FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Using Ordinary Color Video Cameras for Contactless Vital Sign Monitoring in Challenging Conditions

Diogo Terleira Malafaya Baptista

WORKING VERSION

Mestrado Integrado em Bioengenharia

Supervisor: Hélder Filipe Pinto de Oliveira, PhD

Co-supervisor: Sara Campos Monteiro Sabino Domingues, MD

June 28, 2020

Using Ordinary Color Video Cameras for Contactless Vital Sign Monitoring in Challenging Conditions

Diogo Terleira Malafaya Baptista

Mestrado Integrado em Bioengenharia

June 28, 2020

Resumo

A Monitorização de recém-nascidos é uma tarefa árdua que é executada continuamente em cada Unidade de Cuidados Intensivos Neonatais. Devido ao delicado estado de equilíbrio na saúde dos recém-nascidos, a monotorização de sinais vitais é essencial, uma vez que permite uma deteção atempada de problemas médicos, contribuindo ativamente para o bem estar e saúde do indivíduo pré-termo. De forma convencional, os recém-nascidos são monitorizados através de sondas fixas na pele. No entanto, estes instrumentos podem danificar a epiderme e aumentar o risco de infeção, bem como causar desconforto ou dor à criança.

Evoluções recentes nas técnicas de Visão por Computador possibilitaram o desenvolvimento de algorítmos de monotorização baseados em imagem, que representam uma alternativa sem contacto para a extração de sinais fisiológicos tais como batimento cardiaco e ritmo respiratório. Vantagens da monotorização sem contacto incluem redução nos danos causados à epiderme, limitação do número de sondas e monitores usados (deixando maior área de superfície corporal para outros cuidados).

Neste estudo, um método para extração continua e sem contacto do batimaneto cardiaco e respiratório foi desenvolvido, fazendo uso de câmeras de video comuns. A técnica desenvolvida baseia-se na deteção de variações subtis na cor da luz refletida pela pele. Ao bater, o coração induz o fluxo de sangue que se propaga até aos capilares mais superficiais. Por esse motivo, o volume de sangue dentro desses mesmo capilares varia intermitentemente, resultando no fenómeno de Blood Volume Pulse (BVP). Uma vez que a absorvância do sangue difere da dos tecidos envolventes, diferentes volumes sanguineos nos capilares induzirão diferenças na cor da luz refletida pela pele, que pode ser registada possibilitando a extração de um sinal temporal. Este sinal é equivalente ao sinal PPG que pode ser extraido por aparelhos como o oximetro de pulso, no entanto a sua extração é feita sem contacto e por isso é comummente referido como remote PGG (ou rPPG). O sinal rPPG possui informação valiosa que pode ser usada para a extração dos ritmos cardiaco e respiratório. Para além do desenvolvimento do método para extração dos sinais vitais, foi também empregue um método para amplificação de variação de cores em video. Este método permite a enfatização do BVP, fenómeno que permite a extração dos ditos sinais vitais. Para além disso, a mesma técnica foi utilizada para magnificar os movimentos respiratórios.

O método desenvolvido provou ser bem sucedido não só nas condições mais simples, mas também em algumas das condições desafiantes impostas pelo dataset usado. Estas condições, as mais prováveis de serem encontradas aquando dos testes numa NICU, são a presença de batimentos cardiacos elevados, uma vez que os recém-nascidos apresentam um batimento cardiaco muito superior ao dos adultos em repouso, e condições de luz não homogenéas. Para estas condições, o método apresentou um RMSE de 1.05 e 1.55 bpm, respetivamente, o que se traduz em erros relativos de 1.6% and 2.3%. No que diz respeito ao ritmo respiratório, os resultados para os mesmos desafios foram 3.56 e 3.37 bpm.

Abstract

Monitoring of newborns is a challenging task, which is carried daily at every Neonatal Intensive Care Unit (NICU). Due to the delicate state of equilibrium in neonates' health, vital sign monitoring is important, as it allows for early detection of medical issues and therefore actively contributes for the infant's well-being and health. Conventionally, newborns are monitored via probes affixed to their skin. However, such instruments may cause damage to the epidermis and increase the risk of infection as well as promote great discomfort or even pain to the infant.

Recent discoveries and developments in Computer Vision techniques made it possible to develop image-based monitoring algorithms, which represent a non-contact alternative to record physiological signals such as heart and respiratory rates. Advantages of contactless monitoring methods include reduction in skin breakdown, minimizing the number of probes and monitors used, which leaves more body surface-area for other care.

In this study, a framework for contactless and continuous Heart and Respiratory Rates extraction using ordinary color video cameras was developed. The technique employed is based on the detection of subtle variations in the color of the light reflected by the skin. The pulsating action of the heart induces blood flow, which propagates to the most superficial capillary vessels. For that reason, as the blood travels back and forth, its volume inside the said capillaries varies intermittently, a phenomenon referred to as Blood Volume Pulse (BVP). As the absorbance of the blood differs from that of the surrounding tissues, different volumes of blood at the most superficial vessels, result in differences in the hue of the light reflected by the skin, which can be extracted over time to form an extremely informative time signal. The time signal inherent to these hue variations is equivalent to the PPG signal, which can be extracted with medical devices such as the pulse oximeter. Its extraction does not require contact with the patient and, hence the signal is often called remote PPG signal (or rPPG). The rPPG signal provides valuable information, which can be used for the extraction of the Heart and Respiratory Rates, among other medically relevant information. In addition to the contactless vital sign extraction method, a technique for subtle change magnification in video was employed. This method allowed the visualization of BVP, phenomenon which makes the extraction of the vital signs possible. Furthermore, the same technique was used to magnify the respiratory movements.

The framework developed with the a database specific for adult subjects and later validated in neonatal subjects proved to be successful not only in simple conditions but also in some of the challenges covered in the dataset. These challenges, which are the most probable to be faced when dealing with the NICU environment, are the presence of high and fluctuating heart rates, due to the increased heart rate of newborns when in comparison with that of adults and uneven lighting conditions, since the lighting distribution inside a NICU is not necessarily homogeneous and may cast shadows on the infants' faces. For these challenges, concerning the heart rate extraction, the framework presented a RMSE of 1.05 and 1.55 bpm, respectively, which translates to relative errors of 1.6% and 2.3%. With respect to the respiratory rate, the results for the same challenges were of 3.56 AND 3.37 bpm.

Agradecimentos

Todo o trabalho realizado durante este ano letivo, do qual resultou esta dissertação de mestrado, foi fruto de um trabalho conjunto, e por esse motivo gostaria de agradecer a todos os que para ele contribuiram. Em primeiro lugar ao professor Hélder e à Doutora Sara, não só pela excelente orientação e apoio prestado, mas também pelo voto de confiança de que estaria à altura deste desafio.

Em segundo lugar, gostaria de agradecer a todos aqueles que de algum modo contribuiram para o meu desenvolvimento pessoal e profissional e por isso se tornaram responsáveis por moldarem a minha personalidade. Principalmente aos meus pais e ao meu irmão, pela educação que me deram e em particular por desde sempre terem estimulado a minha curiosidade e interesse pelas atividades a que me dedico, bem como a manter sempre um espirito crítico. Aos meus avós pelos mimos e pelas ótimas condições que sempre proporcionam. Quero também agradecer à Bia, que é sem dúvida o meu maior pilar, com quem posso contar para me ajudar em tudo e com quem ainda tenho muito a aprender e viver. Queria também agradecer ao Álvaro e ao Vitória, por puxarem por mim e partilharem comigo a experiência da faculdade e aos "Amigos dos Almoços/Viagens" por serem um exemplo e me terem acompanhado ao longo de toda a vida. Um agradecimento à Rita por ter conseguido resolver um problema técnico que dois "quase-engenheiros" não conseguiram. Fica também um grande obrigado a todos os que dedicam o seu tempo a ensinar e com quem tive a sorte de me aprender ao longo dos últimos 22 anos: educadoras; professores dos ensinos básico, secundário e superior; treinadores; professores de música; professores de inglês e língua gestual e quaisquer outros que a minha memória não permita enumerar.

Por fim, gostaria de deixar um breve agradecimento aos pais que autorizaram a participção dos seus filhos neste estudo e, da mesma forma, a todos aqueles que procuram contribuir para o desenvolvimento da ciência da forma que lhes é possível.

Diogo Malafaya

"Seek discomfort."

Yes Theory

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

AAM	Active Appearance Models
CMIN	Centro Materno-Infantil do Norte
CPAP	Continuous Positive Airway Pressure
CRIES	Neonatal Postoperative Pain Assessment Score
ECG	Eletrochardiogram
EDR	ECG-Derived Respiration
EVM	Eulerian Video Magnification
FACS	Facial Action Coding System
FFT	Fast Fourrier Transform
HR	Heart Rate
HRV	Heart Rate Variability
IBI	Inter-beat Intervals
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
kNN	k-Nearest Neighbours
LBP	Local Binary Pattern
LC	Light Condition
LDA	Linear Discriminant Analysis
NIR	Near Infrared
NFCS	Neonatal Facial Coding System
NICU	Neonatal Intensive Care Unit
NIPS	Neonatal Infant Pain Scale
PCA	Principal Component Analysis
PIPP	Premature Infant Pain Profile
PPG	Photoplethysmography
RMSE	Root Mean Squared Error
ROI	Region of Interest
rPPG	Remote Photoplethysmography
RR	Respiratory Rate
RSA	Respiratory Sinus Arrhythmia
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
WHO	World Health Organization

Chapter 1

Introduction

Preterm birth, defined by delivery before the completion of 37 weeks of gestation [1], represents a huge risk factor for neurological impairment and disability [2]. Its complications are the second leading cause of death among children under 5 years of age, responsible for approximately 1 million deaths in 2015 [3]. In addition, preterm birth not only prejudices the infants and their families but also the health services, once the infant may spend months in hospital increasing related costs [4, 5].

Repercussions induced by preterm birth may vary according to the length of the gestational period. Shorter gestational periods bring higher risks, due to greater immaturity of the organs and body functions, thus being associated with increasing mortality, disability and intensity of neonatal care required. This being said, preterm birth may be segmented into: extremely preterm (<28 weeks), very preterm (28 to 32 weeks), and moderate or late preterm (32 to 37 completed weeks of gestation) [6]; the latter representing approximately 75% of all preterm births [7]. Preterm birth which results from a gestational period bellow 22 to 25 weeks (depending on the country and institution) may be considered beyond the limit of viability and therefore should be careful discussed by parents and health care providers as it may not justify the high mortality rates and inevitable complications [8].

Data from 184 countries revealed that the global average preterm birth rate was 11.1%, which sums up to roughly 15 million babies born prematurely in the year of 2010 [6]. From the countries taken in consideration it can be inferred that, prematurity is undoubtedly considered a global problem. Overall preterm birth rate has been rising steadily, due to several premises of the current society. Such premises are the rise in rates of multiple births, as well as greater used of assisted reproduction techniques and more obstetric intervention [2], which have been proven as causal factors for this type of delivery.

Other causes inherent to preterm birth are distributed along diverse etiologic pathways, such as maternal medical conditions, obstetric complications, major congenital anomalies or isolated spontaneous deliveries. Particularly in late preterms, medically indicated elective cesarean sections were responsible for the majority of all deliveries [9]. When it comes to maternal medical conditions, a panoply of factors have been proved to contain strong causal correlations with an

increasing risk of premature delivery. Such factors may have distinct origins, namely behavioral, psychological and social. It is widely known that some unfavorable lifestyle practices are associated with less propitious pregnancy outcomes, among which high risk of prematurity. These include not only tobacco [10], alcohol and drug use during pregnancy but also unhealthy nutritional habits and improper physical efforts [11]. Psychosocial status may also be associated with increased rates of preterm birth. This is influenced by stress, anxiety, depression, mastery, and self-esteem among others [12].

In spite of the increase in survival rates for preterm babies throughout the years, premature delivery is still a associated with complications both while in the Neonatal Intesnive Care Unit (NICU) and throughout adult life [13]. Compared with infants with regular gestation periods, preterms tend to have higher rates of temperature instability, hemodynamic instability, respiratory distress, apnea, hypoglycemia, seizures, jaundice, kernicterus, feeding difficulties, periventricular leucomalacia, and re-hospitalisations [7].

To fight the health repercussions of preterm birth, the World Health Organization (WHO) has proposed 10 main recommendations for both the mother and the newborn. These include Antenatal cortico steroids, Magnesium sulfate for fetal protection against neurological complications and Antibiotic administration, recommended for women with preterm prelabour rupture of membranes [1].

One other factor, which can improve the outcomes of preterm birth is an effective monitoring of the infant's vital signs in the first weeks after delivery. This action can provide useful insights on the baby's state of health and foresee a wide range of complications. Vital signs, commonly referred to as vitals, are a group of the most important medical signs, which are indispensable for monitoring the patient's progress during hospitalisation [14]. These consist of blood pressure, temperature, pulse rate (also known as heart rate) and respiratory rate, though it has been suggested that they could be complemented with other parameters, such as nutritional status, pulse oximetry [15] and even pain measures [16].

Heart rate and respiratory rate are two of the most informative vital signs. They represent an integral part of standard clinical assessment of children with acute illnesses [2] and are also used in routine checkups form infancy to adulthood. By carefully observing these two parameters, and in particular their comparison with the reference ranges, it is possible to anticipate the occurrence of several complications [17]. For these specific vital signs, reference ranges are not static throughout life. Respiratory rate declines from birth to early adolescence, with the steepest fall apparent in infants under 2 years of age, while median heart rate increases from 127 beats per min at birth to a maximum of 145 beats per min at about 1 month, before decreasing to 113 beats per min by 2 years of age [18]. Figure 1.1 shows the median and centiles of respiratory rate and heart rate for healthy children from birth to 18 years of age, obtained and published by Fleming *et al.* [18].

[1] WHO recommendations on interventions to improve preterm birth outcomes. Published in November 2015

[2] Fever in under 5s: assessment and initial management. Guidelines published by NICE (National Institute for Health and Care Excellence) in November 2019

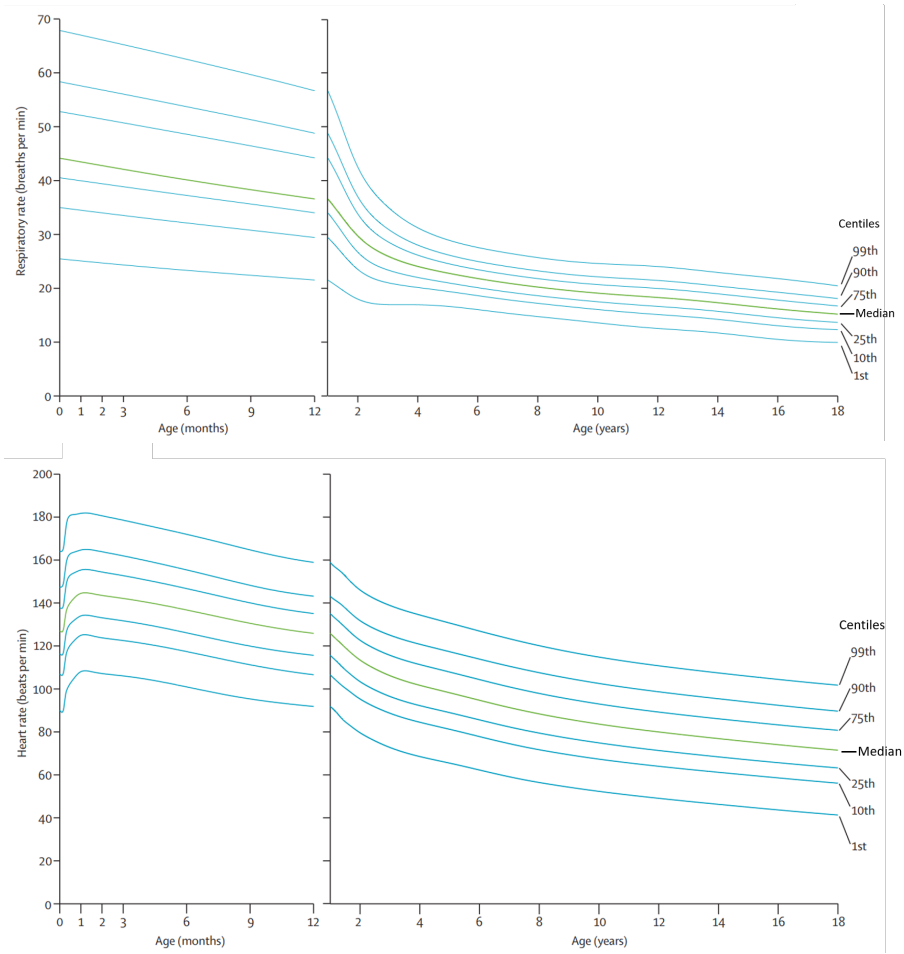


Figure 1.1: Median and Centiles of Respiratory Rate and Heart Rate for healthy children from birth to 18 years of age, adapted from [18]

Body temperature, the third vital sign, is extremely important and its variability is often the first cue for many health problems. Hypothermia, for an instance, causes a decrease in core body temperature, while infections usually result in temperature rising from the widely recognized normal value : 37°C [19], in a phenomenon commonly known as fever. In NICUs, body temperature monitoring is particularly important. While, adults, children and even full-term newborns are able to regulate their body temperature, babies born prematurely have no such ability, once their thermoregulation system only matures in the last trimester of pregnancy [20]. For that reason both incubators and temperature monitoring play an essential role in providing conditions for the neonate's development, by controlling the newborn's body temperature and stimulating the womb's conditions.

