table 7.1 are described as follows for the Tensor case.

7.2.2.2/ STAGE 1

The first stage of reducing all the 3×3 blocks of the $3 \times 3 \times W^2$ tensors $\mathcal{A}$ and $\mathcal{B}$ to upper Hessenberg and to triangular form respectively using Tensor Householder Transformations is simple and provides a computational advantage similar to the Block Householder implementation. To reiterate, the sequence of Tensor Householder transformations followed in this stage are analogous to those in the first stage of algorithm 7.1. This step reduces all the 3×3 blocks $\mathcal{B}_{(:,:,s)}$ to upper triangular and all the $\mathcal{A}_{(:,:,s)}$ to upper Hessenberg

form respectively.

7.2.2.3/ STAGE 2

As expected, the iterative strategy benefits from this new formulation. The scalars a_{10}, a_{20} and a_{30} representing the first three elements of the fictitious first column of each 3×3 block of $\mathcal{A}$ become vectors now. Sequential Tensor Householder transformations following the strategy listed in algorithm 7.1 are applied and the whole process repeated for the matrices $A_i = \mathcal{A}_{(:,:,s)}$ which do not yet have consecutive non-zero elements in their subdiagonals. This ensures that the matrices which have already been reduced to the desired forms can be easily deselected and redundancy be avoided. Of course, the number of iterations is still equal to those corresponding to the 3×3 matrix that requires the maximum iterations, but unnecessary operations on the other matrix blocks are avoided as soon as they are reduced. After the iterative step, all the 3×3 blocks of $\mathcal{A}$ are reduced to a quasi-triangular form, while all the 3×3 blocks of $\mathcal{B}$ still staying upper triangular. As mentioned earlier, the quasi-triangular form has no consecutive non-zero subdiagonals. This means that for a 3×3 block, either the elements $(2,1)$ and $(3,1)$ are zero, or the elements $(3,1)$ and $3,2$ are zero as shown below in equation 7.24.

$$A_i = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \quad or \quad \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & 0 & \times \end{bmatrix} \tag{7.24}$$

7.2.2.4/ STAGE 3

This simplifies the problem to reducing the remaining 2×2 blocks, either at position $(2,2)$ or at position $(1,1)$, i.e. at position 2 or 1 in the diagonal for the matrices shown in equation 7.24. However, these positions will definitely vary for each A_w in $\mathcal{A}$. Fortunately, the indices of the 2×2 block in each A_w can be easily found by a vector p of length W^2 having its values $p_i \in [1,2]$ based on the position of the 2×2 block in each A_i. This vector can be obtained by the following algorithm, which can be implemented easily using vectorized comparisons, for instance, in MATLAB.

Algorithm 3 Find the position of the 2×2 blocks in $\mathcal{A} \in \mathbb{R}^{3\times3\times W^2}$

1: $a_{21} = \mathcal{A}_{(2,1,:)}$
2: $a_{32} = \mathcal{A}_{(3,2,:)}$
3: For all the index positions z_{21} where $a_{21} == 0$, set $p(z_{21}) = 1$
4: For all the index positions z_{32} where $a_{32} == 0$, set $p(z_{32}) = 2$

Using the indices in p, a sub-tensor $\mathcal{A}_s$ of size $2 \times 2 \times W^2$ can be operated upon for the reduction of said 2×2 blocks. Then, the computation of the scalar eigenvalue λ using equation 7.6 is as simple as the element-wise operations similar to those used in the Tensor and Block Householder algorithms.

Next, norms are calculated of each pair of rows in the sub-tensor $\mathcal{A}_s$ along the third dimension giving a norm matrix of size $2 \times W^2$ using which the rows of each block of $\mathcal{A} - \lambda_\diamond \odot \mathcal{B}$ can be compared vectorally similar to algorithm 3, where $\lambda_\diamond$ is the vector representing all the λ in equation 7.6 described in stage 3 of the QZ algorithm in section

7.1.2. As earlier, this comparison gives the indices for each block depicting the larger of the two rows. A Tensor Householder operation with $\mathcal{Z} \in \mathscr{H}_2$ is applied to eliminate the first elements of all these rows. Next, the Hadamard products $\mathcal{A} \odot \mathcal{Z}$ and $\lambda\mathcal{B} \odot \mathcal{Z}$ are calculated and their block-wise norms compared vectorally. The resulting indices indicate which matrix blocks between the two tensors are smaller. A Tensor Householder $\mathcal{Q} \in \mathscr{H}_2$ is applied to nullify the $(2,1)$ elements of these blocks. The tensors $\mathcal{A}$ and $\mathcal{B}$ are now reduced such that the generalized eigenvalues can be calculated from the ratio of the respective diagonal elements α_η and β_η in each pair of matrix blocks A_w and B_w, where $\eta \in [1, \cdots, N]$ and $s \in [1, \cdots, W^2]$.

7.2.2.5/ STAGE 4

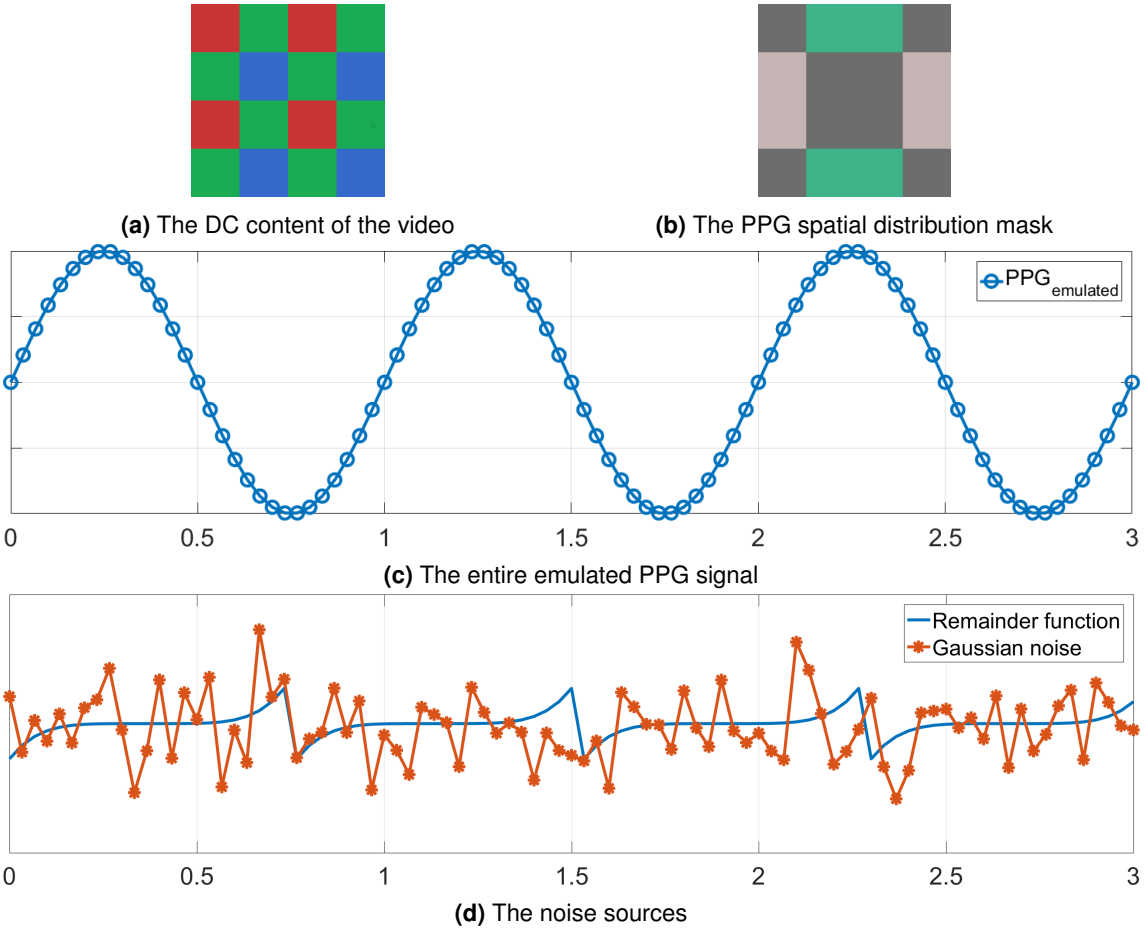
Finally, as in the scalar case, the calculation of the generalized eigenvectors is done on the fly, while applying the $\mathcal{Q}$ and $\mathcal{Z}$, along $w \in [1, W^2]$, on the tensors $\mathcal{A}$ and $\mathcal{B}$ in the last step of the Tensor Householder algorithm. Finally, the Tensor QZ algorithm can be used to estimate the spatial rPPG distribution of a video containing skin pixels using spatial covariance and periodic covariance tensors $\mathcal{C}_\mathbf{x}$ and $\mathcal{P}_\mathbf{x}$. The feasibility analysis, experiments and results are next presented in section 7.3.