Incubators also allow pulse oxymeters for oxygen saturation monitoring. These devices can calculate the percentage of arterial oxyhemoglobin based on the distinct characteristics of light absorption in the red and infrared spectra by oxygenated versus deoxygenated hemoglobin by taking advantage of the variation in light absorption caused by the pulsatility of arterial blood. In spite of its current limitations, pulse oximetry is regarded as an essential element of patient

monitoring in pediatric intensive and perioperative care [21].

1.1 Motivation

According to the World Health Organization (WHO), the number of preterm births per year is continuously rising[3]. Once preterm birth complications are estimated to be responsible for 35% of the world's 3.1 million annual neonatal deaths [6], efforts should be continuously made to address this global problem. While measures that can reduce preterm birth rates are being studied worldwide, it is still relevant to develop methods, which can alleviate the suffering that goes hand in hand with preterm birth. Preterm infants are at higher risks of having temperature instability, respiratory distress, apnea, and seizures [7], and therefore exist in fragile state of health. That being said, accurate monitoring should be preformed in order to guarantee that the babies' health status does not deviate severely from the desired, since this deviation could translate to short and long-term complications or even death.

It is therefore, convenient the development of a more advantageous method of newborn monitoring. Currently, monitoring is performed via probes affixed to the neonate's skin, which may cause damage to the epidermis and increase the risk of infection [22]. By using contactless imaging monitoring, there would be a reduction in the number of probes used, which would decrease skin wearing and overall discomfort and also leave more skin area for other care.

There have been several studies on contactless imaging monitoring. Apart from the scarcity of algorithms developed specifically for newborns, the majority exploits Far-Infrared light (thermal imaging) [23], which implies expensive cameras and may require adaptations to the incubator, as its material is often opaque to radiation in these wavelengths [24].

The development of a method capable of using standard imaging for extraction of neonate's HR and RR requires the use of a database comprising of videos and corresponding ground truth measures. The nonexistence of public databases which focus on neonatal subjects, inflicts the need for creation of a private database. For that reason, we intended to design, collect and put together a database which would be used for the development and testing of the mentioned algorithm. This acquisition would count with a partnership with *Centro Materno Infantil do Norte* (CMIN), where recordings would take place.

Notwithstanding, the global health crisis, which arose in the midst of this academic year and the social constraints imposed in order to fight it, made the acquisition of the said database impossible, as it severely delayed the approval of this study by the Ethic's Committee of the *Centro Hospitalar do Porto*. In an attempt to surpass this obstacle, adaptations had to be made to this master's thesis. At first, the public databases most used in contactless vital extraction problems were searched. However, legal and bureaucratic limitations did not allow for their use. Another dataset of adult subjects was therefore chosen and used [25]. Despite the fact that this database consisted of adult subjects only, its use for the development of the method was always done with

[3] https://www.who.int/news-room/fact-sheets/detail/preterm-birth (Acessed on the 4th of February of 2020)

the final objectives of this thesis in mind, in the sense that, every decision in the pipeline was taken considering what seemed best for the neonate specific application.

In parallel with the development of the methodology to solve the referred problem, efforts were made throughout the duration of the academic semester to move forward with the acquisition at CMIN and database creation, once the extrinsic factors allow. Such efforts included the creation and design of the study, which accompanied by a data acquisition protocol was submitted to the ethics committee of the *Centro Hospitalar do Porto*. A leaflet to be given to the the neonates' parents was also created and can be seen in Appendix B.

1.2 Goals

This thesis project will focus on addressing contactless monitoring of vital signs using visible light, as this topic is of extreme potential and its applicability and validation may result in several benefits. In spite of the dataset used and its limitations, the goal of this thesis is to develop a framework capable of continuous contactless extraction of the HR and RR in newborn subjects. For that reason, all methods put together as well as the analysis of the results were performed having in mind the framework's applicability for the target subjects. Once global conditions allow, the acquisition and creation of a database consisting solely of preterm infants participants, will concede the opportunity to validate the said framework with clinical data. Secondary goals include applying a video magnification method, allowing for a more immediate monitoring of the babies' well-being through color changes associated with the beating of the heart. In case the ultimate goal of replacing the probes used in NICUs is met, benefits would arise, such as decrease in the pain and discomfort felt by the infant while in the NICU. This would translate to an improvement in current monitoring methods, better experience for preterms during hospitalization and ultimately less medical complications.

1.3 Contributions

The development of this master thesis resulted in several outcomes, which will be beneficial for the scientific community in general:

- Algorithm capable of continuously extracting heart rate and respiratory rate in challenging lighting conditions.
- Algorithm for contactless monitoring of heart-beats.

1.4 Outline

Apart from the Introduction, this monograph contains 7 more Chapters.

Chapter 2 describes the current paradigm of NICUs and the devices and methods used, whose comprehension is of great importance for this thesis.

Chapter 3 constitutes a literature review on contactless monitoring and pain assessment methods.

Chapter 4 describes the dataset used.

Chapter 5 discusses the methodology and results for Heart Rate extraction.

Chapter 6 elaborates on the topic of Eulerian Video Magnification and its applications.

Chapter 7 explains the methodology and results for Respiratory Rate extraction.

Chapter 9 summarizes the findings and elaborates on proposed future work.

Chapter 2

Background

Whenever a new birth occurs, it is essential that the newborn is assessed and their health status and individual needs are determined. According to the outcomes of this evaluation, the infant will be assigned to the location most adequate for the type of care needed. While healthy-appearing newborns should be kept near the mother, infants who require specialized medical attention, such as preterm or ill infants, are usually admitted to the special care nursery or NICU [26].

This need for special levels of care in babies whose gestational age is lower than 37 weeks derives from the immaturity associated with underdevelopment by insufficient time in the womb. Such immaturity jeopardizes a wide range of organs and body functions, in particular thermoregulation capabilities. Babies who have not had a full gestational period are reported to have a smaller ratio of body fat to lean mass, once accumulation of this type of tissue will only occur in early post-natal life (what would be the final weeks of the gestation period in case the baby was born in term). Since the absence of body fat is associated to greater heat loss, preterms' ability to regulate body temperature is impaired in their first weeks of life [27]. Reduced time in the womb also induces immaturity of the physiological systems, whose outcome is higher sensitivity to stimuli. This may result in a state of sensory overload in the infant when in a complex and overstimulating environment such as that of the NICU. Consequences of the so called sensory overload may comprise undesirable fluctuations in heart rate, respiratory rate, blood pressure, motor and state systems stability [28].

The respiratory system is affected by prematurity as well. Deficient prenatal lung development often results in respiratory disease, which is the single greatest cause of illness and death in preterm infants [29]. Other insuficciency associated to underdevelopment by insufficient time in the womb is immaturity of the innate immune system. Preterm infants have reduced quantities of monocytes and neutrophils, which makes them highly susceptible to neonatal pathogens and permanent disabilities due to organ damage resulting from either the infection itself or from the inflammatory response created [30, 31]. The combination of these and other flaws in the bodily functions of preterm infants makes these individuals extremely vulnerable to the external conditions imposed by the NICU's environment [32].

2.1 Neonatal Intensive Care Unit

So as to assure a healthy early development for preterm infants, it is important that all their limitations are taken in consideration. Optimal conditions and an adequate environment for neonates' growth is provided by incubators, that being the reason why they spend most of their time inside one. An incubator is a self-contained, crib-like unit, whose purpose is to mimic the conditions inside the womb by contributing with the maintenance of a rigorously controlled environment.

2.1.1 Incubators

One can find several types of devices inside a NICU, which serve a similar purpose as that of the incubator. In spite of being open to the air, the radiant warmer is an apparatus, which actively tries to maintain the infant's body temperature by providing radiant heat below (through the surface where the preterm is laying) or above the baby (through infrared emitters). Advantages of the radiant warmer include open access to infants in need of resuscitation or other procedures, while still providing sufficient exogenous heat to replace natural body heat losses [33]. this reason it is commonly used to stabilize infants following delivery room resuscitation and for transportation of the infant for surgical interventions [34]. Nevertheless, the piece of equipment where the preterm infant spends the most time is the closed box incubator. This type of incubator differs from others by holding a closed hood made of a single or double-layered clear dome (usually made of Acrylic or Plexiglas). Despite allowing high visibility, this hood poses as a physical barrier between the infant and the family, which is not advantageous. Nevertheless, incubators' benefits far outweigh the disadvantages as it has been established that the use of these devices leads to better growth and improved survival rates for preterm infants [35, 36, 37].

As mentioned above, the inefficiency to regulate body temperature is one of the biggest challenges a preterm infant has to face in their early days of life. Therefore, a closed box incubator's main purpose is to assure temperature stability, by minimising heat loss. Such stability is achieved passively (by having the closed hood as an insulating element which prevents heat from escaping the incubator) and actively (by injecting heated air to its interior with the help of a ventilator device). This mechanism protects the preterm against external thermal fluctuations until their thermoregulatory mechanisms become fully efficient, thus preventing states of hyper or hypothermia [38]. Furthermore, closed box incubators are characterized by having a fresh air filtration system. Besides minimizing the risk of infection, this mechanism allows for regulation of the relative humidity as well as the percentage of oxygen in the air inside the incubator. Other perks associated to the incubator include isolation of the neonate from the outside world, thus protecting them from infection or allergens and even from external stimuli by dampening the sound generated by the practitioners, families, devices and other sound emitters present in the NICU [39, 40].

Figure 2.1: Closed box incubator used in the NICU of Centro Materno-Infantil do Norte

2.1.2 Sensors

Like most incubators, the ones used in the NICU of the *Centro Materno-Infantil do Norte* (CMIN) integrate several sensors, which are useful for controlling a vast range of parameters inherent to the incubator's environment [1]. Furthermore, other sensors are used to measure the signs which have to do with the neonates' themselves. Table 2.1 summarizes the variables measured as well as the sensors used to do so.

Table 2.1: Common variables measured and respective sensors both integrated and coupled to the incubator

Variable Measured	**Sensor**	**Integrated**
Body Mass	scale incorporated under the mattress	yes
Air Temperature	thermistor	yes
Oxygen Saturation in Air	eletro-galvanic sensor	yes
Relative humidity	hygrometer	yes
Skin Temperature	thermistor (inserted in probe)	no*
Heart Rate	eletrodes	no
Respiratory Rate	eletrodes	no
Oxygen Saturation in Blood	pulse oxymeter	no
CO2 concentration	transcutaneous CO2 monitor	no

** Can be connected to the incubator and used for air temperature servo-control.*

It is common practice to register the evolution of the infants' weight over time, which is followed on a daily basis in intensive care and once every two days in intermidiate care. For this purpose, a scale is incorporated under the mattress of the incubator. This device is able to preform periodic weighings of the neonate as well as to continuously measure differences in the infant's weight. Although the weighing mode provides the actual value of the infant's mass excluding that

[1] V-2100G Infant Incubator: Operation Manual by Atom Medical Corporation

of the mattress platform, the mattress and the bed sheet, it requires the nurses' intervention, who have to lift and lay down the baby for accurate measure. On the other hand, the weight difference mode does not require a practitioner's help. The device does not behave as a force plate in the sense that it can not discretize distributed weight. Therefore, it is advised that the infant is placed in the center of the mattress for accurate readings.

In order to maintain the oxygen levels inside the incubator, an oxygen flowmeter is usually placed between an oxygen source and the incubator. Upon measuring the percentage of oxygen inside the incubator using an electro-galvanic sensor, the nurse can regulate the flow of oxygen that enters the incubator through the flowmeter. Oxygen saturation inside the hood can range between 21% (not receiving external flow of oxygen apart from the oxygen contained in atmospheric air) and 65%. Furthermore, the relative humidity of the air inside the incubator is strictly regulated. A hygrometer allows for constant measure and in case of low values, a deposit full with distilled water introduces water vapor in the hood increasing relative humidity.

Apart from the sensors integrated in the incubator itself, all devices allow the insertion of probes, which are used to monitor signs intrinsic to the neonates themselves, namely the four vital signs. When it comes to temperature, it can be measured in several distinct ways. The most common practice in NICUs is to use a skin temperature probe [41]. This sensor is usually attached anywhere between the navel and the xiphoid process of the preterm and is composed by a regular thermistor. It can be connected to the incubator and its output used as a parameter for the servocontrol of the incubator's active temperature maintenance by providing feedback control to regulate the heated air environment. [34]. In contrast, control can be also performed manually by periodically measuring the infant's temperature using a regular thermometer or using the thermal sensor integrated in the incubator to measure the temperature of the air and manually setting the incubator's temperature accordingly [41].

Regarding heart and respiratory rates, 3 electrodes affixed to the infant's chest are shared for the extraction of the two vital signs. These sensors are the ones which occupy the most body surface-area, as can be seen in Figure 2.2, and thus, their replacement for a contactless substitute would be the most advantageous. Apart from measuring the two rates, the electrodes also allow for formation of a simple electrocardiogram (ECG).

Finally, two other variables measured have two do with gas exchange. Oxygen saturation in the preterm's blood is measured through a common pulse oxymeter placed on the infant's hand or foot. This sensor relies on a technology called photopletismography, which will be mentioned in the next chapter, as its comprehension is crucial for the understaning of this study. CO_2 saturation levels are also extraced through two different possible methods. The two methods are used distinctly and depend on the type of ventilation needed by the infant. If the infant is under invasive ventilation, meaning a tube which conducts oxygen directly to airways usually inserted through the nostrils, the extraction of CO_2 levels is made easy once the tube itself is able to detect the flow of the said gas in expiration. On the other hand, some preterms do not require ventilation at all or are ventilated using a continuous positive airway pressure (CPAP) machine [42]. This device, which contacts with a infant through a mask, does not have the capability to quantify the

Figure 2.2: Standard monitors and probes used inside a NICU from [23]

CO_2 expelled. Therefore, in theses cases a transcutaneous CO_2 sensor is attached to the neonate's chest [43]. For these sensors to function properly they must be heated [44]. This may induce skin burns and therefore, nurses must constantly switch the position of the probe. Besides, its results are claimed not to be very precise.