7.3/ EXPERIMENTS AND RESULTS

7.3.1/ SYNTHETIC EXAMPLE

To analyze the effectiveness of the Tensor QZ algorithm, we tested it against a synthetic example as a proof of concept. A synthetic video that has an underlying quasi-periodic temporal signal embedded into its frames would serve our requirement. To keep things simple and easy to verify, we started with a static image of 4×4 pixels representing the static or DC signal content as shown in figure 7.8a. The periodic or AC component, $PPG_{emulated}$ was generated using a sinusoid of 1 Hz with a sampling frequency of $F_s = 30$ Hz and an amplitude of 0.2, with a duration of 3 seconds, resulting in a signal of $N = 90$ samples, which is depicted in figure 7.8c.

This periodic source $PPG_{emulated}$, representing the cardiac signal was then perturbed using two noise sources, $N_1 = .2*((rem(v, 23) - 11)/9)^5$, where $v = [0, N-1]$ and $rem(\cdot)$ is the remainder function, and a uniform random noise signal N_2, with a standard deviation of 0.5 which are shown in figure 7.8d. Note that N_1 is itself a periodic signal which provides a basis for robust validation for the algorithm. The perturbation was performed by mixing the three signals using a random weighting matrix of 3×3 to obtain a mixture of size $3 \times N$ as depicted in figure 7.9a. This emulated RGB signal was then repeated over the static image matrix of 4×4 pixels, scaled according to a spatial distribution pattern, shown in figure 7.8b, and overlaid onto the static image. Note that there are three distinct regions in this spatial PPG pattern which our algorithm should recover from the mixture. One cycle corresponding to 1 second of the resultant video and the corresponding temporal signals are shown in figures 7.9a and 7.9b. This source video was then analyzed using the Tensor QZ algorithm and compared against pixel-wise application of the PVM algorithm which is discussed next in section 7.3.2.



(a) The DC content of the video

(b) The PPG spatial distribution mask

(c) The entire emulated PPG signal

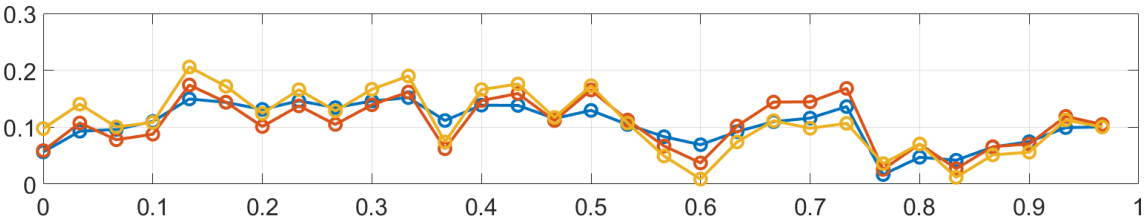
(d) The noise sources

Figure 7.8: Synthetic example to test the Tensor QZ Algorithm, a PPG signal of 1 Hz at a sampling rate of $F_s = 30$ with 3 periods.

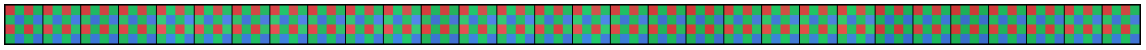
7.3.2/ TENSOR QZ ON THE SYNTHETIC VIDEO

In this section, we describe the process used to apply the Tensor QZ algorithm on a synthetic video. The procedure is same for applying the algorithm on a real video to perform SrPDE, except for the change in matrix size W and the corresponding increase in computation times.

1. **Acquisition:** The first step is to prepare the input data. For this, the frames from the synthetic video are loaded into a tensor $\mathcal{X}_{\mathbf{p}}$, from which the temporal average tensor $\overline{\mathcal{X}}_{\mathbf{p}} = \frac{1}{N}\sum \mathcal{X}_{\mathbf{p}}$ is then calculated, where N is the total number of frames, and $\mathcal{X}_{\mathbf{p}}$ and $\overline{\mathcal{X}}_{\mathbf{p}}$ are tensors of size $W \times W \times 3 \times N$ and $W \times W \times 3$ respectively. The suffix $\mathbf{p}$ represents the padded version of these tensors based on the requirement of calculating the lagged covariance, and in order that the output video be the same size as the input video. The Tensor QZ algorithm is applied on the padded input data as a result of which the final output shall be a subset of this padded output video. Of course, this would not be necessary for a real video where the whole process would be repeated over a temporal window of, say 5 seconds, in which case, there is an initial delay in the estimation of the SrPDE. In the synthetic case, frames from the start of the video are padded at the end to maintain quasi-periodicity and the



(a) One cycle (1s) of the sources $PPG_{emulated}$, N_1 and N_2 mixed using a random weighting matrix of size 3×3



(b) One cycle of the synthetic input video

Figure 7.9: Synthetic example to test the Tensor QZ Algorithm, the source mixtures and video.

final output video extracted from the padded output. In order to accommodate the estimation of all the possible pulse rates, the number of frames to be padded is determined by $N_p = \tau_{max} F_s$ where $\tau_{max} = (\frac{60}{HR_{min}})^{-1}$ corresponding to the period of the minimum heart rate of 40 bpm used in all our experiments. Accordingly, τ_{min} corresponds to the lag related to the maximum heart rate of 200 bpm, giving the range of lags as $[\tau_{max} F_s, \tau_{min} F_s]$ for the heart rate range of $[40, 200]$ bpm.

2. Preprocessing: The next step is to pre-process the input data by performing centering by $\widetilde{\mathcal{X}}_{\mathbf{p}} = \mathcal{X}_{\mathbf{p}} - \overline{\mathcal{X}}_{\mathbf{p}}$, followed by whitening, using $\widetilde{\mathcal{X}}_{\mathbf{p}} = \widetilde{\mathcal{X}}_{\mathbf{p}} / \sigma_{\diamond}$, where $\sigma_{\diamond}$ is a tensor of size $W \times W \times 3$ representing the standard deviation of each pixel along the temporal dimension, for each of the three RGB channels. Note that whitening is not a necessary step here, although it helps to scale the resultant pixel-wise rPPG signals for better visualization. Next, the tensors $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{X}}_{\tau^*}$ are extracted from $\widetilde{\mathcal{X}}_{\mathbf{p}}$ representing a a window of frames and its lagged version. τ^* here corresponds to the period obtained by applying PVM on the spatially averaged $\widetilde{\mathcal{X}}_{\mathbf{p}}$ based on the assumption that the underlying cardiac pulse is fairly uniform in a small spatial region such as the face. This reasoning is further elaborated in section 7.3.3. The preprocessing is then finalized by reshaping $\widetilde{\mathcal{X}}$ and $\widetilde{\mathcal{X}}_{\tau}$ from $W \times W \times 3 \times N$ to $3 \times N \times W^2$ tensors as required by the Tensor QZ algorithm.

3. $\mathcal{P}_{\mathbf{x}}$ and $C_{\mathbf{x}}$: Next, the lagged covariance and covariance matrices are calculated as

$$\mathcal{P}_{\mathbf{x}} = \widetilde{\mathcal{X}} \diamond \widetilde{\mathcal{X}}_{\tau}^{\mathcal{T}} \quad \text{and} \quad C_{\mathbf{x}} = \widetilde{\mathcal{X}} \diamond \widetilde{\mathcal{X}}^{\mathcal{T}} \tag{7.25}$$

where $\diamond$ represents the block-wise product of two third order tensors along the third (temporal) dimension and $(\cdot)^{\mathcal{T}}$ is the transpose on each 3×3 block performed along the temporal dimension over $[1, N]$.

4. Tensor GEVD: Finally, Tensor GEVD is performed using the Tensor QZ algorithm on the pair $\mathcal{P}_{\mathbf{x}}$ and $C_{\mathbf{x}}$ which gives the eigentensor $\mathcal{V}^*$ representing the block-wise eigenvectors that maximize periodic variance of each block in the temporal dimension. $\mathcal{V}^*$ is of size $1 \times 3 \times W^2$ on which the original data tensor $\widetilde{\mathcal{X}}$ is then projected to obtain the output tensor, $\widetilde{\mathcal{X}}_{\mathbf{o}} = \mathcal{V}^* \diamond \widetilde{\mathcal{X}}$. This is done by applying each 1×3 vector (representing W^T in the PVM algorithm) on each 3×3 block of the tensor $\widetilde{\mathcal{X}}$ [3].

[3] As mentioned earlier in this chapter, all the matrix multiplications on the blocks inside the tensors were performed using the MTIMESX [55] library which provides the possibility to perform matrix multiplications on select dimensions of tensors.

7.3.3/ RESULTS ANALYSIS

Using the steps listed in the previous section, the underlying quasi-periodic signal was successfully extracted using the Tensor QZ algorithm using the lagged-covariance and covariance tensors $\mathcal{P}_{\mathbf{x}}$ and $\mathcal{C}_{\mathbf{x}}$. It was mentioned that the tensor $\mathcal{P}_{\mathbf{x}}$ was built by applying the τ^* obtained by using the PVM algorithm on the spatially averaged signal since this was a proof of concept experiment. This formulation corresponds to the assumption that in a small patch of the skin, e.g. the facial region, there is hardly any phase difference between the rPPG obtained from different parts of this patch. However, a complete application of the SrPDE can equally be performed for each pixel, over the range of τ corresponding to the human heart rate $[40, 200]$ bpm, and the pixel wise best weighting matrix can be subsequently selected in the same manner as the iterative step of the Tensor QZ algorithm in section 7.2.2. This *Tensor PVM* algorithm can be thought of as the higher dimensional analogue of the PVM algorithm that maximizes the pixel-wise periodic variance using the corresponding pixel-wise τ^* instead of one obtained from the spatially averaged temporal signal. This aspect of the experiment is however left for the future work and shall be the most obvious course of further validation of the Tensor QZ algorithm.

That being said, the results of the Tensor QZ algorithm on the synthetic experiment are indeed as desired. Specifically, the comparison with the pixel-by-pixel application of the PVM algorithm confirmed the expected accuracy of the Tensor QZ algorithm. Furthermore, regarding computation times, even with the substantially small size of 4×4, the Tensor QZ algorithm reports improvements of more than 5 times taking 0.35 seconds to process a temporal window of 3 seconds as compared to the 1.96 seconds taken by the pixel-wise application of PVM. Moreover, the improved scalability of the Tensor QZ algorithm and the corresponding reduction in the computational complexity for larger matrices has already been discussed in section 7.2.1.1, where performance improvements of up to 10 and 14 times for the first step of the algorithm, as compared to the sequential application of PVM, were reported for larger matrices of sizes 90×90 and 300×300 respectively. These improvements shall undoubtedly be carried over to the application on real video data as well. We next describe the results obtained by the Tensor QZ algorithm on the synthetic example set up in the previous section.

It has been established that the Tensor QZ algorithm is superior to application of pixel-wise PVM. But this improvement in computational efficiency must be coupled with appropriate accuracy to validate the overall effectiveness of the Tensor QZ algorithm. In this context, the following criteria should be fulfilled.

- The pixel-wise temporal signals should be similar to those obtained by the pixel-wise application of the PVM algorithm. This similarity is clearly visible in figures 7.10d and 7.10f. Furthermore, an FFT analysis of the pixel-wise output signals is also presented in figure 7.11, where the frequency of the underlying quasi-periodic signal (shown in blue) has been correctly extracted for each pixel at 0.99 Hz which is in accordance with the pixel-wise FFT of the signal obtained using PVM (shown in green). Furthermore, the spatial average of all the extracted signals should be close to the $PPG_{emulated}$. This is indeed clearly visible from figure 7.10d and 7.10f.
- The spatial rPPG distribution should be revealed in the resulting video. In our synthetic example, the spatial distribution of the PPG was emulated using amplitude scaling across pixels using the pattern shown in figure 7.8b. This should result in distinct rPPG signals in $\widetilde{\mathcal{X}}_{\mathbf{o}}$ corresponding to different spatial regions. This result is

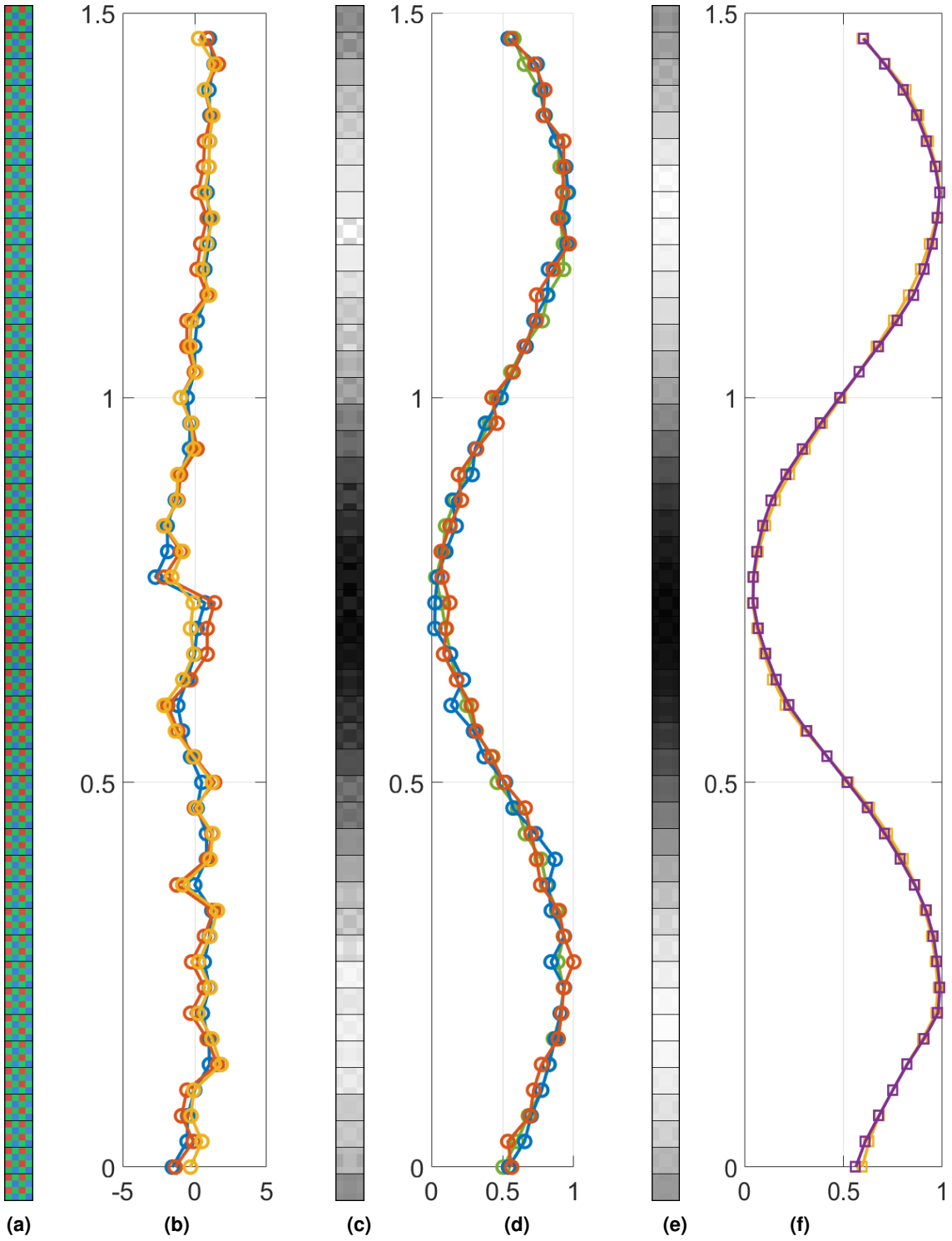


Figure 7.10: Synthetic example for emulating SrPDE. The vertical frames correspond to 1.5 cycles (45 frames) of the total 3 cycles (91 frames). (a) The source video (b) The spatially averaged RGB waveforms (c) The result of the Tensor QZ (d) The pixel-wise waveforms after Tensor QZ (e) Frames after pixel-wise PVM application (f)Pixel-wise PVM waveforms

clearly visible in figure 7.10d where there are three distinct rPPG waveforms corresponding to the three regions of the spatial pattern in figure 7.8b.