All of the variables extracted, both inherent to the incubator and the baby, are followed regularly by nurses and physicians, who take use of the external monitors coupled to the incubator to visualize the evolution of the measured parameters. Figure 2.2 exhibits the external probes as well as monitors and other devices essential to the incubator's well functioning.

2.1.3 Pain assessment

Apart from the uninterrupted monitoring of the mentioned vital signs, other chores compose the daily routine of every NICU and its workforce. Pain assessment is an important task, which is performed periodically and consists of evaluating and quantifying both chronic and acute pain felt by an infant at a given time. In order to do so, nurses take in consideration several different factors such as facial expression, body movement and crying [45]. In addition to such behavioural indicators, neonates exhibit a wide range of physiological responses to painful stimuli as a result of the activation of the sympathetic nervous system [45, 46]. Changes in physiological indicators include increased heart rate, respiration, blood pressure, and oxygen saturation. By using pre-established tools, such as scales which take in consideration several of these responses, it is possible for nurses to quantify and formalize the pain measured. The scales most frequently cited are the Premature Infant Pain Profile (PIPP): Neonatal Postoperative Pain Assessment Score (CRIES), Neonatal Facial Coding System (NFCS), and the Neonatal Infant Pain Scale (NIPS) [47, 48]. These scales aim specifically for neonates. Since these individuals do not have the ability to express themselves verbally, its use relies on the interpretation of nurses, therefore being considered hetero-evaluation

methods. There are, naturally, auto-evaluation scales for use in pediatric patients, such as the *Faces Pain Scale* [49], but these will not be addressed as they deviate from the theme of this study.

When it comes to preterm-specific scales, the reference scales and the ones used at *CMIN* are the *EDIN* and *N-PASS* scales. Once again its use is determined by whether the infant is ventilated or not, the latter being for newborns under ventilation. Most infants, specially the older ones are not ventilated invasively nor through the *CPAP* device, therefore *EDIN* represents best the method used by Nurses for acute pain assessment in the NICU in question.

EDIN (*Échelle Douleur Inconfort Nouveau-Né*) relies on five behavioural indicators of prolonged pain to deduce the overall level of pain felt. These indicators, entirely observational, are: Facial activity; Body Movements; Quality of Sleep; Quality of contact with nurses; Consolability. Each of these variables are scored on a four point scale, 0 indicating well-being and 3 severe prolonged pain. The values for the five variables are then added up. If the sum is equal zero, the baby is considered to be under no pain. In case the sum falls between 1-4 the pain is classified as light, 5-8 as moderate, 9-12 severe and 12-15 extremely severe. For values between 5 to 15 therapeutic intervention is advised [50]. Common practice suggests that this assessment should be performed in 8 hour intervals. In Table 2.2 one can see the EDIN pain scale.

Table 2.2: EDIN pain assessment scale

Indicator	**Description**
Facial activity	**0.** Relaxed facial activity **1.** Transient grimaces with frowning, lip purse and chin quiver or tautness **2.** Frequent grimaces, lasting grimaces **3.** Permanent grimaces resembling crying or blank face
Body movements	**0.** Relaxed body movements **1.** Transient agitation, often quiet **2.** Frequent agitation but can be calmed down **3.** Permanent agitation with contraction of fingers and toes and hypertonia of limbs or infrequent, slow movements and prostration
Quality of sleep	**0.** Falls asleep easily **1.** Falls asleep with difficulty **2.** Frequent, spontaneous arousals, independent of nursing, restless sleep **3.** Sleepless
Quality of contact with nurses	**0.** Smiles, attentive to voice **1.** Transient apprehension during interactions with nurses **2** Difficulty communicating. Cries in response to minor stimulation **3** Refuses to communicate. No interpersonal rapport. Moans without stimulation
Consolability	**0.** Quiet, total relaxation **1.** Calms down quickly in response to stroking or voice, or with sucking **2.** Calms down with difficulty **3.** Disconsolate. Sucks desperately

In contrast, the *N-PASS* differs from the previous one not only because it is used on preterms who are under ventilation but also because it incorporates vital sign variability in addition to

behavioural indicators. This scale can also assess sedation apart from pain.

Chapter 3

Literature Review

In order to develop a framework for contactless vital sign extraction that can be adapted to integrate the NICU and serve the healthcare of preterm infents, it is important to discern the scientific and technical knowledge already unveiled by the scientific community on the topic. In this chapter, a detailed description of the current panorama of contactless HR monitoring (Section 3.2), contactless RR monitoring (Section 3.3) and video change magnification (Section 3.5) will be presented.

3.1 Contactless Monitoring

Most methods currently used in a clinical context for HR and RR monitoring are considered non-invasive. Nevertheless, this does not mean such methods are contact free. In fact, gold standard methods for HR monitoring, such as Electrocardiography (ECG), Phonocardiography (PCG), Echocardiography (Echo) and Photoplethysmography (PPG) among others, require contact of the used instrument with the patients body.

A tendency to evolve to contactless solutions has been rising, partially due to advances in image capturing technologies and Computer Vision techniques. A contactless approach for vital sign monitoring presents several advantages over its contact-dependent competition. In spite of the benefits exhibited throughout this document, contactless solutions come hand in hand with some drawbacks, namely the fact that the effectiveness of these methodologies in real-life scenarios depends on various factors such as variation in illumination, motion artifacts, distance from the camera and quality of the imaging sensors [51]. That being said, there is still a need for clinical validation of such methods in order for a possibility of them being introduced not only in NICUs but hospitals everywhere.

This literature review will lean towards computer vision methods for contactless monitoring of vital signs using exclusively visible and near infra-red light. Tests performed with the incubators at *Centro Materno Infantil do Norte* revealed that the material which constitutes their hood is opaque to thermal radiation, *i.e* infrared radiation of longer wavelengths, thus making impossible the use of thermal cameras, despite their wide range of benefits for vital sign monitoring [52, 53, 54]

3.2 Contactless Heart Rate Monitoring

Recent computer vision guided methods for human pulse estimation, either in infants or adults, broadly fall in one of two categories [51]:

- **Color Guided Techniques:** Use the variation in intensity levels of the different color channels over time to build the feature trajectory, which is fed into a statistical model for HR estimation.
- **Motion Guided Techniques:** Pixel tracking over time to detect subtle periodic motion caused by cardiac pumping action to be used as a feature for pulse estimation.

State of the art methods for both ramifications follow a generic framework divided into three blocks (*Signal Extraction*, *Signal Estimation* and *HR Estimation*), which are schematically represented in Figure 3.2. Differences between the two categories of algorithms reside solely in *Signal Extraction* (first block), in particular in the steps regarding *Region of Interest (ROI) tracking* and *Raw Signal Estimation*, since the underlying principles to do so are divergent. Differences across studies inside each technique are intrinsic to individual steps, in which distinct but equivalent algorithms are used.

3.2.1 Color Guided Techniques

Blood Volume Pulse (BVP) is a concept which refers to the changes of the volume of blood inside the the microvascular bed of tissue, caused by the rhythmic pulsating action of the heart. When the ventricles contract, blood is pumped out of the heart and carried to the peripheral vascular system, filling the capillaries and thus, increasing the volume of blood inside them, momentarily. Due to the difference in light absorption of blood and surrounding tissue, blood volumetric variations lead to periodic change in the amount of light absorbed by the region and consequently in light reflected. These rhythmic fluctuations of the the light intensity are therefore correlated to the HR and can be easily detected in the skin, fingertips and ears [55]. Figure 3.1 illustrates how PPG may be used to exploit the BVP phenomenon.

Photopletismography (PPG) is an optical measurement technique, which can be used to detect BVP. To exploit this phenomenon, a light source (usually operating at red or near infrared wavelengths) illuminates the tissue, while a photodetector captures the light that has passed trough the tissue (transmission mode operation) or reflected by it (reflection mode operation). There are three main reasons for the use of these wavelengths: the first is due to the main constituent of the human tissues being water, which absorbs light very strongly in the ultraviolet and far infrared wavelengths. If this is added to the fact that melanin absorbs the shorter wavelenghts of visible light, only a small window in the absorption spectra is left, which allows measuring blood volume in the red and near infrared spectra; the second motive refers to these wavelengths being the isobetic wavelengths of haemoglobin, meaning the wavelengths for which there are no differences in absorption between oxyhaemoglobin (HbO_2) and reduced haemoglobin (Hb); the third and last

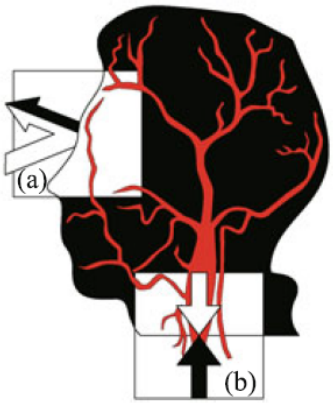


Figure 3.1: Representation of Photoplethysmography to exploit Blood Volume Pulse, extracted from [56]: (a) embodies the variations of reflected light in the skin, a region with a high number of capillaries; (b) portrays the variations in blood caused by the pumping action of the heart.

reason is that the depth to which light penetrates the tissue depends on the light's wavelength, being optimal for this range of wavelenghts [55].

The signal extracted by the photodetector is referred to as the *PPG waveform* and consists of two components: The pulsatile component, which contains information on the HR and and a slowly varying component, related to respiration, vasomotor activity among other factors, which will be explained further on in this chapter [55]. Suitable filtering, amplification and signal processing allow the distinction of both components and subsequent pulse wave analysis [55].

The most common example of the applicability of PPG is the pulse oxymeter, a sensor used to obtain information about the arterial blood oxygensaturation (SpO_2) as well as HR. It functions in transmission mode operation, meaning that the tissue sample (in this case the fingertip) is placed between the light source (usually a light-emitting diode or LED) and the photodetector, which captures the light let through by the finger.

This being said, PPG is the basis of *Color Guided Techniques* for contactless monitoring of the HR. However, since the main purpose of such techniques is to work from a distance, adaptations of PPG had to be performed to allow readings despite the absence of direct contact [57].

Remote Photoplethysmography (rPPG) appeared as a contactless extension of PPG. This technique has gained acceptance among the scientific community, once in 2008 Verkruysse *et al.* proved that reflected ambient light is sufficient to obtain a photoplethysmography signal [58]. Several studies have been published since, which use rPPG as a foundation for HR extraction in human individuals using both commercial and advanced camera equipment.State of the art methods follow a generic framework divided in three blocks, which can be seen in Figure 3.2. Each of these blocks comprises several steps, which may vary across studies.

Fernandez *et al.* in 2015 [59], used the Viola Jones algorithm [60] for detection of the human face as a Region of Interest (ROI) definition initial step. In this study, three variations of the Viola Jones algorithm were trained and integrated to overcome the original's algorithm poor performance when leading with non-frontal faces. Afterwards, and since the output of this method

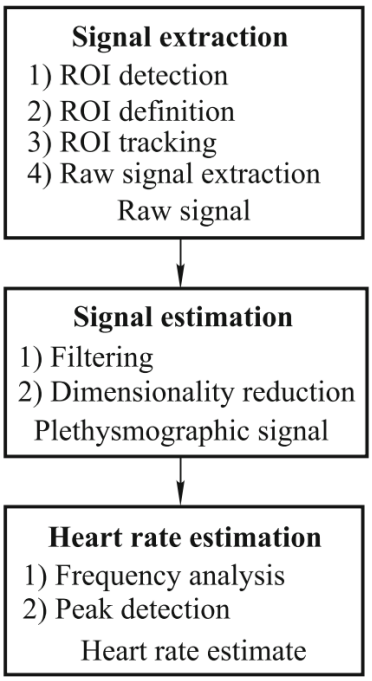


Figure 3.2: Generalized rPPG algorithm framework from [56]

would include non-face pixels from the corners of the rectangles which serve as bounding-box for the face, Fernandez *et al.* added a second step which aims at ROI definition and robust face tracking over-time and exclusion of unwanted regions by focusing on rectangular patches in the subject's forehead. This portion of the algorithm, based on *Deformable Parts Model* [61], detects the corners of the eyes for each frame and aligns the frames in a way that the eyes are always found in the same coordinates. The area above the eyes is then extracted and used as the final ROI for the PPG waveform extraction. Raw signal extraction (the last step in the *Signal extraction* building block) is performed by spatially averaging the intensity values of the ROI pixels for each of the RGB channels, method known as spatial pooling, which results in three signals (one for each channel) which resume the variation of average intensity over time. The *Signal Estimation* block uses these three raw signals as input, which are then smoothed and normalized as part of the filtering step. These signals are then decomposed into three independent source signals using Independent Component Analysis (ICA) and only the range of frequencies of interest is maintained by applying a temporal filter. In the case of adult individuals this frequency corresponds to roughly 1Hz as the normal range for HR in healthy humans individuals is between 60 and 100 bpms. In the final block, or *Heart Rate Estimation*, this study calculated the inter-beat intervals by analysing the distance between peaks and, hence obtaining the number of pulsations per minute, commonly known as heart rate.

Other state of the art methods differ from the one proposed by *Fernandez* in the several steps intrinsic to the general framework. For an instance, despite the Viola Jones algorithm being the most used algorithm in this type of applications [56], other methods for *ROI detection* were also

presented, such as face landmark detection or even combinations between Viola Jones and other methods, for instance Active Appearance Models (AAM) [57] or skin detection algorithms [62]. In addition, although *Fernandez* performs spatial pooling on the three channels, *Verkruysse* showed that the green channel contains the strongest plethysmographic signal, clearly indicating the fundamental HR frequency. Despite ICA being the most common method for dimension reduction, other methods have been presented in state of the art studies, namely *Principal Component Analysis* (PCA) [63]. The aspect in which *Fernandez's* study differs the most form other *state of the art* methods is the technique used to extract the HR itself. While the current study uses peak detection to calculate the time between beats, most literature uses the Fast Fourrier Transform (FFT) to extract the maximum response in frequency domain [64, 65, 66].

3.2.2 Motion Guided Techniques

Instead of relying on color changes, caused by the variations in blood volume at the peripheral blood vessels, *Motion Guided Techniques* extract the HR from periodic motion of the subject's head (generally not observable through naked eyes [67]), which occurs because of the influx of blood to the head [56]. According to *Rouast et al.* [56], studies which adopt these methods over *Color Guided Techniques* are scarce, representing only 9% of the published studies. This is mainly due to this technique's susceptibility to noise due to natural head movements which are inevitable in a real-world problem.

In this type of algorithms, only the *ROI tracking* and *Raw signal extraction* steps differ in the entirety of the pipeline. In order to take advantage of the pulsatile motion caused by blood pressure due to caridac pumping, these techniques must incorporate a method to track points within the ROI. A study conducted by *Irani et al.* in 2014 [68] used the Viola Jones algorithm once more to detect the face of the subject in the frame. This algorithm's output will include all areas of the face, namely the eyes and the mouth area, which are prone to movement due to changes in facial expressions. Since this movement does not reflect the beating of the heart, only the most stable regions of the face, *i.e.* the forehead and nose regions, were isolated. For these purpose portions of the area of Viola and Jones bounding box were manually set and optimized for inclusion of such regions. Afterwards, *Good Feature Tracking* algorithm, was used to select the keypoints, from subsections of the ROI preestablished by the Viola and Jones algorithm. This method extracts each pixels' eigenvalues, rejecting corners for which their value is minimal or too close too stronger corners. Lucas Kanade's tracking algorithm [69] was then used to to extract the x and y components of the trajectories of said keypoints. Only the y components were further carried on for processing since the relevant motion is the one caused by the naturally vertical flow of blood to and from the head. Similarly to *Color Guided Techniques* this method uses filtering to dump unwanted frequencies. In this particular case, a pass band filter is used (8th order Butterworth filter) with cutoff frequency interval of [0.75 5] Hz. Once again, in accordance with the general framework for contactless HR estimation methods, methods were used to reduce the dimensionality and to extract the heart rate from the plethysmographic signal. In this case, PCA was used rather than ICA and Discrete Cosine Transform (DCT) rather than the usual FFT.