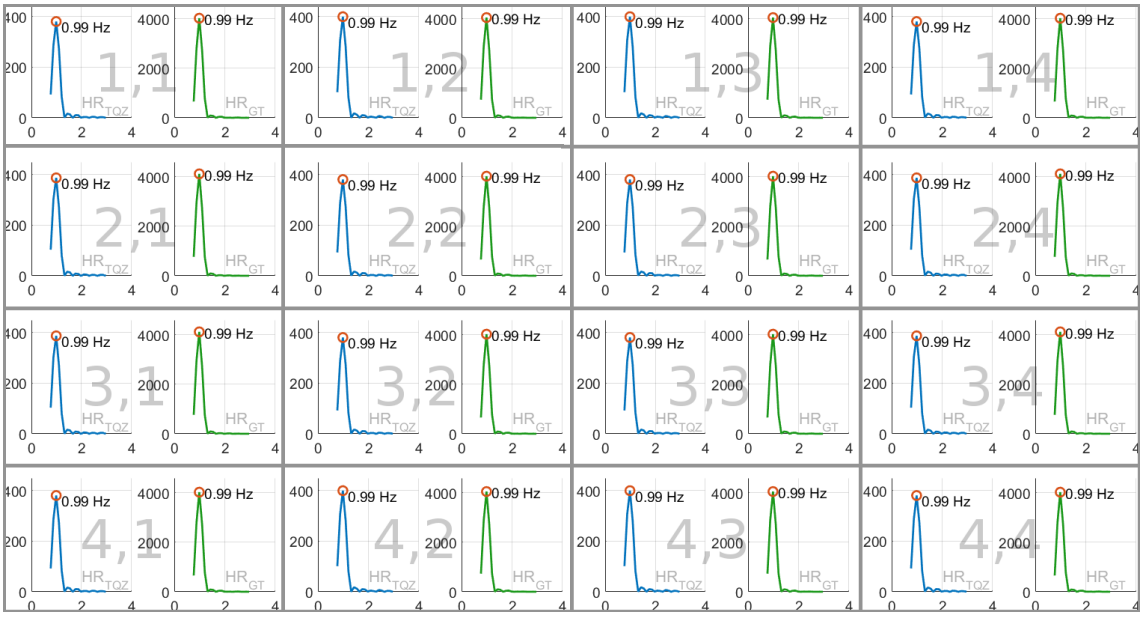


Figure 7.11: FFT analysis for the temporal signal corresponding to each pixel of the output video obtained by the Tensor QZ algorithm. Each block is annotated with its position r, c in the image where $r, c \in [1, 4]$.

- The correct spatial distribution pattern should also be visible in the resultant video. First, looking at figures 7.10c and 7.10e, the values of the pixels are coherent with the respective temporal signals and their corresponding waveforms in figures 7.10d and 7.10f. This means that parts of the waveforms with higher amplitude appear brighter in the video strip and parts with lower amplitudes appear darker. Second, if watched closely, the spatial PPG pattern can also be deciphered in both the video strips, although the Tensor QZ algorithm in figure 7.10d seems to better extract the spatial PPG pattern.

Thus, the Tensor QZ algorithm is successful in extracting the spatial distribution while simultaneously maximizing the pixel-wise periodic variance of the synthetic video. With this validation in place, the algorithm can be attempted on real video data, perhaps also live video data to perform SrPDE, which shall be undertaken in a future work.

7.4/ SUMMARY

In this chapter, we developed the higher dimensional analogues of the QZ algorithm which is used at the core of the PVM algorithm of chapter 6 to perform GEVD on the pair of lagged covariance and covariance matrices, $P_{\mathbf{x}}$ and $C_{\mathbf{x}}$. We formulated two higher dimensional variants, the Block QZ and the Tensor QZ algorithm, concluding with the advantages of the tensor version over the block version. Finally, we demonstrated the efficiency of the Tensor QZ algorithm with a synthetic example emulating the SrPDE problem, to perform GEVD simultaneously across multiple elements, in order to extract the underlying periodic signals over a spatial region, while maintaining spatial coherence. The synthetic experiment validates the capability of the Tensor QZ algorithm in extracting the underlying quasi-periodic source simultaneously maintaining the spatial coherence. This

validation paves the way for application of this algorithm to the SrPDE which is one of the essential future works to be undertaken. Another related future work is going to be the implementation of the Tensor PVM algorithm.

III

CONCLUSION

8

Conclusions and future work

The problem of physiological signals estimation is ubiquitous, of which the cardiac signal is one of the most important biomedical signals. Today, remote monitoring of heart rate and related physiological signs is on the rise and owing to the advanced computing capabilities of devices, long-term monitoring and measurement of these signals is becoming more and more feasible, which calls for conception of novel ideas and related methods for measurement of said physiological signals.

Consequently, much research has been done in the domain of remote photoplethysmography, to which our research contributes some significant advancements in the form of new algorithms. Our algorithms exploit the periodicity information, which is quite pervasive in almost all biomedical signals, to improve existing blind source separation methods in order to obtain better signal estimates. It is this omnipresence of periodicity in biomedical signals enables the methods developed in this thesis to be able to be extensively practicable not only in the applications presented, but with other biomedical signal modalities as well. The novel methods that have been developed in this thesis exploit this periodicity information by quantifying it in the form of autocorrelation in different manners.

All the methods developed in this thesis were validated against two publicly available databases, one of which actually is our contribution and is specifically tailored for analysis of remote photoplethysmography methods. The creation of this inhouse database, UBFC-RPPG, marks a significant contribution of this thesis and is promising in use to the research community. The database comprises of two datasets, SIMPLE and REALISTIC, with 9 and 46 videos each comprising of around 21k and 94k frames respectively. Since it is specifically created keeping rPPG measurement analysis in mind, the database has proven extremely useful to us in the analysis of our novel methods and their comparison with other state of the art methods. The database being made publicly available has already enabled other researchers to employ it in their works, proving its utility. We are positive that this database shall be consequential in urhter advancing research in the domain of rPPG signal measurement.

Regarding novel algorithms, two semi blind source extraction methods, Multi-objective optimization using Autocorrelation and Independent Component Analysis (MAICA) and Constrained Independent Component Analysis (cICA) have been developed. The MAICA algorithm combines the mean squared autocorrelation as an objective function with the objective function of traditional ICA, negentropy, in a multi-objective optimization scheme. Its validation against the UBFC-RPPG and the public MMSE-HR database [84] has proved its accuracy in rPPG signal estimation. This algorithm also served as an initial corroboration for the employment of periodicity as a priori information in order to improve

BSS methods, thereby enabling them to *see* and extract the underlying cardiac signal, in the case of rPPG signal estimation.

Another variant of the MAICA method, the parameter-free adaptive step-size MAICA, was also developed, which focuses on improving the optimization part of the algorithm by choosing adaptive-step sizes for the two functions, negentropy and autocorrelation in the gradient ascent optimization, which traditionally uses a fixed step size. The introduction of this adaptive step-size provided improved results as compared to MAICA with a fixed step sized, MAICAfix, along with the simplification provided by the gradient ascent method as opposed to the Newton method used in MAICA, which requires the calculation of the second derivatives of the objective functions. Consequently, MAE values of up to 3.18 bpm were reported against 3.34 bpm from regular MAICA which are both slightly better than the 3.81 bpm provided by the CHROM method, which in our opinion is one of the most robust and simplest methods used for rPPG signal estimation.

In contrast to the incorporation of a priori information as an objective function, in the cICA algorithm, the periodicity information was formulated as a constraint using autocorrelation. This formulation was then further coupled with a priori information corresponding to the physical properties of the skin, in the form of the chrominance constraint. The cICA method extracts the underlying cardiac signal successfully, and provides a slightly improved accuracy of 3.14 bpm as compared to MAICA and parameter-free MAICA. The essential improvement, though, is the capability to incorporate multiple constraints in order to reduce the solution search space for the rPPG signal estimation problem. In this sense, the cICA method conforms to the traditional formulation of a semi-blind source extraction method, that incorporates a priori information in the form of constraints. For the implementation of cICA, we used the interior point algorithm, that provides the best of Newton and conjugate gradient methods, in an attempt to perform the optimization efficiently. The algorithm is not without limitations though. As is the case with any constraint based optimization, one such limitation is the requirement of using thresholds that indicate whether specific constraints are satisfied or not. An automatic threshold selection scheme as formulated in the parameter-free adaptive step-size MAICA method can be used to overcome this limitation.

In spite of this tentative solution of automating the threshold estimation, another major limitation of cICA, and even MAICA, remains that they are optimization based methods. As a result, these methods are prone to issues with indeterminacy of the solution, which required us to perform regression analysis on the obtained solutions. To circumvent issues related to optimization, the Periodic Variance Maximization (PVM) method was formulated, which while providing substantial accuracies in rPPG signal measurement of up to 4.47 bpm, is generic enough that it can be adapted to other modalities such as ECG and neural signals. The concept of periodicity as a priori information is also used in this algorithm, albeit in the form of lagged covariance matrix $\mathcal{P}_{\mathbf{x}}$, which can be thought of as the analogue of autocorrelation. The PVM method uses Generalized Eigenvalue Decomposition at its core which not only makes it extensive enough to be used in applications other than rPPG signal estimation, but also eliminates the optimization step. Furthermore, the calculation of the derivatives of the autocorrelation and chrominance constraints is also not required, making PVM significantly faster than MAICA and cICA, and consequently providing the possibility of using it in a real-time scenario. With respect to the accuracy, MAE values of 1 bpm are acceptable in the case of PVM due to its improved computational complexity and adaptability with problems other than rPPG signal estimation.

Finally, substantiating the extendability of GEVD, it was adapted to the higher dimensional case, especially for scenarios requiring analysis of temporal signals coupled with the requirement of keeping the spatial coherence of the underlying data. Specifically, the QZ algorithm used for performing GEVD was adapted to higher dimensions, tailored to the domain of temporal signals. This novel Tensor QZ algorithm is formulated to handle tensor data, to simultaneously perform GEVD on multiple blocks corresponding to single spatial data points. For instance, in the application of Spatial rPPG Distribution Estimation (SrPDE), the Tensor QZ algorithm can be applied to pixel-wise lagged covariance and covariance matrices simultaneously, instead of pixel by pixel, to extract the subspace that maximizes the periodic variance of the pixel-wise temporal signals. The application of the Tensor QZ algorithm on a synthetic example aimed to emulate the SrPDE problem are extremely promising, with respect to high accuracy as well as improved computation times. This method consequently promises extensive use, owing to its generic nature and the comprehensive utility of GEVD.

This thesis can be useful for the implementation of semi-blind source extraction methods, not only in the domain of rPPG measurement, but also in biomedical signals analysis corresponding to other scenarios. The Tensor QZ alorithm is interesting and novel in its conception, especially because such a tool does not exist in current literature. It can be used in scenarios where temporal signals have to be analyzed while keeping the spatial relations between the signals intact. That being said, the method can undoubtedly improved upon and extended. One definite course of action for future work is the application of the Tensor QZ algorithm on real data from the UBFC-RPPG and MMSE-HR databases, and possibly even in real-time scenarios where the visualization of the spatial rPPG distribution of a skin patch using only a generic web camera shall be extremely practical.

Speaking of generic scenarios that expand the gamut of applications that use semi-blind source extraction methods, it would be desirable to verify the utility of the algorithms developed to other modalities of biomedical signals, such as neural and ECG signal analysis. For instance, periodicity is an essential property of EEG signals, and research in this field comprises of solving many interesting problems. One such problem is the estimation of intent information from EEG signals, that indicate the intention of performing an activity or thinking of a simple action, such as moving a mouse pointer on a screen. In such a scenario, the idea of periodicity information can be combined with a machine learning approach to create novel algorithms.

The objectives set out in the beginning of this thesis have been fulfilled. During the course of this thesis, we have developed several semi-blind source extraction methods, some specific to rPPG signal estimation, and some generic enough to warrant use in other biomedical signal analysis problems. The research related to these methods has been validated and published in international journals and conferences, a list of which is presented in the next chapter. It is our hope and belief that the research undertaken in this thesis shall be a valuable addition to the scientific community, not only in the domain of rPPG signal estimation, but also in generic biomedical signal analysis.

AUTHOR'S PUBLICATIONS

INTERNATIONAL JOURNALS

[**a**] MACWAN, R., BENEZETH, Y., AND MANSOURI, A. **Heart rate estimation using Remote Photoplethysmography with Multi-objective Optimization**. *Biomedical Signal Processing and Control* (2018).

[**b**] MACWAN, R., BENEZETH, Y., AND MANSOURI, A. **Remote photoplethysmography with constrained ICA using periodicity and chrominance constraints**. *BioMedical Engineering Online 17*, 1 (2018), 22.

[**c**] BOBBIA, S., MACWAN, R., BENEZETH, Y., MANSOURI, A., AND DUBOIS, J. **Unsupervised skin tissue segmentation for remote photoplethysmography**. *Pattern Recognition Letters*,0 (2017), 1-9

INTERNATIONAL CONFERENCES

[**i**] MACWAN, R., BENEZETH, Y., AND MANSOURI, A. **Parameter-free adaptive step-size multiobjective optimization applied to remote photoplethysmography**. *IEEE Biomedical and Health Informatics Conference* (2018).

[**ii**] MACWAN, R., BENEZETH, Y., MANSOURI, A., NAKAMURA, K., AND GOMEZ, R. **Remote Photoplethysmography measurement using Constrained ICA**. *IEEE E-Health and Bioengineering Conference* (2017).

[**iii**] MACWAN, R., BOBBIA S., BENEZETH, Y., DUBOIS J., AND MANSOURI, A. **Periodic Variance Maximization using Generalized Eigenvalue Decomposition applied to Remote Photoplethysmography estimation**. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2018), pp. 1445–1453.

[**iv**] BENEZETH, Y., LI, P., MACWAN, R., NAKAMURA, K., GOMEZ, R., AND YANG, F. **Remote heart rate variability for emotional state monitoring**. *IEEE EMBS International Conference on Biomedical and Health Informatics* (2018), pp. 153-156

PATENTS

1. BENEZETH Y., BOBBIA S., DUBOIS J., MACWAN R., MANSOURI A. **Surveillance de la qualité du sommeil d'un bébé par combinaison de données multimodales incluant la photopléthysmographie**. *publié le 14/12/17 sous le numéro WO 2017/212174 A1.*

BIBLIOGRAPHY

[1] HERTZMAN, A. B. **Photoelectric Plethysmography of the Fingers and Toes in Man**. *Experimental Biology and Medicine 37*, 3 (1937), 529–534.

[2] CHERRY, E. C. **Some Experiments on the Recognition of Speech, with One and with Two Ears***. *The Journal of the Acoustical Society of America 25*, 5 (1953), 975–979.

[3] HOUSEHOLDER, A. S. **Unitary Triangularization of a Nonsymmetric Matrix**. *Journal of the ACM 5*, 4 (1958), 339–342.

[4] BRENNAN, D. G. **Linear Diversity Combining Techniques**. *Proceedings of the IRE 47*, 6 (1959), 1075–1102.

[5] FRANCIS, J. G. F. **The QR Transformation, A unitary analogue to LR Transformation - Part I**. *The Computer Journal 4*, 3 (1961), 265–271.