3.3 Contactless Respiratory Rate Monitoring

RR monitoring techniques can also be distinguished into color based (rPPG) or motion based. Color based methods can be seen as an extension of the *Color Guided Techniques* for contactless monitoring of the HR. For an instance, in the previously described study, *Fernandez et al.* presents a RR estimation module apart form the HR one already exposed. The value for the RR is derived from the previously extracted HR, in particular the Heart Rate Variability (HRV). HRV represents the variation among the intervals between heartbeats and is calculated by performing Power Spectral Density (PSD) estimation using the *Lomb periodogram*, a method of estimating a frequency spectrum, based on a least squares fit of sinusoids to data samples, from which is possible to detect the RR. Methods based on motion for the detection of RR rely on different principles of those in *Motion Guided Techniques* for extraction of the HR, partially due to the evident chest motion caused by respiration. A method which utilizes motion cues for extraction of the RR is described below.Other methods, such as the one proposed by Iozza *et al.* [70] combine both rPPG derived RR with motion information obtained from video.

3.4 Newborn specific applications

Although HR and RR extraction has been widely explored in adult individuals, newborn specific studies are not abundant, and most of those have been published use infrared thermography imaging [23]. Nevertheless, recent studies using visible light images have been divulged with promising results. Non-contact estimation of HR methods for newborns using standard images was first addressed in 2012 by *Scalise et al* [71]. In this study, seven infants were monitored with a webcam and resorting a special external illumination source.It emitted green light, which was reflected by the infant's skin and measured by the camera. One year later *Aarts et al.* managed to successfully [72] monitor the HR of 19 newborns without dedicated external light sources. In 13 out of the 19 infants the extracted HR matched that of the gold-standard methods fore more than 90% of the time.Since then, efforts have been made towards improving newborn specific applications for HR extraction, which ultimately led up to methods which integrate both HR and RR extraction.

In 2015 *Fernando et al.* applied *Wang et al.*'s work [63] to the NICU. In this study, two regular cameras recording at a resolution of 768x576 and 20 fps were used to capture images of two different regions of the infant. [73] One camera aimed at the neonate's face and its purpose was to estimate HR through color change observation, while the second camera observed the motion of the chest from which RR it would possible to extract the RR, although this study does not specify the proceedings. Regarding HR extraction, instead of applying *spatial pooling* to the ROI as described in most publications, the author considered keypoints inside the ROI which were treated as independent rPPG sensors. To make this possible, an online object tracker was used to track the infant's face over time. Dense optical-flow was then applied to align skin-pixels in consecutive frames to make sure each sensor (pixel) was continuously considering the same skin region. Subsequently, a chrominance based rPPG algorithm was used to extract pulse intervals

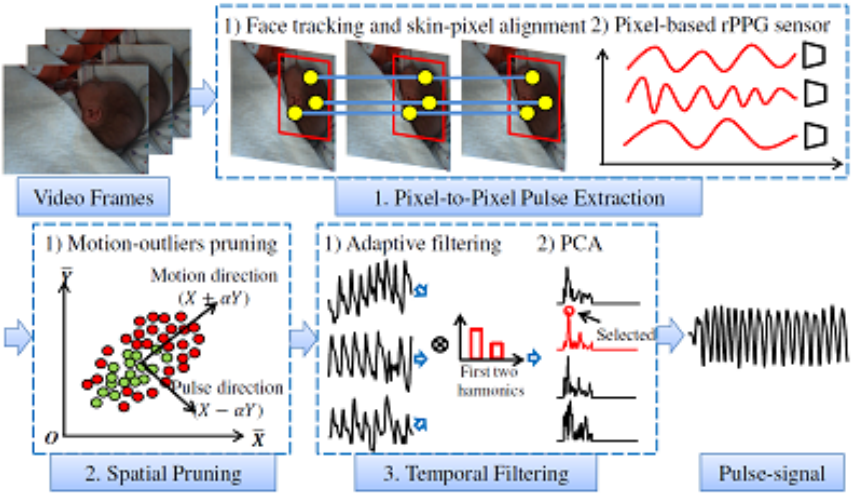


Figure 3.3: Framework of motion robust rPPG method for HR extracted from [73]

from RGB values for every skin-pixel. This method, described in [74], proved to be more robust to subject motion (in comparison with ICA and PCA) since it can split the variation in reflected light intensity caused by either blood volume differences or differences in specular reflection due to motion. By spatially representing these two components it is possible to eliminate motion-induced outliers. This study innovates in its temporal filtering step. While most methods use a static bandpass filter, *Fernando et al.* innovate by applying an adaPtive filter which strengthens the frequency inherent to the pulse. Finally, and similarly to other state of the art methods, FFT is applied to convert the signal to the frequency domain and select the peak as the subject's heart rate. This process is performed under a 8s time window allowing renovation and update of the subject's HR.

A more recent method, which comprises HR and RR extraction in neonate's was published by *Antognoli et al.* in 2019 [75]. For the benefit of robustness several simplifications were imposed, such as the use of a dedicated light source to inhibit the influence ambient light. This study was innovative in the sense that it applied a well known technique, Eulerian Video Magnification (EVC) to the NICU environment. EVM is a technique which amplifies motion or color variations in time, enabling the visualisation of imperceptible information to the naked eye [67]. *Antognoli et al.* start by selecting the 10 second portion of each video which includes less subject motion or ligh variations. A ROI is then selected manually in such a way that the infant's thorax is captured in its full extension. The output of the EVC algorithm is an equivalent to the input video in which motion due to respiration as well as color changes due to pulse are magnified and clearly visible. This resulting video is then used in a similarly pipeline as all other methods previously described. The RGB values for the pixels inside the ROI are averaged into 3 distinct raw signals, of which the desired component was extracted. Since a Butterworth pass-band filter had already been already used in the magnification step, filtering the raw signals for the intended frequencies would be redunant. *Power Spectral Density* was then applied to the said component and the resulting peaks

in the desired frequency ranges were detected as HR and RR values for each video.

Although the results of this study allegedly match or even outperform the gold-standard methods when in comparison with measures directly taken by the physicians its performance is limited in real world conditions. First of all, EVM based methods are computationally expensive and therefore make real-time analysis impossible, specially when using big-sized temporal windows. Secondly, and in the same way as *Motion Guided Methods* for HR extraction, this method is extremely sensitive to variations external conditions, namely lighting variations or even camera and crib shaking, which can nullify the measures when they occur.

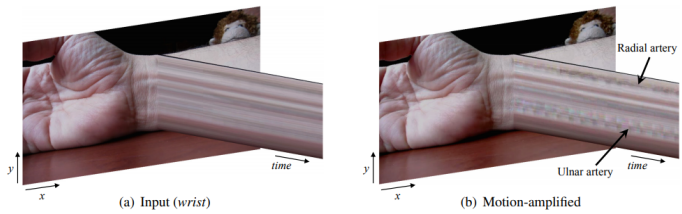
3.5 Change detection and Magnification

Video magnification techniques are useful for visualizing small changes in videos, whether caused by motion or color alterations. Various approaches have been considered in order to reach the goal of subtle change magnification.

In 2005, *Liu et al.* [76] created a technique capable of analyzing and amplifying subtle motions and visualize deformations that would otherwise be invisible. The framework presented firstly segments a reference frame into regions grouped by proximity, color homogeneity, and correlated motions. The user identifies the portion to be amplified, allowing for the video to be re-rendered with the subtle changes of the desired segment magnified. In a similar way, *Wang et al.* [77] proposed in 2006 the *Cartoon Animation Filter*, a method to exaggerate motion within a video sequence in a perceptually appealing manner. Both methods follow a Lagrangian perspective, a concept commonly used in fluid dynamics. In this prespective, the trajectory of particles is tracked over time. For that reason, both methods rely on accurate motion estimation, which results in a computationally expensive framework.

Eulerian Video Magnification (EVM) appeared as an alternative to Lagrangian approaches for both motion and color variation magnification in videos. This method, published by *Wu et al.* [67], makes use of Laplacian pyramids decomposes the input video sequence into different spatial frequency bands. The same temporal filter is then applied to all bands, which are then amplified by a given factor and added to the original signal. All levels of the pyramid are then collapsed, generating the output video. This method, which is described in more detail in Chapter 6, is capable of amplifying small motion even though motion is not tracked as in the Lagrangian methods previously presented. Figure 3.4, shows an example of the aplicability of EVM. By applying EVM to the video represented it is possible to amplify the movement of the arteries caused by BVP, a phenonmenon explained in Section 3.2.

Due to the increased popularity of EVM and its relatively low computation cost, recent studies on subtle change magnification have focused on improving the method created by *Wu et al.* [67]. For instance, *Liu et al.* [78] developed a method which makes use of EVM as a spatio-temporal motion analyzer to get the pixel-level motion mapping. It then magnifies the temporal video motion by warping the images based-on the previous motion mapping. This technique was proved



(a) Input (*wrist*) (b) Motion-amplified

Figure 3.4: Eulerian video magnification used to amplify subtle motions of blood vessels arising from blood flow, extracted from [67].

to improve the results generated by traditional EVM, once it supports larger amplification factors while being significantly less influenced by frame noise, as it does not involve pixel value modifying.

3.6 Summary

Scientific research in the last decade has resulted in countless contributions in the field of contactless extraction of vital signs, particularly HR and RR. Although two different types of techniques have been explored by the scientific community, the latter prevails as the most used methodology by far. As regards neonatal specific applications, rPPG methods have gained popularity due to its higher robustness to subject and camera movement when in comparison with motion-based methods. Nevertheless this applications have drawbacks such as the need for a technique to detect which of the signal's independent components reflects the pulsatile signal, which may not be trivial. EVM appears as a solution with high potential, despite its high computational cost (still lower than other alternatives), which may be a hinder for real-time analysis. Its advantages are the fact that it allows visualisation of the magnified movement, being particularly useful to help families and caregivers notice the breathing patterns of the infant.

Chapter 4

Dataset

Since the global situation posed as a preclusion to the acquisition of the videos at CMIN (and consequent creation of the neonate dataset), the need for a public dataset arose. However, the nonexistence of datasets with infant subjects has obliged to the use of datasets composed by videos of adult subjects. The most Common public datasets used in rPPG problems are the MAHNOB-HCI-Tagg [79]) and the COHFACE dataset[1]. The first was not originally created specifically for rPPG algorithms, but for characterisation of multimedia content based on human emotions. For that purpose, video and physiological data from 30 subjects was collected while being exposed to different audio-visual stimuli. Among those physiological parameters are the ECG and respiration amplitude signal, which allows the use of this dataset for contactless Heart and Respiratory Rate algorithm development. For the acquisitions, professional cameras and lighting setups were used. The second dataset, contains 160 one-minute long RGB video sequences of 40 healthy subjects (12 females and 28 males) in 2 different lighting conditions: natural light and studio lighting. Despite the advantages, such as the high number of participants and the possibility of comparing our results with most rPPG papers, which are inherent to these datasets' usage, legal constraints prevented their utilization.

Alteratively, the dataset used for the work developed was the *Public Benchmark Dataset for Testing rPPG Algorithm Performance* created at the *Eindhoven University of Technology* [25]. Unlike the previously mentioned, this dataset aims to test rPPG tools' performance in challenging conditions. It is stated that the capabilities of rPPG technologies and its underlying theory is well established for simple environments, but not ready for real-world applications. Therefore, more recent studies on the area focus on improving the technique's robustness to external factors, specifically trying to reduce or negate the influence of the said factors in the algorithm's performance. This dataset was created in order to evaluate and benchmark algorithm's robustness by including videos under challenging conditions or factors. The factors incorporated in the dataset are the most challenging and more common in real world scenarios. According to the literature, these factors are lighting conditions, subject skin color, high and fluctuating heart rates and presence of motion, and therefore those are the conditions covered in this dataset. These conditions were

[1] https://idiap.ch/dataset/cohface

used to design 3 challenges, which intend to answer a set of research questions about the tool to be tested:

- *Challenge 1* has to do with lighting conditions and skin tone and is supposed to infer how light intensity, light temperature, uneven light and skin tone variations affect the measurement accuracy of the rPPG tool.
- *Challenge 2* considers the influence of both global and rhythmic subject motion on the accuracy of the rPPG tool.
- *Challenge 3* investigates how high and fluctuating heart rates affect the tool's performance.

The dataset consists of a total of 21 videos of three healthy male participants, and simultaneous ECG measures to serve as Ground Truth. In order to evaluate the tools' response to each condition independently (and therefore its performance for each challenge), each of the 21 videos is addressed to one and one challenge only:

- Challenge 1 (Lighting Conditions and Skin Colour) consists of 17 videos: P1LC1, P1LC2, P1LC3, P1LC4, P1LC5, P1LC6, P1LC7, P2LC1, P2LC2, P2LC3, P2LC4, P2LC5, P3LC1, P3LC2, P3LC3, P3LC4, P3LC5.
- Challenge 2 (Motion) includes 3 videos: P1M1, P1M2, P1M3.
- Challenge 3 (High and Fluctuating Heart Rates) contains 1 video: P1H1.

While *Challenge 1* is represented much more extensively than the other two, accounting for more than 80% of the dataset, *Challenge 2* and *Challenge 3* only possess 3 and 1 videos, all of which of the same participant. Each video's identifier is defined by concatenating the patients' ID (P1, P2 or P3), with the ID of the condition to be tested (LC1 to LC7 for the lighting conditions, H1 for the high and fluctuating heart rates and M1 to M3 for subject motion). From this moment forward, each video will be referred to by its identifier.

All 21 videos were recorded with with the JVC GZ-VX815BE HD video camera. The participants were seated behind a table making sure that they were in the centre of the lab. A head rest was used for the videos in *Challenges 1 and 3* to eliminate movements of the head. The chin holder of the head rest had a height of 31 centimetres relative to the desk it was placed on and the camera was positioned exactly opposite to the head rest at a distance of 80 centimetres. The camera was placed on a platform 31.5 centimetres high relative to the desk. The videos were recorded at 25 fps, with UXP video quality and at a resolution of 12 MP. The camera was equipped with a F1.2 bright lens making it suitable for recordings at low lighting intensities. After editing, the videos were exported as *avi* video files with a resolution of 1080 x 1920 and a frame rate of 30fps using the H264 video compressor. A 1-lead ECG was provided, from which the ground-truth HR curve was extracted. An example of the 1-lead ECG provided can be seen in Figure 4.1. The ECG was measured using the Mobi electrocardiograph, which has a sampling frequency of 1024 Hz. This device was connected to a laptop with a Bluetooth receiver especially tested for the Mobi. The

Mobi is CE certified (class 2A, type CF), meaning it is cleared for medical usage in the EU. The data was synchronised using the sound of the button on the Mobi in the Video, which indicates the beginning of the recording of the ECG device. All videos, except for that of the third challenge (video P1H1) had a duration of three minutes. The video entitled P1H1 addressed high heart rates and for that purpose had a duration of five minutes: the time it took for the patient's heart rate to stabilize.

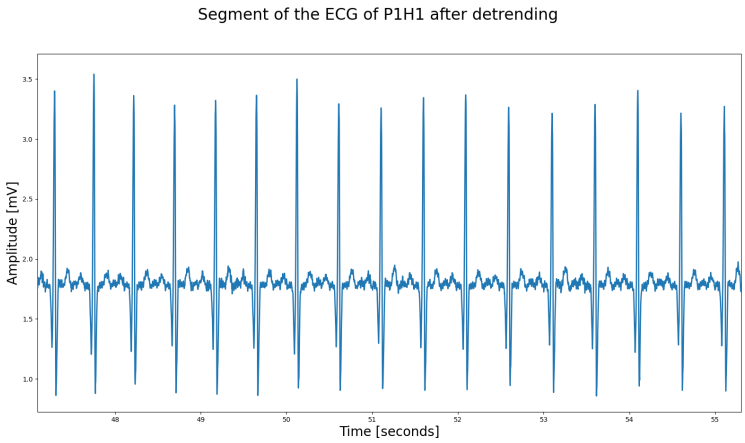


Figure 4.1: Segment of the one-lead ECG provided as Ground Truth for video P1H1 after detrending.

For the first 17 videos, *i.e.* those regarding *Challenge 1*, different lighting conditions were simulated. All recordings took place at a lab of the Technical University of Eindhoven, which is sealed from external light sources and equipped with 6 *Philips Savio* wall fixtures (4 in the wall facing the subject and 2 in the wall at their right) and 30 *Savio* fixtures in the ceiling, whose intensity can be set for values between 87 and 255. In order to obtain the different lighting conditions (7 in total: LC1, LC2, LC3, LC4, LC5, LC6 and LC7) different combinations of lights were turned on at a time and their intensity regulated. While LC1 to LC5 address increasing overall light intensity, LC6 alters light temperature (2700K versus 6500K for all other videos) and LC7 makes use of the physical distribution of the wall fixtures to create uneven light distribution. The Light intensity level for each LC was determined using the pocket-lux device by *Lichtmess Techniek Berlin*. Table 4.1 summarizes the characteristics of every LC included in the dataset.

Still regarding the first challenge, for the first five LC's, videos were recorded of all three participants (of ages 21, 27 and 31), who have different complexions: P1 having light skin, P2 intermediate skin and P3 dark sin. By combining these two factors (Light condition and Skin tone) it is possible to determine their influence in performance both separately and jointly. In addition, for Light conditions 6 and 7 videos were recorded of Patient 1 (P1). Figure 4.2 exhibits frames extracted for each video associated to *Challenge 1* (Light Conditions and Skin Tone).

When it comes to *Challenge 2*, three videos were recorded. Only patient P1 took part in this portion of the protocol and Light Condition 4 was set for all three videos, as it was considered to be the most neutral condition for rPPG measurements. The subject was seated behind the desk

Table 4.1: Detailed description of all Lighting Conditions comprised in the dataset.

Lighting Condition (LC)	**Light temperature (Kelvin)**	**Ceiling fixtures**	**Southern wall fixtures**	**Western wall fixtures**	**Light intensity (lux)**
LC1	6500	Off	87	Off	.052 x 100
LC2	6500	87	87	Off	.363 x 100
LC3	6500	143	143	Off	1.870 x 100
LC4	6500	199	199	Off	7.20 x 100
LC5	6500	255	255	Off	27.2 x 100
LC6	2700	100	100	Off	.349 x 100
LC7	6500	Off	Off	199	.180 x 100

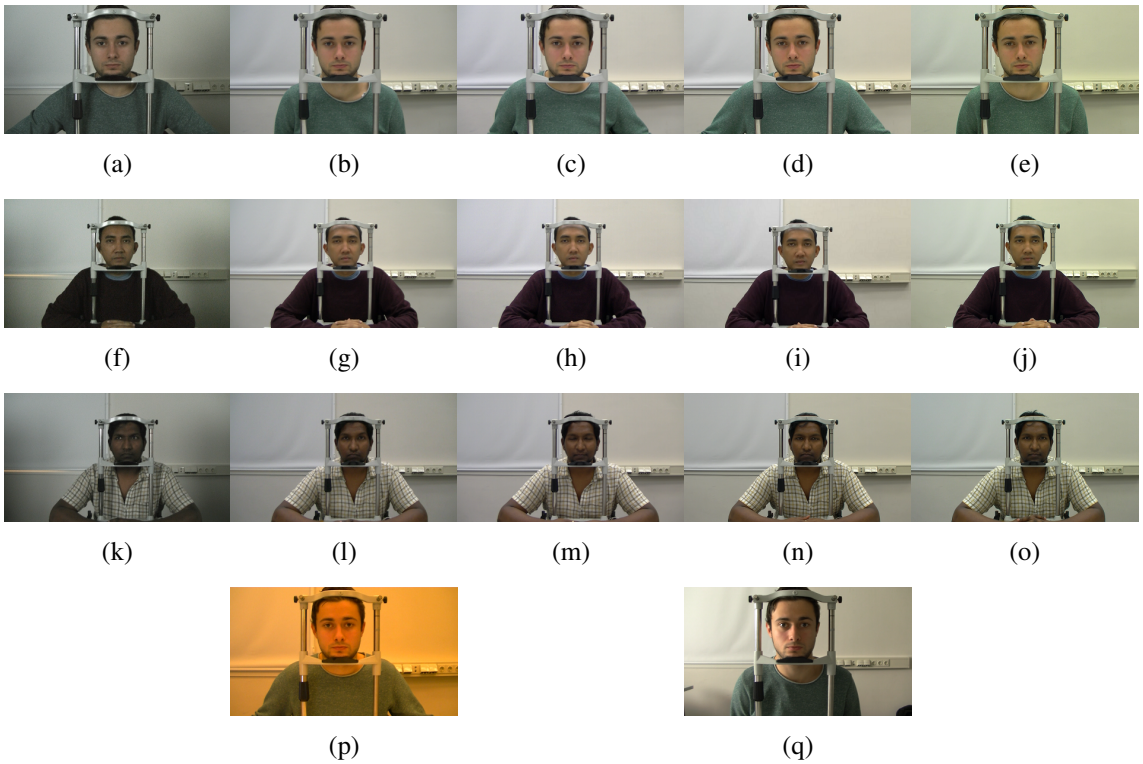


Figure 4.2: Sample frames from all videos comprised in *Challenge 1*: (a) P1LC1, (b) P1LC2, (c) P1LC3, (d) P1LC4, (e) P1LC5, (f) P2LC1, (g) P2LC2, (h) P2LC3, (i) P2LC4, (j) P2LC5, (k) P3LC1, (l) P3LC2, (m) P3LC3, (n) P3LC4, (o) P3LC5, (p) P1LC6, (q) P1LC7.

without the head rest and asked to move his head freely for video P1M1. This video aimed at testing simple movements and rotations of the head as if the participant was looking around. For the other two videos, the subject was asked to make continuous nodding movements at specific frequencies. In P1M2 the motion frequency was set to 60 bpm (a frequency that on average matches the participants' HR), while in P1M3 the nodding frequency was set at 90 bpm, which should fall outside of the selected frequency bandwidth. These videos will test whether the algorithm sees the signal originated from the rhythmic motion as noise or mistakes it for the HR. In both recordings the participant was guided by a metronome played on a smartphone.

For *Challenge 3*, only Participant 1 was recorded and the LC was once again set to LC4. Immediately before recording the video the participant is asked to run back and forth for 3 minutes before being asked to seat and place his head on the head rest.

Logical reasoning is enough to draw predictions on how each condition will influence the tools' performances. When it comes to light intensity, it is hypothesized that increasing intensity will have a strong positive effect on the accuracy of the rPPG measurements, since lower light intensities translate into noisier images. Uneven light conditions may also reduce signal quality since it will introduce heterogeneity in the subject's face and ROI, which will possibly affect the HR measures. When it comes to light temperature, although it has been discovered that rPPG signal strength can differ between the RGB channels due to the differences in the spectral light intensity distribution, it is expected that the obstacle of light temperature can be surpassed by using Blind Source Separated signals or Chrominance Signals instead of the classic signal from the green channel only.

4.1 Additional Material

In order to enable variety in testing and to informally test the framework in its early stages of development, four videos were recorded with a DSLR camera (*Canon EOS 70D*) coupled with a lens with a fixed focal distance of 50 mm and an aperture of f/1.8 positioned at the same height of the subject's eyes and a distance of 1 meter. These videos consisted of only one subject under both frontal and artificial light and uneven natural light conditions. For both lighting conditions, one of the recordings was performed immediately after a short exercise session which intended to increase the subject's heart rate. Once the ground truth was obtained from peripheral pulse palpation at the carotid region, method which is not considered to be unreliable [80], the results obtained from this video were not considered.

4.2 Validation Dataset

After the period defined for the writing of this dissertation had finished, social constraint policies were partially withdrawn, which allowed to start the acquisition process. For that reason, a session was held at CMIN to acquire videos of newborns inside both cribs and incubators alongside with the respective ground truth at the Intermediate Care Unit. This acquisition session resulted in the creation of a database which counts with the authorized participation of 6 newborns. For each of the individuals, one or more five minute videos were recorded with a *Kinect V2 for Windows* and the acquisition protocol can be seen in Appendix C. As the pulse oximeter was the only sensor available this was used to obtain the ground truth. Because of this, no measures could be taken to infer about the Respiratory Rate and for that reason the extraction of this vital sign could not be validated in newborn individuals. A second camera was used to record the pulse oximeter's monitor, and the HR ground truth was obtained by using the *Tesseract OCR engine* [81] to extract the digits inherent to this vital sign for every frame. The two videos were later synced.

As opposed to dataset used for the tool's development, in this dataset each video integrates a combined set of challenges as a result of the uncontrolled acquisition conditions. Furthermore, in addition to the challenges which were expected to be found in a NICU and are covered in the development dataset, other challenges were encountered and can be found in this dataset. For these reasons, this dataset represents a much more challenging set of samples, which best represents the real-world conditions. For legal purposes no images of the acquired videos may be presented as these represent sensitive and personal information.

4.3 Future Testing

Nevertheless, both datastes used present flaws and their use should be complemented with other datasets. Regarding the dataset used for development, the fact that each condition is only represented by one video makes it impossible to assess reproducibility and consistency. Furthermore, by only including three participants, the dataset may miss out on certain conditions that might have impact on real-world applications, such as age, scarring or even intermediate skin tones. Regarding the actual scope of this master thesis, which is applying contactless HR and RR extraction methods to neonatal participants, one could argue that the predominant challenges to be faced differ from those represented in this dataset. For instance, although head motions may occur, its magnitude is not usually as exaggerated as those exhibited in this dataset. Most head movements performed by a baby in an incubator are sparse rather than rhythmic and occur when changing positions. Similarly, increasingly darker skin tones should not be so challenging firstly because neonates' skin (in particular in preterm infants [82]) is thinner than that of adults and therefore the Blood Volume Pulse phenomenon should be much more evident across all complexions. Secondly, the darkening effect of the skin with age [79] makes the difference in skin tones more evident between adults than between neonates.

On the other hand, since the reference ranges of HR and RR of newborn infants (and children in general) are much higher than that of adult's, the third challenge imposed by this dataset (High and Fluctuating Heart Rates) should be addressed more carefully, once the tool must be able to perform for elevated pulse values. Furthermore, lighting conditions, in particular uneven lighting, would be one of the most probable challenges encountered in a NICU environment and thus, it is important to assure that the developed tool does not falter in such conditions. One untested challenge which might induce major performance issues is specular reflection caused by the incubator's hood. To test these premises would require extensive recordings of neonatal participants in a regular NICU environment.

Chapter 5

Heart Rate

The framework developed in this study relies in its entirety on the HR extraction process and its secondary products, in the sense that the signals and measures resulting from it will be used as inputs for other portions of the algorithm, namely for RR extraction, as described in Chapters 7. For this reason, the development of the algorithm regarding HR extraction was more extensively developed, since the accomplishment of its purpose would define the success of the following portions of the framework. Furthermore, the nature of the dataset used, namely the fact that it was designed specifically for evaluating contactless HR extraction tools, allowed a wider range of tests than those which could be performed for the other portions of the framework presented.

HR is one of the most informative vital signs, and its acquisition and monitoring is indispensable to assure every preterm infant's well-being inside the NICU. It is true that the benefits of contactless HR extraction extend way beyond the NICU, however the impact caused by substitution of the regular methods is way more significant for newborns than for adults. Besides, despite the abundance of studies covering contactless HR extraction for adults, there are still few studies on its effectiveness in newborns, specially focusing on videos acquired by low-cost cameras. For that reason, despite having been developed on a dataset consisting solely of adult subjects, the construction of the concerned method was guided on every decision by what would be most beneficial for preterm subjects.

5.1 Methodology

As mentioned in Chapter 3, the typical contactless HR extraction tool is composed of a series of sequential processes each of which has a particular purpose. Every process (from now on referred as module) makes use of the result of the previous module to determine its own output which will be then used by the following module, forming a linear combination of techniques which ultimately result in the extracted HR. Since rPPG-based tools for HR extraction first appeared, the order and the purpose of all modules have been firmly established. As a consequence, among recent published methods there is little variation in the modules used and their order. What does change between studies are the techniques used for each of the modules. For instance, all published

studies present a module for Region of Interest (ROI) definition following a face detection and tracking module. Nonetheless, new methods for ROI definition, including different combinations of skin patches arise frequently.

For that reason, the typical framework was adopted and several techniques were applied and compared for each module. Figure 5.1 shows the overall pipeline of the HR extraction method, specifying the modules utilised as well as the various techniques tested for each of these. All techniques employed will be further explained in more detail. In order to provide the continuous monitoring of the HR, the proposed method uses a temporal sliding window whose length is defined by the user. For each position of the window, the rPPG signal is extracted, from which a single HR value is determined. The window is then moved by a step also defined by the user and the process is repeated. In the end, the output of this pipeline is a continuous HR evolution curve, whose length is equal to the number of positions taken by the window until the end of the video. It is important to notice that this tool has a buffer period with the same length of the window.

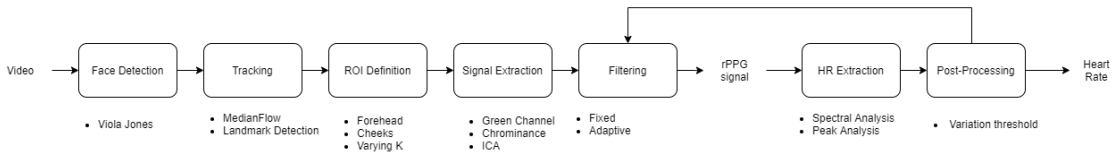


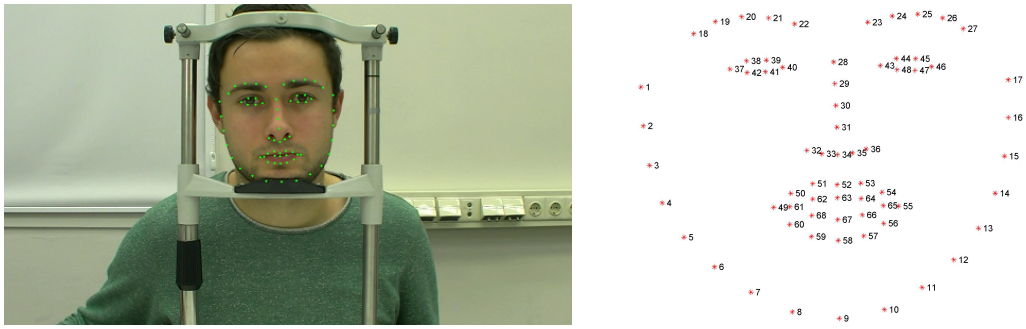
Figure 5.1: Framework for Heart Rate measurement from face videos. For each module, the various techniques employed are enumerated.

5.1.1 Face Tracking and Region of Interest Definition

The first step in any contactless HR extraction algorithm is face detection. This task, performed by the acclaimed Viola-Jones algorithm [60], results in a bounding-box around the subjects face for the first frame of the video. The position of that bounding box is then tracked using Median Flow [83]. Despite the widespread acceptance of the Viola-Jones algorithm, its use comes with a series of disadvantages, the main one being that it can only detect frontal faces. This fact compromises this technique's use in a NICU environment, since the infants usually have their head rotated laterally while inside the incubator. For this reason, an alternative method had to be tested. Furthermore, the use of Median Flow is not robust enough for big movements (such as those presented in P1M1, P1M2 and P1M3), owing to the fact that this technique struggles to find its target once lost, which happens frequently with sudden movements [84].