[6] WILKINSON, J. H. **The algebraic eigenvalue problem**. Oxford University Press, Oxford, 1965.

[7] PETERS, G., AND WILKINSON, J. H. **Eigenvectors of real and complex matrices by LR and QR triangularizations**. *Numerische Mathematik 16*, 3 (1970), 181–204.

[8] J.E, D., TRAUB, J. F., AND WEBER, R. P. **On Matrix polynomials, lambda-matrix and Block Eigenvalue problems**. Tech. rep., Carnegie-Mellon University, 1971.

[9] EISENFELD, J. **Block diagonalization and eigenvalues**. *Linear Algebra and Its Applications 15*, 3 (1976), 205–215.

[10] KUROIWA, Y., AND CELESIA, G. **Clinical significance of periodic eeg patterns**. *Archives of Neurology 37*, 1 (1980), 15–20.

[11] BERTSEKAS, D. **Constrained Optimization and Lagrange Multiplier Methods**. Academic Press, 1982.

[12] VICTÓRIA, J. **Block Eigenvalues of block compound matrices**. *Linear Algebra and Its Applications 47*, C (1982), 23–34.

[13] SHIEH, L. S., TSAY, Y. T., AND WANG, C. T. **Matrix Sector Functions and Their Applications to System Theory**. *IEE Proc. 131*, 5 (1984), 171–181.

[14] MAGNUS, J. **On Differentiating Eigenvalues and Eigenvectors**. *Econometric Theory 1* (1985), 179–191.

[15] SHIEH, L. S., TSAI, J. H., AND YATES, R. E. **The generalized matrix sector function and the separation of matrix eigenvalues**. *IMA Journal of Mathematical Control and Information 2*, 3 (1985), 251–258.

[16] GLOVER, F. **Paths for Integer Programming**. *Computers and Operations Research 13*, 5 (1986), 533–549.

[17] STRANG, G. **Linear Algebra and its Applications**, 3rd ed. Cengage Learning, New York, 1988.

[18] JUTTEN, C., AND HERAULT, J. **Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture**. *Signal Processing 24*, 1 (1991), 1–10.

[19] TOMINAGA, S. **Dichromatic reflection models for a variety of materials**. *Color Research & Application 19*, 4 (1994), 277–285.

[20] DOUGLAS, S. **Generalized gradient adaptive step sizes for stochastic gradient adaptive filters**. *1995 International Conference on Acoustics, Speech, and Signal Processing 2*, 3 (1995), 1396–1399.

[21] HAYKIN, S., HILL, H., AND HAYKIN, S. **Adaptive Filter Theory**, vol. 90. Prentice Hall, 1995.

[22] EDWARDS, G., TAYLOR, C., AND COOTES, T. **Interpreting face images using active appearance models**. *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition* (1998), 300–305.

[23] CARDOSO, J. F. **High-order contrasts for independent component analysis**. *Neural Computation 11*, 1 (1999), 157–192.

[24] HYVARINEN, A. **Fast and robust fixed-point algorithm for independent component analysis**. *IEEE Transactions on Neural Networks and Learning Systems 10*, 3 (1999), 626–634.

[25] BYRD, R. H., GILBERT, J. C., AND NOCEDAL, J. **A trust region method based on interior point techniques for nonlinear programming**. *Mathematical Programming, Series B 89*, 1 (2000), 149–185.

[26] HYVÄRINEN, A., AND OJA, E. **Independent component analysis: Algorithms and applications**. *Neural Networks 13*, 4-5 (2000), 411–430.

[27] PEREIRA, E., AND VITÓRIA, J. **Deflation for block eigenvalues of block partitioned matrices with an application to matrix polynomials of commuting matrices**. *Computers and Mathematics with Applications 42*, 8-9 (2001), 1177–1188.

[28] HASAN, M., AND HASAN, J. **Block eigenvalue decomposition using nth roots of the identity matrix**. *Proceedings of the IEEE Conference on Decision and Control 2*, December (2002), 2119–2124.

[29] TARVAINEN, M. P., RANTA-AHO, P. O., AND KARJALAINEN, P. A. **An advanced detrending method with application to HRV analysis**. *IEEE Transactions on Biomedical Engineering 49*, 2 (2002), 172—-175.

[30] LEARNED-MILLER, E. G., AND FISHER III, J. W. **ICA Using Spacings Estimates of Entropy**. *The Journal of Machine Learning Research 4* (2003), 1271–1295.

[31] DEN UYL, M., VAN KUILENBURG, H., AND LEBERT, E. **FaceReader: an online facial expression recognition system**. *Measuring Behavior 2005 5th International Conference on Methods and Techniques in Behavioral Research 2005*, September (2005), 589–590.

[32] LU, W., AND RAJAPAKSE, J. C. **Approach and applications of constrained ICA**. *IEEE Transactions on Neural Networks 16*, 1 (2005), 203–212.

[33] WIERINGA, F. P., MASTIK, F., AND VAN DER STEEN, A. F. W. **Contactless multiple wavelength photoplethysmographic imaging: A first step toward "spO 2 camera" technology**. *Annals of Biomedical Engineering 33*, 8 (2005), 1034–1041.

[34] LU, W., AND RAJAPAKSE, J. C. **ICA with Reference**. *Neurocomputing 69*, 16-18 (2006), 2244–2257.

[35] WU, H., RUBINSTEIN, M., SHIH, E., GUTTAG, J., DURAND, F., AND FREEMAN, W. **Eulerian Video Magnification for Revealing Subtle Changes in the World**. *ACM Transactions on Graphics 31*, 4 (2006), 1–8.

[36] CONAIRE, C., O'CONNOR, N., AND SMEATON, A. **Detector adaptation by maximising agreement between independent data sources**. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2007).

[37] DELORME, A., PALMER, J., OOSTENVELD, R., ONTON, J., AND MAKEIG, S. **Comparing Results of Algorithms Implementing Blind Source Separation of EEG Data**. *Neural Networks* (2007).

[38] VAN DER AA, N., TER MORSCHE, H. G., AND MATTHEIJ, R. R. M. **Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem**. *Electronic Journal of Linear Algebra 16* (2007), 300–314.

[39] HÜLSBUSCH, M. **An image-based, functional method for the opto-electronic detection of skin perfusion**. *Anesthesia & Analgesia 105*, 6S Suppl (2008), S31.

[40] LU, H., PLATANIOTIS, K., AND VENETSANOPOULOS, A. **MPCA : Multilinear Principal Component Analysis of Tensor Objects**. In *IEEE Transactions on Neural Networks* (2008), vol. 19, pp. 18–39.

[41] SAMENI, R., JUTTEN, C., SHAMSOLLAHI, M., AND ELECTROCARDIOGRAM, M. **Multichannel Electrocardiogram Decomposition Using Periodic Component Analysis**. *IEEE Transactions on Biomedical Engineering* (2008), 1935–1940.

[42] VERKRUYSSE, W., SVAASAND, L. O., AND NELSON, J. S. **Remote plethysmographic imaging using ambient light.** *Optics express 16*, 26 (2008), 21434–21445.

[43] ZARZOSO, V., AND COMON, P. **Optimal step-size constant modulus algorithm**. *IEEE Transactions on Communications 56*, 1 (2008), 10–13.

[44] KOLDA, T. G., AND BADER, B. W. **Tensor decompositions and applications**. *SIAM Review 51*, 3 (2009), 455–500.

[45] TSALAILE, T., SAMENI, R., SANEI, S., JUTTEN, C., AND CHAMBERS, J. **Sequential blind source extraction for quasi-periodic signals with time-varying period**. *IEEE Transactions on Biomedical Engineering 56*, 3 (2009), 646–655.

[46] C . B . MOLER AND G . W . STEWART. **An Algorithm for Generalized Matrix Eigenvalue Problems**. *Society for Industrial and Applied Mathematics Journal on Numerical Analysis 10*, 2 (2010), 241–256.