The use of the previously mentioned combination of algorithms was therefore compared to the detection of facial keypoints, which should be more accurate while not compromising in robustness. The keypoint detection algorithm used was the DLIB python library's implementation of Kazemi et al's [85] method, which had previously been trained on the iBUG 300-W face landmark dataset [86]. This method uses a a cascade of regressors to estimate the position of 68 facial landmarks in a computationally efficient manner. Figure 5.2 shows the result of the keypoint detection as well as the identifier of each fiducial point for further referencing. The fact that the

keypoint locator takes less than $\frac{1}{30}$ seconds, makes it possible to perform keypoint detection for every frame, instead of a tracking alternative, which prevents cumulative error propagation from frame to frame.



(a) (b)

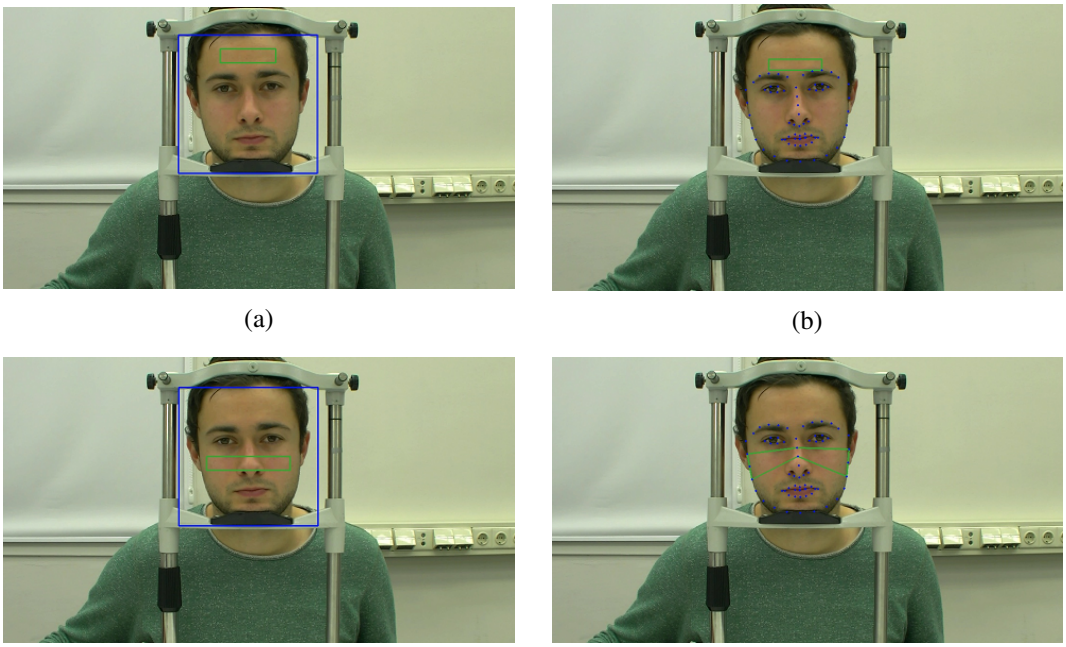
Figure 5.2: (a) Example of facial keypoint detection performed in a frame of video P1LC5; (b) Identifiers of the 68 facial keypoints detected.

Once the location of the face or its fiducial points is known for each frame, one can situate particular regions of interest within the frame relative to the known points' locations. The precision of this ROI definition is highly dependant on the previous tracking steps and its robustness. In order to define the ROI location within the frame it is first important to select which facial structure it is that deserves our methods attention. Both the forehead, the inferior portion of the face, the cheeks region have been reported in the literature, as well squares containing the entirety of the face, although the latter has passed out of use due to the amount of non-skin pixels it encompassed, namely in facial hair, eyes and nostrils. Both the forehead and cheek regions were tested for being the ones less-susceptible to non-rigid motions, induced by facial expressions, as they exclude the mouth and eyes areas. This detail is particularly detereminative when it comes to dealing with newborns since their facial expressions are more exaggerated than that of an adult and they spent a considerable amount of time crying and frowning.

In addition to the choice of the facial region itself, it is also possible to define the ROI relatively to different sets of points. A simpler approach would be to define ROI's location with reference to the location of the face's bounding box. In this line of thought, the ROI's was defined as a simple rectangle, whose width was set as 60% of the width of the face's bounding box, once the bounding box from the Viola-Jones algorithm typically includes background pixels on either sides. The height of the ROI was set to 20% of that of the bounding box. In case the forehead was chosen to represent the ROI, the rectangle was aligned with the top of the face's bounding box, whereas if the objective was to map the cheeks area the rectangle inherent to the ROI was vertically centered. In both cases, the ROI was horizontally centered within the face's bounding box.

Nonetheless, this approach is evidently less robust than defining the ROI's location according to the position of the facial fiducial points, mainly because the bounding box assumes a frontal face and therefore pays no attention to rotations, preserving the ROI's shape and dimensions when situations when these should change. To define the forehead according to the facial keypoint's

locations, keypoints 19 and 24 longitudes were used to define the ROI's lateral limits and their latitude to define the ROI's inferior limit. Once the highest keypoints from the keypoint library were in the eyebrows, there was no information which could be used to limit the rectangle superiorly. The ROI's height was then set as a function of its width to enable adequate scaling if the subject moves away or to the camera. Figure 5.3 shows the different types of ROIs tested according to the facial region it maps and how they were calculated.



(a) (b) (c) (d)

Figure 5.3: Examples extracted from P1LC5 of all the different ROIs tested and its different definition methods: (a) forehead ROI determined in relation to the face's bounding box; (b) forehead ROI determined in relation to the facial landmarks; (c) cheeks ROI determined in relation to the face's bounding box; (d) cheeks ROI determined in relation to the facial landmarks;

When it comes to the cheek ROI and its definition process relative to the fiducial points, a concave hexagon was defined with keypoints 2, 4, 30, 14, 16 and 28 as vertices. Piece-wise linear wrapping is then applied to wrap the hexagon into a rectangle of fixed shape. For the reason that each facial landmark has a particular semantic meaning, we can assume that the wrapping transformation makes each pixel in the resulting rectangular ROI is aligned. Figure 5.4 displays the cheek ROI after being wrapped to the said rectangle.

Following the wrapping of the cheek area into a rectangle, a skin segmentation step was also included. Although this was not strictly necessary for adult subjects (Figure 5.4 shows that all of the pixels covered represent skin areas), this step will prove to be extremely convenient when dealing with preterm infants, specially those who are under artificial ventilation conditions. It is common for preterm infants who require intensive care to be ventilated using a continuous positive airway pressure (CPAP) device, as mentioned in Chapter 2. This device is usually white, and is connected to the newborn's airways via the nose. It possesses a tube which usually passes in front

Figure 5.4: ROI corresponding to the cheeks after it has been wrapped to a rectangle.

of the babies forehead covering most of it and is fixed to the infant's head through white straps which wrap around the infant's face obstructing part of the cheeks.

Despite being ideal surfaces for visualisation of the BVP phenomenon due to the high levels if irrigation from superficial blood vessels, the selected ROIs map an area which is considerably big and lacks colour homogeneity as is made evident by Figure 5.4. Such incongruity may be mostly due to three factors:

- Localized blushing.
- The fact that the face rounds and therefore cannot be considered a perfect *Lambertian surface*, (i.e. it reflects light differently according to its orientation to the light source).
- The rough relief of the face makes that it receives light in an uneven manner. Specially when the light source is not directly in front of the subject (Uneven light conditions, such as P1LC6 or what would be expected in a NICU environment),the nose will project shadows on the cheeks, for an instance.

All of these result in a high standard deviation for the RGB values of the pixels inside the ROI, which will harm the signals quality, as will be proved in Section 5.2. In order to surpass this, the ROI is divided into K smaller rectangles, which will function as independent ROIs and will be referred to as sensors from this moment on. An increasing number of sensors (K) should translate to smaller average values of standard deviation for each sensor, justified by the fact that they are decreasing in size and therefore should map a much more homogeneous patch of skin, as will do be discussed later on in this chapter. Of course this premise only verifies as long as the sensors are as close to a perfect square as possible. To assure that and at the same time allow flexibility in the values of K, we let K be a user defined parameter and calculate the sensor size according to the following equations:

$$K = \frac{WH}{s^2} \Leftrightarrow s = \left\lfloor \sqrt{\frac{WH}{K}} \right\rfloor \tag{5.1}$$

where W and H denote the width and the length in pixels of the big ROI and s becomes the value (also in pixels) to each the nearly squares' width and height is going to be approximated to.

$$A = \left\lfloor \frac{W}{s} \right\rfloor \wedge B = \left\lfloor \frac{H}{s} \right\rfloor \tag{5.2}$$

One can then obtain the number o columns (represented by A) and number of rows(represented by B) by diving W and H by the size of the side of the ideal square. A and B must be rounded down to the nearest integer to guarantee that the the last row or column does not have a fraction of its the desired size, therefore becoming an elongated rectangle. This operation comes with the cost of the possibility of having a few pixels which are left out bu that drawback is not enough to justify the use of portions of rows and columns.

$$w = \left\lfloor \frac{W}{s} \right\rfloor \wedge h = \left\lfloor \frac{H}{s} \right\rfloor \tag{5.3}$$

The near-squares actual dimensions can then be calculated by dividing the integer number of columns W and rows B by s. The general ROI can then be segmented into K_{approx} sensors distributed over a grid of A columns and B. Each sensor acquires a near-square shape and has w as width and h as height.

5.1.2 Signal Extraction

After the ROI and its sensors are defined, average pooling is performed for each of the three RGB channels and for each sensor independently. This means that $3 \times K_{approx}$ 1D time-signals will be generated. According to the literature, and as mentioned in Chapter 3, of the three RGB channels the one which best reflects the BVP phenomenon is the green channel, which is clearly visible in Figure 5.5 as the peaks in the green channel are more equally spaced and differentiated than in the other two channels. The signal extracted from this channel has been reported to be enough for HR extraction in certain conditions.

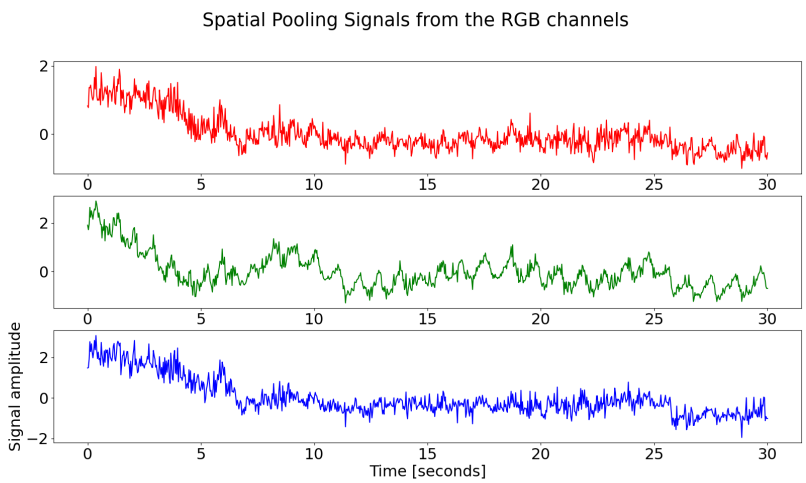


Figure 5.5: Comparison of the raw signals extracted by spatial pooling of each of the RGB channels.The top signal is inherent to the red channel, the middle one to the green channel and the bottom one to the blue channel.

However, using only this channel seems to be fragile in more harsh conditions and presumably will falter in conditions of non-white illumination and darker subject complexions. Therefore, two

other signals will be generated from the spatially pooled RGB signals and tested. These signals are:

- Blind Source Separation (BSS) rPPG signal
- Chrominance rPPG signal

5.1.2.1 Blind Sourse Separation rPPG signal

BSS is a technique used for separating a set of signals into its unobserved sources or original components, while having no prior information about the mixing process [87]. For instance, if three microphones are set in a recording room each of which capturing the overall sound of the room, where three musical instruments are being played simultaneously, BSS can theoretically be used to separate the signals from the three microphones into the three independent signals from each of the instruments. In the context of our problem, BSS can be used to separate the three RGB signals into its source components, which should reflect distinct origins. From the three resulting signals, one should translate the variation induced by BVP, thus being the rPPG signal and the other two signals should encode noise originated by motion or lighting variations. The BSS algorithms most described in the literature is ICA, mentioned in Chapter 3. ICA works by finding a linear representation of non-Gaussian data so that its components are statistically independent, or as independent as possible [88]. Despite all its benefits, the use of ICA presents one flaw, which is that the component which carries the pulse signal is a priori unknown as the components are presented in no particular order. Therefore, for every sample, the need to select (out of the three components) which represents the rPPG signal arises.

That being said, ICA was applied to the three RGB signals in order to decompose them into their independent components. Several methods for component selection were compared, those being:

- Maximum intensity predominant frequency response
- Maximum correlation with reference sine signal
- KNN and SVM classifiers combining frequency spectra and time domain features.

Commonly, the selection process assumes that the pulse signal shows the strongest periodicity. Two distinct selection methods can be used based on this fact. The first one, more abundant in literature, is to chose the component which has the highest frequency response for its predominant frequency. In practice, this means that the FFT is computed for each component and its predominant frequency selected (by locating the frequency response's peak). The component which presents the highest value is considered the rPPG signal.

One other method addressing periodicity is to compare each component to a reference sine as suggested by Feng etal. [89]. For that purpose, the FFT is computed for all three components and its predominant frequency is once again extracted. Three reference since are then created,

each with frequency equal to those extracted. Pearson's correlation coefficient is then calculated for each pair of components and respective reference sign after alignment. The component which correlates the strongest to its reference sine signal is thus the more periodic and adopted as the pulse signal.

Nonetheless, both methods disregard HRV (a healthy phenomenon that represents a source of non-periodicity), which we intend to preserve in the signal for other vital signs extraction as explained in Chapter 7. Furthermore, in case of periodic motion, such as rhythmic head nodding, as portrayed in this dataset by the videos P1M2 and P1M3, these methods may overlook the actual pulse signal and mistake it by the motion induced independent component, which naturally is extremely periodic.

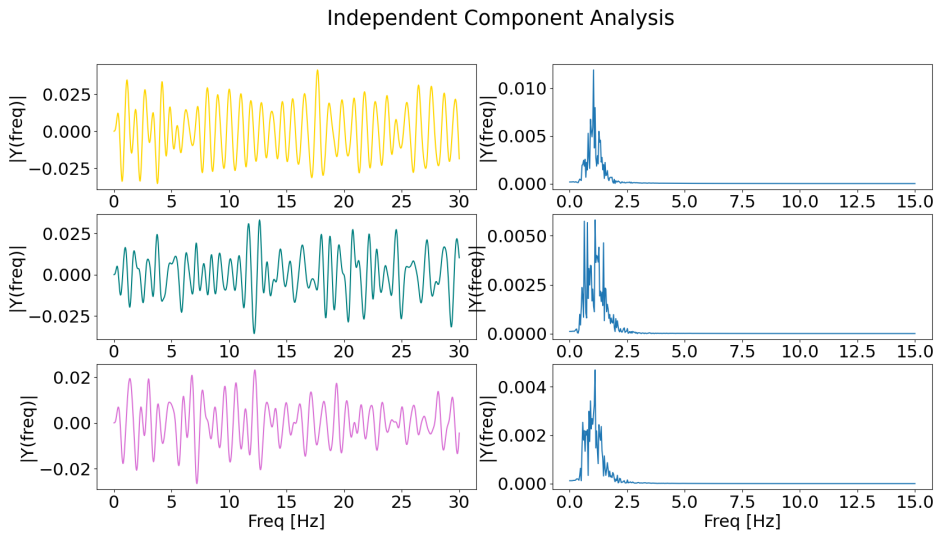
In an attempt to overcome this flaw a third and distinct method was tested. Similarly to what was described by Monkaresi et al. [87], we used Machine Learning classifiers to chose the pulse signal from the three independent components. The classifiers compared were SVM's and k-Nearest Neighbours (kNN). One difference between Monkaresi et al. [87] and the method employed in this work are the features adpoted. While the published paper mentioned the use of 9 features (3 per component), we fed the classifiers a total of 12 features (4 per component), which are:

- The energy of the most predominant frequency;
- The most predominant frequency;
- The correlation coefficient of the component to its reference sine;
- The standard deviation of the height of the component's peaks;

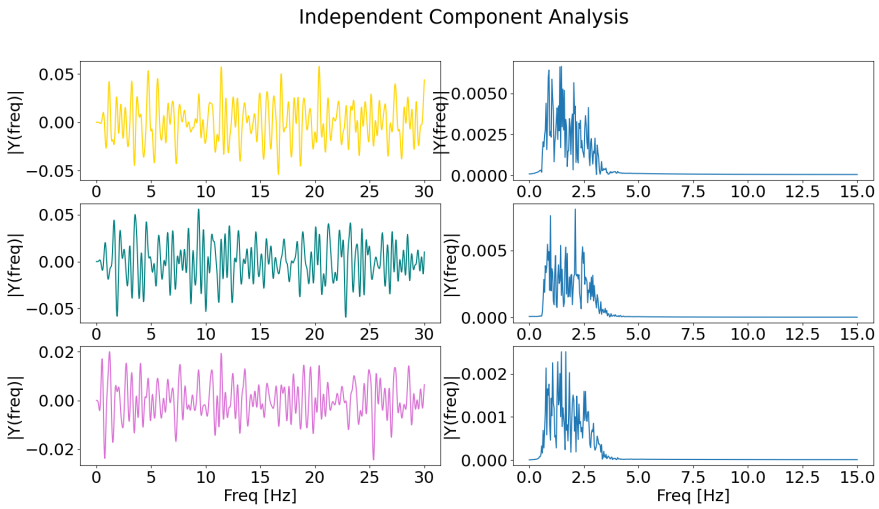
In order to train the classifiers, random samples from videos P1LC5, P2LC5, P3LC5, P1LC6, and P1H1 as well as samples from homemade videos which are not part of the used dataset (as described in , described in Chapter 4) were extracted and annotated. This set of samples was further divided into training and validation. Since it would be impossible to annotate all samples for all videos of the dataset, it was only possible to calculate evaluation metrics for the classifiers' performance using the validation samples. The overall performance of the classifiers was assessed through their impact on the general performance of the framework.

Not only the reduced number of samples used but also the difficulty to annotate some of the instances posed as a hindrance for the classifiers' training. Figure 5.6 shows two different instances which are quite contrasting in terms of annotation difficulty, due to the contrast between the components. For each of the instances (a) and (b) the three independent components (which resulted from ICA) are displayed. In the sample represented by Figure 5.6 (a) it is fairly simple to visually identify which of three components represents the pulse signal. Due to its increased periodicity, greater homogeneity in peak height and higher intensity in frequency response, the first component (on the top) is easily identified as that which represents the rPPG signal. On the other hand, for the sample represented in Figure 5.6 (b) the same task is nearly impossible, since

all three components look alike, and so do their power spectra. Samples such as that indicated in Figure 5.6 (b) are difficult to annotate with a high degree of certainty and may damage the process of classifier training, but should not be disregarded as they appear frequently in any of the videos of the dataset.



(a)



(b)

Figure 5.6: Independent Components and their Spectral Analysis for distinct samples (a) refers to a sample from P1LC5, whereas (b) refers to a sample from P3LC2

5.1.2.2 Chrominance signal

Lastly, chrominance based signals are those which convey color information usually by computing and relating color-difference components. In 2013, *de Haan et al.* [74] conducted a study to test the efficiency of Chrominance based signals in handling challenging conditions in rPPG problems,

in particular subject motion. In that study, they compared the performance of several chrominance-based methods (namely $RoverG$, $XoverY$, $Xmin\alpha Y$ [74]) among each other and with BSS signals (namely those resulting from ICA and PCA) and concluded that $Xmin\alpha Y$ was the best performing of all for the conditions tested. For that reason this will be the signal which will represent the category of Chrominance based signals in this pipeline and the tests to it performed. In order to calculate the $Xmin\alpha Y$ signal one must first obtain the X_s and Y_s signals according to:

$$X_s = 3R_n - 2G_n \tag{5.4}$$

$$Y_s = 1.5R_n + G_n - 1.5B_n \tag{5.5}$$

where R_n, G_n and B_n are the normalized versions of the Red, Green, and Blue channels respectively. The normalization of the signals consists in its division by its highest value in order for all values to be re-scaled to the range between 0 and 1.

$$S = X_f - \alpha Y_f \tag{5.6}$$

After calculating X_s and Y_s, X_f and Y_f are simply their bandpass filtered versions. Finally, the $Xmin\alpha Y$ signal (S) is determined as the difference between X_f and the product of α and Y_f, in which α stands for the ratio between the standard deviations of X_f and Y_f:

$$\alpha = \frac{\sigma(X_f)}{\sigma(Y_f)} \tag{5.7}$$

The final module of the Signal Extraction block is the most simple but probably the most important, without which HR extraction would become a much more difficult problem to be solved. This module consists on filtering the raw rPPG signal. For the Chrominance based signal, the filtering process is already embedded in the creation of the signal itself, therefore there is no need to redo that operation, but for the other two signals this step is done separately from its creation. The filter used was a 8th order Butterworth band-pass filter. When it comes to the definition of its cut-off frequencies two approaches were used. A common wide band fixed filter was used for all samples. In this approach the cut-off frequencies were set to $[0.7Hz - 4Hz]$ once these were the most commonly employed in the literature [56] [89] [90]. A more interesting technique was also employed. This technique, commonly referred to as adaptive filtering, relies on the fact that a limit exists to how much the HR can vary in a given interval. For this reason the adaptive filter's cutt-off frequencies were different for each sample and based on the determined HR of the previous sample. The bandwidth of the filter was therefore defined as $[HR - 30bpm, HR + 30bpm]$. Figure 5.7 shows how the effect of both fixed and adaptive filtering in a 30 second sample of the raw Green rPPG signal.

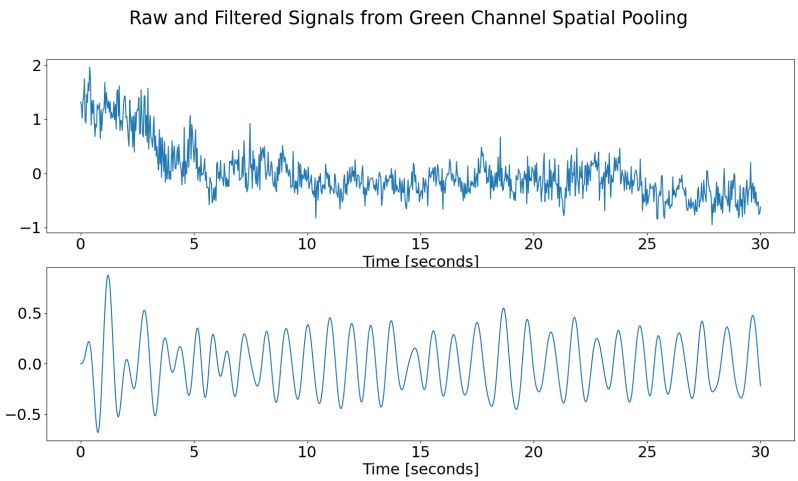


Figure 5.7: Comparison of the raw signal extracted from the green channel (top) with its bandpass filtered equivalent (bottom).

5.1.3 Heart Rate Extraction and Post Processing

The last step in HR extraction from rPPG signals concerns the extraction of the predominant frequency of the sampled signal. In this study, this was tested in two very distinct techniques, which have their own perks and drawbacks. The first and more robust method, most commonly used in the literature, was to perform Power Spectral Analysis [56], through the act of applying a FFT to the signal. The HR is then defined as the frequency with the highest response, determined by peak in the power spectrum of the signal. A second method was to analyse the rPPG signal in respect to its time domain and determine the position of the peaks (i.e. local maxima), which should mirror the beats of the heart. Through the calculation of all the intervals (in seconds) between successive beats, commonly referred as *Interbeat Intervals* (IBI), the frequency corresponding to each IBI was determined as its reciprocal multiplied by 60, in order to obtain a value in beats per minute (bpm):

$$f_i = \frac{60}{IBI_i} \tag{5.8}$$

The HR was then chosen as the median of the frequencies (f_i) inherent to all IBIs. The median was selected over the mean since the former is less sensitive to outliers.

The extraction step can be complemented with a post-processing module, which presents itself the form of a Variation Threshold. The Variation Threshold technique supports itself in the same principle as the adaptive filtering as it limits the possible variation between successive HR samples and assumes the previous sample in case the variation exceeds the threshold., such as had been done by Poh et al. [90].

5.2 Results and Discussion

The modular nature of this framework generates an overwhelming number of possible pipeline configurations to test, since for each module, several techniques were employed. Specifically, there are 96 possible pipeline configurations, each of which has an infinite number of variants according to the amount of different numbers of sensors (K_{approx}) tested. To be able to find the optimal combination of techniques without having to test all possible configurations, a cumulative approach for technique selection was used. This approach, inspired in wrapper feature selection methods often employed in Machine Learning problems, assures that the the optimal configuration which is being achieved considers method interaction, instead of addressing the modules' performance independently.

In our cumulative selection approach (which can be compared to a forward feature selection approach in wrapper methods), we start by evaluating the performance of the least sophisticated configuration of techniques (*Baseline Configuration*). After the *Baseline Configuration* is evaluated, module analysis is performed sequentially from the pipeline's starting module (Face detection and tracking) to its finishing module (post-processing). For each module, the various described techniques are ranked according to how they influence the pipeline's performance. Once the best technique for a given module is found, it replaces its homologous in the *Baseline Configuration*. As we advance through the modules, we create new configurations, which consist of the best techniques for the modules evaluated so far and the *Baseline configuration's* techniques for the modules yet to be evaluated. In the end, after evaluation is performed for every module, the optimal configuration is established. Table 5.1 summarizes the configurations created during this process and provides each of them with an identifier.

Since each module delivers a distinct contribution to the general framework, it would be inappropriate to evaluate their influence according to the same the standards. For instance, the purpose of the first modules (Face Detection and Tracking; ROI definition; Signal Extraction and Filtering) is to maximize the quality of the signal extracted so that the the last two modules (HR extraction and Post-Processing) can use it to accurately obtain the HR. For that reason, different metrics were used to evaluate different modules, according to the modules' purpose.

5.2.1 Face Detection and Tracking

The first two modules are responsible for locating the face within the frame and tracking its position over time. This is an extremely important step, since the location of the ROI, from which the rPPG signal will be extracted, will be obtained from the information resulting from this module. Failure in accurately and consistently defining the position of the face will resulted in a deficient ROI definition and consequent extraction of a meaningless signal, which will completely invalidate the extraction of the HR measures and therefor the whole framework.

For that reason, it is expected of the technique used in this module to be robust enough to locate the face's position and its orientation in a wide variety of conditions, which will for sure be encountered in a real-world NICU environment. The technique employed should be able not only

Table 5.1: Description and identifiers of the pipeline configurations tested

Identifier	Description	Similar to
Baseline	Viola-Jones and Median Flow + Forehead ROI + $K_{approx} = 1$ + Green channel rPPG signal + Fixed filtering + FFT + No post-processing	1
C1	Landmark Detection + Forehead ROI + $K_{approx} = 1$ + Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C2	Landmark Detection + Cheeks ROI + $K_{approx} = 1$ + Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C3	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C4	Landmark Detection + Cheeks ROI + $K_{approx} = 16$+ Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C5	Landmark Detection + Cheeks ROI + $K_{approx} = 50$ + Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C6	Landmark Detection + Cheeks ROI + $K_{approx} = 100$ + Green channel rPPG signal + Fixed Filtering + FFT + No post-processing	1
C7	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + Chrominance rPPG signal + Fixed Filtering + FFT + No post-processing	1
C8	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + BSS rPPG signal + Fixed Filtering + FFT + No post-processing	1
C9	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + Green channel rPPG signal + Adaptive Filtering + FFT + No post-processing	1
C10	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + Green channel rPPG signal + Adaptive Filtering + Peak Analysis + No post-processing	1
C11	Landmark Detection + Cheeks ROI + $K_{approx} = 9$ + Green channel rPPG signal + Adaptive Filtering + Peak Analysis + Variation Threshold	1

to detect and track and the face in challenging light conditions but more importantly to cope with head movements and eccentric head positions, once newborn's will be able to rotate their head while lying inside the incubator. Furthermore, the more stable the tracking method is, the better, as even slight flickering of the determined face's position may introduce noise in the rPPG signal, specially when dealing with small ROI.

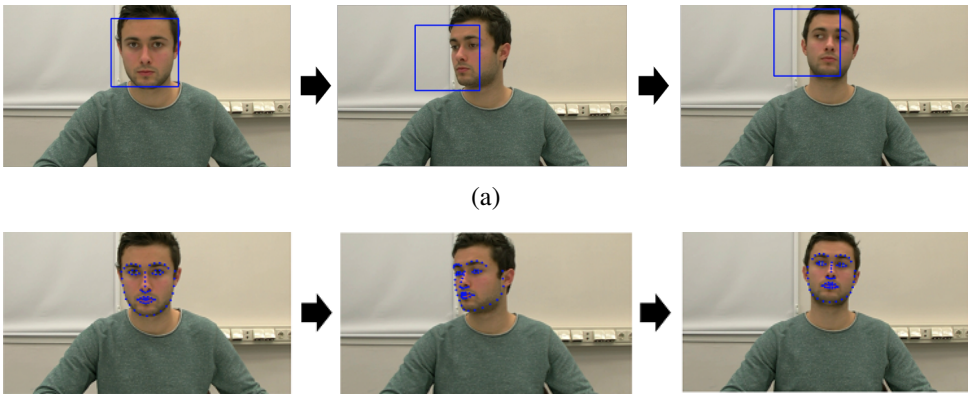
The two techniques tested were a combined use of the Viola Jones algorithm (for face detection) and subsequent tracking by Median Flow algorithm and the sole use of a facial landmark detection library for every frame. Because there is no ground truth for facial position included in the dataset, and annotation of the face's position for every frame was impractical, performance comparison for these two methods had to be done on the basis of domain knowledge and observation.

As mentioned in Chapter 4, only three of the videos included in the dataset addressed head motion, which makes these the most adequate to use for this module's evaluation. In all other videos, the participant's had their head supported by a head rest, which ruled out any type of rigid motion. For that reason, these videos could only be used for facial detection performance evaluation and not for evaluation of the precision in tracking itself. For all participants (and thus skin complexions) as well as for all lighting conditions both methods presented no failures and were entirely accurate in detecting the face for all frames of every video.

However, when it comes to videos P1M1, P1M2 and P1M3 (those which include head motion) considerable discrepancies were found between the performance of the two techniques. Since the Viola Jones algorithm is only capable of detecting frontal faces, there is an imminent drawback to the use of this technique. This drawback means that the algorithm will only work if the camera is positioned strictly in front of the subjects face, at least at the initial time of recording, when face detection is performed. Although this factor may not influence the algorithms' overall performance for this dataset, since all videos start with the subject facing the camera, it will most certainly pose an impediment when dealing with newborns inside an incubator, as one cannot restrain the neonate's head position.