[47] LOU, X.-J., CHEN, H.-J., AND JIE, L. **Block Diagonalization and a Numerical Method of the Eigenproblem for an nxn Normal Matrix**. *2010 First ACIS International Symposium on Cryptography, and Network Security, Data Mining and Knowledge Discovery, E-Commerce and Its Applications, and Embedded Systems* (2010), 31–33.

[48] NAKAJIMA, H., NAKADAI, K., HASEGAWA, Y., AND TSUJINO, H. **Blind source separation with parameter-free adaptive step-size method for robot audition**. *IEEE Transactions on Audio, Speech and Language Processing 18*, 6 (2010), 1476–1485.

[49] POH, M. Z., MCDUFF, D. J., AND PICARD, R. W. **Non-contact, automated cardiac pulse measurements using video imaging and blind source separation.** *Optics express 18*, 10 (2010), 10762–10774.

[50] SAMENI, R. **A Review of Fetal ECG Signal Processing Issues and Promising Directions**. *The Open Pacing, Electrophysiology & Therapy Journal* (2010).

[51] DEB, K. **Multi-objective optimization using evolutionary algorithms: an introduction**. *Multi-objective evolutionary optimisation for product design and manufacturing* (2011), 1–24.

[52] GAUTHAAM, M., AND SATHISH KUMAR, S. **EMG controlled bionic arm**. *Proceedings of National Conference on Innovations in Emerging Technology, NCOIET'11* (2011), 111–114.

[53] LEWANDOWSKA, M., RUMINSKI, J., KOCEJKO, T., AND NOWAK, J. **Measuring pulse rate with a webcam; A non-contact method for evaluating cardiac activity**. *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)* (2011), 405–410.

[54] POH, M. Z., MCDUFF, D. J., AND PICARD, R. W. **Advancements in noncontact, multiparameter physiological measurements using a webcam**. *IEEE Transactions on Biomedical Engineering 58*, 1 (2011), 7–11.

[55] TURSA, J. **MTIMESX: Fast Matrix Multiply with Multi-Dimensional Support, version 1.10.0.0**, 2011.

[56] TILAKARATNA, P. **How pulse oximeter works**, 2012.

[57] AARTS, L. A., JEANNE, V., CLEARY, J. P., LIEBER, C., NELSON, J. S., BAMBANG OETOMO, S., AND VERKRUYSSE, W. **Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit - A pilot study**. *Early Human Development 89*, 12 (2013), 943–948.

[58] DE HAAN, G., AND JEANNE, V. **Robust pulse rate from chrominance-based rPPG**. *IEEE Transactions on Biomedical Engineering 60*, 10 (2013), 2878–2886.

[59] HOLTON, B. D., MANNAPPERUMA, K., LESNIEWSKI, P. J., AND THOMAS, J. C. **Signal recovery in imaging photoplethysmography**. *Physiological Measurement 34*, 11 (2013), 1499–1511.

[60] ALIAN, A., AND SHELLEY, K. **Photoplethysmography**. *Best Practice and Research: Clinical Anaesthesiology 28*, 4 (2014), 395–406.

[61] DE HAAN, G., AND VAN LEEST, A. **Improved motion robustness of remote-PPG by using the blood volume pulse signature.** *Physiological measurement 35*, 9 (2014), 1913–1926.

[62] HSU, Y., LIN, Y. L., AND HSU, W. **Learning-based heart rate detection from remote photoplethysmography features**. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (2014), 4433–4437.

[63] LI, X., CHEN, J., ZHAO, G., AND PIETIK, M. **Remote Heart Rate Measurement From Face Videos Under Realistic Situations**. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 4264–4271.

[64] MCDUFF, D., GONTAREK, S., AND PICARD, R. **Improvements in remote cardiopulmonary measurement using a five band digital camera**. *IEEE Transactions on Biomedical Engineering 61*, 10 (2014), 2593–2601.

[65] TARASSENKO, L., VILLARROEL, M., GUAZZI, A., JORGE, J., CLIFTON, D. A., AND PUGH, C. **Non-contact video-based vital sign monitoring using ambient light and auto-regressive models.** *Physiological measurement 35*, 5 (may 2014), 807–31.

[66] TASLI, H. E., GUDI, A., AND DEN UYL, M. **Remote PPG based vital sign measurement using adaptive facial regions**. *2014 IEEE International Conference on Image Processing, ICIP 2014* (2014), 1410–1414.

[67] FENG, L., PO, L. M., XU, X., LI, Y., CHEUNG, C. H., CHEUNG, K. W., AND YUAN, F. **Dynamic ROI based on K-means for remote photoplethysmography**. *ICASSP 2015-Augus* (2015), 1310–1314.

[68] FLETCHER, R. R., CHAMBERLAIN, D., PAGGI, N., AND DENG, X. **Implementation of smart phone video plethysmography and dependence on lighting parameters**. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2015-Novem* (2015), 3747–3750.

[69] GUNTHER, J., RUBEN, N., AND MOON, T. **Model-based (passive) heart rate estimation using remote video recording of moving human subjects illuminated by ambient light**. *International Conference on Image Processing (ICIP)* (2015), 2870–2874.

[70] KUMAR, M., ASHOK VEERARAGHAVAN, AND ASHUTOSH SABHARWAL. **DistancePPG: Robust non-contact vital signs monitoring using a camera**. *Biomedical Optics Express 6*, 5 (2015), 1565.

[71] MCDUFF, D. J., ESTEPP, J. R., PIASECKI, A. M., AND BLACKFORD, E. B. **A Survey of Remote Optical Photoplethysmographic Imaging Methods**. In *37th Annual International Conference of the IEEE EMBC* (2015), pp. 6398–6404.

[72] OSMAN, A., TURCOT, J., AND KALIOUBY, R. E. **Supervised learning approach to remote heart rate estimation from facial videos**. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG 2015* (2015).

[73] WANG, W., STUIJK, S., AND DE HAAN, G. **Exploiting spatial redundancy of image sensor for motion robust rPPG**. *IEEE Transactions on Biomedical Engineering 62*, 2 (jan 2015), 415–425.

[74] WANG, W., STUIJK, S., AND DE HAAN, G. **Unsupervised subject detection via remote PPG**. *IEEE Transactions on Biomedical Engineering 62*, 11 (2015), 2629–2637.

[75] ALWOSHEEL, A., ALASAAD, A., AND ALQARAAWI, A. **Heart rate variability estimation in photoplethysmography signals using Bayesian learning approach**. *Healthcare Technology Letters 3*, 2 (2016), 136–142.

[76] BOBBIA, S., BENEZETH, Y., AND DUBOIS, J. **Remote Photoplethysmography Based on Implicit Living Skin Tissue Segmentation**. *PRL* (2016).

[77] BUTLER, M. J., CROWE, J. A., HAYES-GILL, B. R., AND RODMELL, P. I. **Motion limitations of non-contact photoplethysmography due to the optical and topological properties of skin.** *Physiological measurement 37*, 5 (2016), N27–37.

[78] KIM, J., LEE, J. K., AND LEE, K. M. **Deeply-Recursive Convolutional Network for Image Super-Resolution**. *IEEE Conference on Computer Vision and Pattern Recognition* (2016), 1637–1645.

[79] SUN, Y., AND THAKOR, N. **Photoplethysmography Revisited: From Contact to Noncontact, from Point to Imaging**. *IEEE Transactions on Biomedical Engineering 63*, 3 (2016), 463–477.

[80] TSOURI, G. R., KYAL, S., DIANAT, S., AND MESTHA, L. K. **Constrained independent component analysis approach to nonobtrusive pulse rate measurements**. *Journal of biomedical optics 21*, 7 (jul 2016), 071105.

[81] TULYAKOV, S., ALAMEDA-PINEDA, X., RICCI, E., YIN, L., COHN, J. F., AND SEBE, N. **Self-Adaptive Matrix Completion for Heart Rate Estimation from Face Videos under Realistic Conditions**. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 2396–2404.

[82] VAN GASTEL, M., STUIJK, S., AND DE HAAN, G. **New principle for measuring arterial blood oxygenation, enabling motion-robust remote monitoring**. *Nature 6*, December (2016), 1–16.

[83] WANG, W., DEN BRINKER, A., STUIJK, S., AND DE HAAN, G. **Algorithmic Principles of Remote-PPG**. *IEEE Transactions on Biomedical Engineering 64*, 7 (2016), 1479 – 1491.

[84] ZHANG, Z., GIRARD, J. M., WU, Y., ZHANG, X., LIU, P., CIFTCI, U., CANAVAN, S., REALE, M., HOROWITZ, A., YANG, H., COHN, J. F., JI, Q., AND YIN, L. **Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis**. *IEEE Conference on CVPR* (2016), 3438–3446.

[85] LATIFIS, I., PARASHAR, K., DIMITROULAKOS, G., CAPPELLE, H., LEZOS, C., MASSELOS, K., AND CATTHOOR, F. **A MATLAB Vectorizing Compiler Targeting Application-Specific Instruction Set Processors**. *ACM Transactions on Design Automation of Electronic Systems 22*, 2 (2017), 1–28.

[86] WANG, W., DEN BRINKER, A. C., STUIJK, S., AND DE HAAN, G. **Amplitude-selective filtering for remote-PPG**. *Biomedical Optics Express 8*, 3 (2017), 1965–1980.

[87] WANG, W., DEN BRINKER, A. C., STUIJK, S., AND DE HAAN, G. **Robust heart rate from fitness videos**. *Physiological Measurement 38*, 6 (2017), 1023–1044.

[88] BOBBIA, S., LUGUERN, D., BENEZETH, Y., NAKAMURA, K., GOMEZ, R., AND DUBOIS, J. **Real-Time Temporal Superpixels for Unsupervised Remote Photoplethysmography**. In *IEEE Conference on Computer Vision and Pattern Recognition* (2018), vol. v, pp. 1454–1461.

[89] DEMIREZEN, H., AND ERDEM, C. E. **Remote photoplethysography using Nonlinear Mode Decomposition**. In *IEEE Conference on Acoustics, Speech and Signal Processing* (2018), pp. 1060–1064.

[90] GASTEL, V., HAAN, D., GASTEL, M. V., AND HAAN, G. D. **Fully-automatic camera-based pulse-oximetry during sleep**. In *IEEE Conference on CVPR* (2018).

[91] LIU, XUENAN; YANG, XUEZHI;JIN, JING; LI, J., LIU, X., YANG, X., JIN, J., AND LI, J. **Self-adaptive signals separation for non-contact heart rate estimation from facial video in realistic environments**. *Physiological measurement* (2018), 0–18.

[92] MCDUFF, D. **Deep Super Resolution for Recovering Physiological Information from Videos**. In *IEEE Conference on CVPR* (2018), pp. 4321–4328.

[93] NOWARA, E. M., MARKS, T. K., MANSOUR, H., AND VEERARAGHAVAN, A. **SparsePPG: Towards Driver Monitoring Using Camera-Based Vital Signs Estimation in Near-Infrared**. In *IEEE Conference on CVPR* (2018).

[94] PARK, J., SABHARWAL, A., AND VEERARAGHAVAN, A. **Direct-Global Separation for Improved Imaging Photoplethysmography**. In *IEEE Conference on CVPR* (2018), pp. 1456–1465.

[95] MATLAB AND STATISTICS TOOLBOX RELEASE 2016B, THE MATHWORKS, INC., NATICK, MASSACHUSETTS, U. S. **Vectorization**.

List of Figures

1.1 PPG setup . . . 4
1.2 PPG measurement with a pulse oximeter . . . 5
1.3 Absorption spectra of Hemoglobin . . . 6
1.4 Remote Photoplethysmography measurement framework . . . 7

3.1 Central Limit Theorem . . . 28
3.2 Non-guassianity = Independence . . . 28
3.3 rPPG measurement framework . . . 31
3.4 Autocorrelation of a sinusoid vs a random signal . . . 32

4.1 System framework for MAICA and its parameter free variant . . . 42
4.2 Typical rPPG signal . . . 43
4.3 Weights analysis for MAICA . . . 44
4.4 Correlation Analysis for MAICA . . . 45
4.5 ICA vs MAICA for certain videos . . . 47

5.1 Mean-squared Autocorrelation . . . 53
5.2 Combination of periodicity and chrominance constriants . . . 54
5.3 CICA methodology . . . 57
5.4 Correlation analysis of ICA vs cICA . . . 59
5.5 Regression analysis of cICA . . . 60

6.1 Typical distribution of the periodicity metric, $\mathcal{P}(\tau)$ . . . 66
6.2 Periodicity Variance Maximization methodology . . . 67
6.3 Correlation Analysis for PVM . . . 68
6.4 Regression analysis of PVM . . . 70

7.1 SNR map of stationary face videos . . . 74
7.2 Tensorization . . . 75
7.3 Interlacing of block matrices . . . 78
7.4 Householder reflection . . . 85

7.5 The Block Householder transformation 88
7.6 Matrix product using Hadamard product 89
7.7 The Vectorized tensors 91
7.8 Synthetic example for Tensor QZ, the emulated PPG signal 94
7.9 Synthetic example for Tensor QZ, source video 95
7.10 Synthetic example for Tensor QZ, output video 97
7.11 FFT analysis for Tensor QZ 98

List of Tables

4.1 Performance compraisons for MAICA 46
4.2 Performance compraisons for parameter-free MAICA 48

5.1 Performance metrics 58
5.2 Effect of scale and window length 59

6.1 Performance metrics for PVM 69

7.1 Overview of the QZ Algorithm, I 81
7.2 Overview of the QZ Algorithm, II 82

List of Algorithms

1 Optimum Basis Estimation 65
2 Householder transformation 86
3 Subdiagonal elements position estimation 92

List of Abbreviations

- **2SR:** Spatial Subspace Rotation
- **AAM:** Active Appearance Models
- **AMPD:** Automatic Multi-scale based Pulse Detection
- **APBV:** Adaptive PBV
- **AR:** AutoRegressive
- **BR:** Breathing Rate
- **BSS:** Blind Source Separation
- **BVP:** Blood Volume Pulse
- **CHROM:** Chrominance
- **cICA:** Constrained Independent Cimponent Analysis
- **DRCN:** Deeply Recursive Convolution Network
- **ECG:** Electrocardiography
- **EEG:** Electroencephalography
- **EMG:** Electromyography
- **FFT:** Fast Fourier Transform
- **fMRI:** Functional Magnetic Resonance Imaging
- **G-R:** Green Minus Red
- **GEVD:** Generalized Eigenvalue Decomposition
- **HOSVD:** Higher Order Singular Value Decomposition
- **HRV:** Heart-Rate Variability
- **IBI:** Inter-Beat Interval
- **ICA:** Independent Component Analysis
- **IPPG:** Imaging PPG
- **JADE:** Joint Approximate Diagonalization of Eigenmatrices
- **MAE:** Mean Absolute Error
- **MAICA:** Multi-objective optimization using Independent Component Analysis

- **MMSE:** MultiModal Spontaneous Emotion
- **MPCA:** Muli-linear Principal Component Analysis
- **MSF:** Matrix Sector Functions
- **NIR:** Near Infra Red
- **NLMS:** Normalized Least Mean Square
- **NMD:** Non-linear Mode Decomposition
- **PBV:** Pulse from Blood Volume
- **PCA:** Principal Component Analysis
- **POS:** Plane-Orthogonal-to-Skin
- **PPG:** Photoplethysmography
- **PPGI:** PPG Imaging
- **PSD:** Power Spectral Density
- **PVM:** Periodic Variance Maximization
- **RADICAL:** Robust Accurate Direct Independent Component analysis ALgorithm
- **RGB:** Red Green Blue
- **ROI:** Region Of Interest
- **rPPG:** Remote Photoplethysmography
- **SASS:** Self-Adaptive Signal Separation
- **Semi-BSE:** Semi Blind Source Extraction
- **SNR:** Signal to Noise Ratio
- **SpO_2:** Blood Oxygen Saturation
- **SrPDE:** Spatial rPPG Distribution Estimation
- **TFR:** Time-Frequency Representation

Document generated with LaTeX and:
the LaTeX style for PhD Thesis created by S. Galland — http://www.multiagent.fr/ThesisStyle
the tex-upmethodology package suite — http://www.arakhne.org/tex-upmethodology/