The mentioned limitation, combined with the fact that Median Flow relies on the bounding box from one to obtain the bounding box's position in the following frame, makes this technique extremely fragile for dealing with head motion. As can be seen in Figure 5.8, once the head is turned sideways, this technique fails to accurately detect the face's position. Features from inside the bounding box are then tracked between consecutive frames, meaning that if the bounding box of one frame is slightly deflected from its supposed position, it will include other structures other than the face, possibly background. As a consequence, the tracking portion of the algorithm will not only attempt to track the face but also the other structures contained in the bounding box, degrading its content and propagating the error which had been introduced. This error will then directly affect all forthcoming frames, even those where the face is again facing the camera, as made evident by Figure 5.8 (a).

Landmark detection was therefore considered superior to the combination of Viola-Jones and Median Flow, since it can cope with a much higher angle of head rotation and at the same time



(a)

(b)

Figure 5.8: Comparison of Tracking Methods' accuracy when dealing with Head Motion: (a) The use of Viola Jones for *Face Detection* and *Median Flow* for subsequent tracking may propagate error if the bounding box happens to be shifted; (b) When using *Landmark Detection* error propagation is impossible since keypoint detection is performed for each frame independently.

has no possibility of accumulating error, shown in Figure 5.8 (b), since the fiducial points are detected for each frame independently. Moreover, landmark detection provides much more precise and complete information on the position of the facial structures, which will enable a much more flexible ROI definition. For all those reasons, Landmark Detection replaced Viola-Jones and Median Flow in the *Face Detection and Tracking* modules of the *Baseline Configuration*, creating configuration C1, whose composition can be consulted in Table 5.1.

5.2.2 ROI selection

To address the impact of using distinct facial structures as ROI we compare the performance of the previously created configuration C1 (which uses the forehead as the ROI) with a new configuration C2, which only differs from the former by having ROI defined around the cheeks region. Since the most immediate goal of this module is to provide the best conditions to extract the signal, its performance should reflect more directly on the quality of the extracted signal rather than the accuracy of the final HR extraction. For that reason, a metric which reflects the extracted signal's quality needs to be introduced, that metric being Signal-to-Noise Ratio (SNR).

SNR compares the level of a desired signal to the level of background noise it contains and can be calculated by the ratio of the energy around the desired frequencies and the remaining energy contained in the spectrum, which reflects noise. This metric was calculated according to the following equation:

$$SNR = 10\log_{10}\left(\frac{\sum((U(f)\times\hat{S}(f))^2}{\sum((1-U(f)\times\hat{S}(f))^2}\right) \tag{5.9}$$

where $S(f)$ is the spectrum of the rPPG signal, f is the frequency in beats per minute, $U(f)$ is a binary template window and the value is presented in dB. When analysing the SNR for a given

signal, negative values will mean the energy of the noise frequencies is higher than that of the desired frequencies.

The choice of which frequency is considered as desired, and therefore goes in the numerator, is usually defined by the most predominant frequency (i.e. the peak in the power spectrum). However, in the context of this problem, the frequency regarded as desired should be defined by the HR extracted from the ECG as this represents the ground truth. This is done to prevent highly periodic signals of random frequencies to be regarded as having high SNR.

For that reason, the template window $U(f)$ is multiplied by the square of the normalized frequency response of the signal. This window's values are set to 1 for the chosen frequency and the 5 bins closest to it as well as that frequency's first harmonic and the 10 bins closest to it, as shown in Figure 5.9. The fact that 5 and 10 bins were used instead of exactly the fundamental frequency and its first harmonic was to not consider HRV as noise, once this is a healthy phenomenon which should not be disregarded, as mentioned before. Besides that reason, by using multiple bins around the ground truth frequency we assure, rPPG signals whose predominant frequency is slightly deviated from its supposed value are not regarded as poor quality signals, distinguishing them from completely arbitrary rPPG signals.

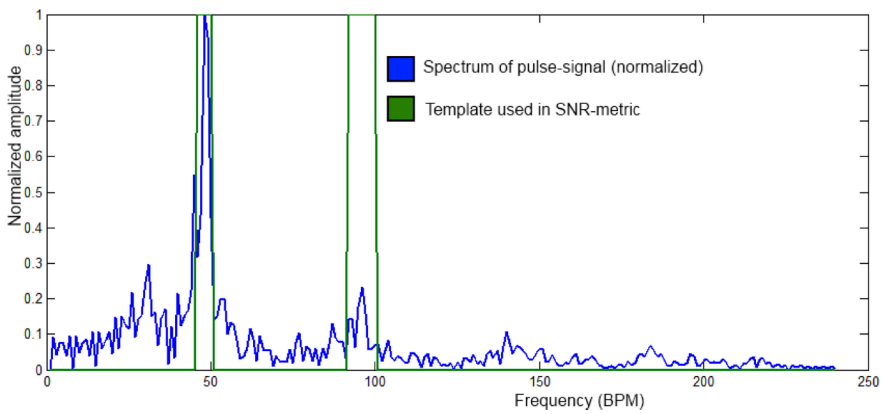


Figure 5.9: SNR calculation uses a template passing 5 bins in the 512 bin spectrum, centered around the contact sensor pulse rate (10 bins around the first harmonic) to allow for heart-rate variability. The SNR is measured by the energy ratio of the components inside and outside the template. from [74]

Table 5.2 displays the average SNR of all 30 second samples extracted using configurations C1 and C2. The average SNR is presented by video to provide a more detailed view on the influence of each ROI in each of the conditions, allowing to assess the performance of these configurations for all three challenges.

After analysing Table 5.2, it is evident that, regardless of the chosen region, participants with darker skin tones consistently present lower SNR. This is a phenomenon which is not exclusive to rPPG and can also be encountered when dealing with PPG signals from pulse oxymeters for an instance, once high degrees of melanin may conceal BVP [91] [92].

Table 5.2: Effect of ROI selection on the rPPG signal SNR

	C1	C2
P1H1	0.511679	-3.042
P1LC1	-2.08301	-1.5578
P1LC2	1.469669	0.006101
P1LC3	4.672141	1.739569
P1LC4	6.573577	3.194916
P1LC5	6.412619	3.396959
P1LC6	0.912877	-1.84869
P1LC7	2.392935	2.974369
P1M1	-0.66871	-0.76924
P1M2	4.120398	-1.25695
P1M3	-7.97976	-1.57556
P2LC1	-8.93493	-4.49823
P2LC2	-7.14645	-5.29417
P2LC3	-7.07065	-4.77595
P2LC4	-6.71044	-4.66715
P2LC5	-6.97931	-5.26607
P3LC1	-5.84706	-5.64595
P3LC2	-7.45471	-5.32919
P3LC3	-5.67233	-5.32685
P3LC4	-6.11048	-5.89743
P3LC5	-6.96168	-4.99319

Although this discrepancy in performance for different skin complexions will be present throughout the analysis of the results, it becomes clear that the use of cheek region as ROI helps to attenuate this difference, as it has distinct impact in the performance for different types of videos. In a general manner, it can be stated that the extraction of the rPPG signal from the cheek area harms the quality of all signals inherent to participant 1, who has a lighter skin complexion, while improving the quality for the other two patients (darker skin tones). This improvement can be justified by the fact that the skin of the cheeks is less thick and more irrigated than the skin of the forehead, thus making BVP more evident for participants who have darker skin tones. Inherently, more evident BVP results in a considerably easier rPPG signal extraction for these subjects, thus shortening the difference in performance for subjects with lighter and darker complexions.

The decrease in performance for the lighter skin tone participant may be justified by the increasing ROI heterogeneity inherent to the use of the cheeks region's. As can be seen in Figure 5.10, the average variance of the pixels inside the ROI is considerably lower for the first participant (P1) when using the forehead in comparison to using the cheeks as ROI. The same is true for the other two participants, as shown in the Appendix A.

In all possibility, by combining the use of the cheek area with a heterogeneity reduction process, one would obtain better results than what could be achieved by using the forehead as a whole. For this reason, C2 is considered to be the configuration with the most potential, in the context of

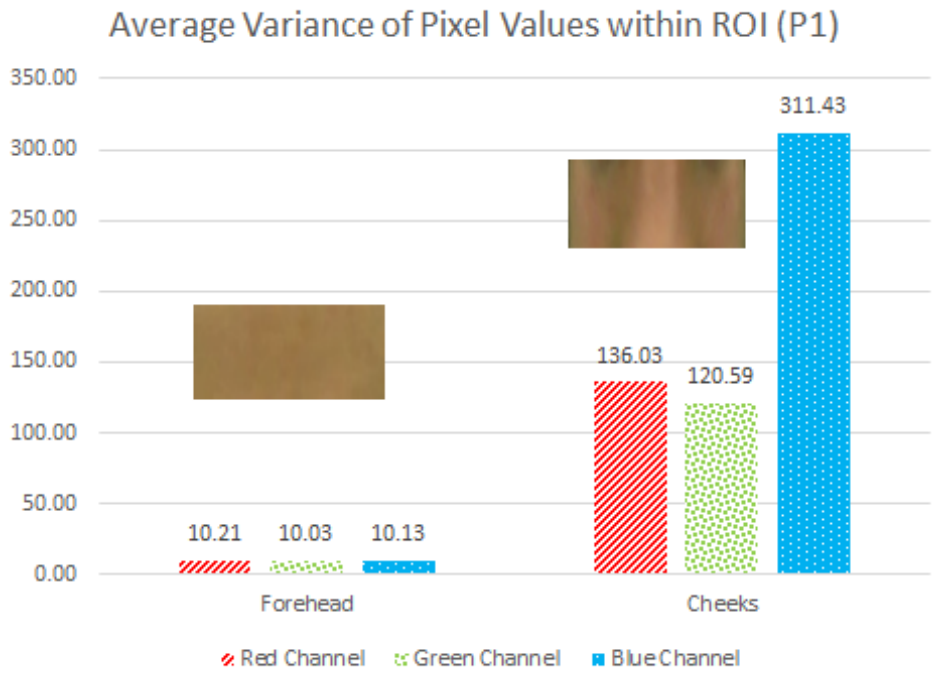


Figure 5.10: Comparison of Tracking Methods' accuracy when dealing with Head Motion.

this problem, once it is pretended from the framework to behave similarly for subjects with different skin tones. It is nonetheless evident that this configuration would benefit from an heterogeneity reduction method.

5.2.3 Sensor size

As mentioned previously, the use of the cheeks region gives rise to an inconvenience, which is the lack of homogeneity inside the ROI. As made evident by Figures 5.4 and 5.10, the average variance within this region is considerate and hence, signal generation by pixel value averaging will be damaged if the ROI is considered as a whole. In order to counteract this effect, the ROI was segmented into smaller, near-square sensors. The consequence of considering smaller sensors independently, rather than utilizing the ROI as a whole is that each sensor will map smaller and more homogeneous regions of the skin. By having less color variance, and being more invulnerable to shadows, movements or lighting variations, these smaller sensors should originate a signal with more quality.

Figure 5.12 proves that dividing the ROI in increasingly smaller sensors decreases the average variance per sensor, as expected, and thus improves the overall sensor homogeneity. However, analysing the frames independently is not enough to guarantee the superiority of the multi-sensor approach, once it only addresses the spatial aspect of the matter. To complement this static analysis, it is essential to consider how using multiple and independent sensors affects the quality of the signal extracted over time.

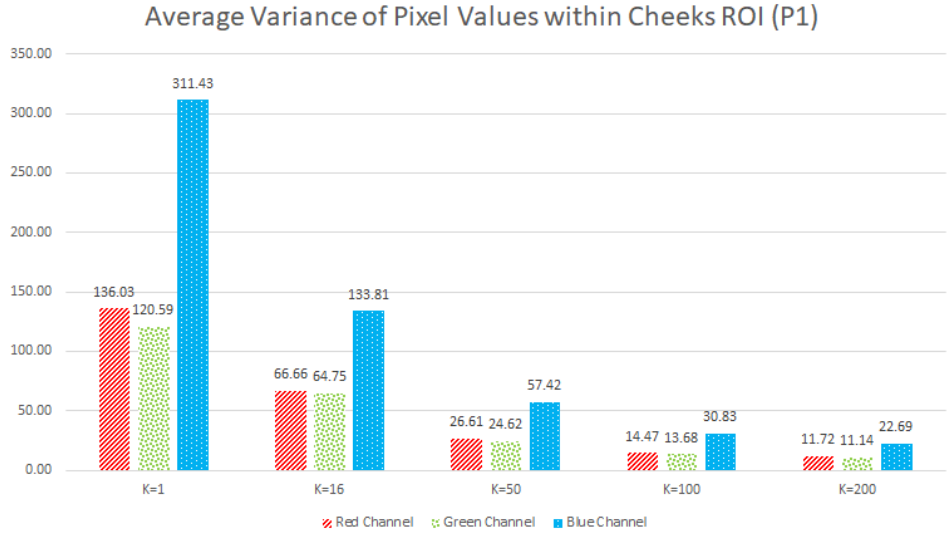


Figure 5.11

Figure 5.12: Comparison of Tracking Methods' accuracy when dealing with Head Motion.

For that reason, the average SNR was once again calculated, to compare the influence of using different sized sensors on the quality of the extracted rPPG signal. As explained in Section 5.1, when using multiple sensors, not all contribute to the calculation of the final HR value. Although the rPPG signal is extracted and their inherent HR frequency computed for each and every sensor, only some of them are selected to define the final HR measure. As proposed by *Niu et al.* [93], the sensors are sorted by their HR value and the l middle HR values are averaged, as they should pose as the most stable. That being said, for every sample, the SNR was calculated as the average SNR of the l sensors which actively contributed to define the final HR value.

Different quantities (and hence sizes) of sensors were compared. To do so, configuration C2, which makes use of the whole ROI as one unique sensor, was compared to other configurations (C3, C4, C5, C6) which were in no way different than the former apart from relying on an increasing amount of sensors: $K = 9$, $K = 16$, $K = 50$, $K = 100$, respectively. Table 5.3

The analysis of the presented results conveys that in a general manner, dividing the ROI in smaller independent sensors enhances the average signal quality, to a certain extent. As portrayed in Table 5.3, the use of nine sensors culminates in a significant increase in the SNR for all videos, with little exceptions. However, as the number of sensors is further extended and their size diminishes, the average quality of the signal becomes progressively worse for 16, 50 and 100 sensors. This can be explained by the fact that for smaller sensors there is a greater influence of the noise induced by the camera's sensor. If the number of pixels to be averaged for rPPG signal extraction is small, the presence of a few pixels whose value is influenced by noise will greatly disturb the average. If the number of pixels were to be higher, the presence of a few pixels affected by noise would be diluted in the totality of values to be averaged. Furthermore, when using smaller sensors, slight head motion may result in a displacement of the same skin regions between adjacent sen-

Table 5.3: Effect of Sensor size on the rPPG signal's SNR

	C2	C3	C4	C5	C6
P1LC1	-1.79	3.69	0.49	-0.42	-0.76
P1LC2	0.18	4.61	1.74	1.08	0.76
P1LC3	1.89	6.00	2.63	2.98	1.90
P1LC4	3.33	5.75	2.09	2.88	1.85
P1LC5	3.49	6.27	3.00	3.15	2.11
P1LC6	-1.82	3.29	0.45	0.41	-0.06
P1LC7	3.07	4.65	0.33	-0.45	-0.35
P2LC1	-5.13	-7.32	-7.63	-7.98	-8.24
P2LC2	-5.25	-4.51	-6.49	-6.89	-7.41
P2LC3	-7.48	-5.25	-6.34	-6.99	-7.39
P2LC4	-7.02	-4.48	-6.40	-6.83	-6.61
P2LC5	-5.43	-4.51	-6.61	-7.17	-7.33
P3LC1	-6.96	-4.40	-4.28	-4.30	-4.12
P3LC2	-5.45	-2.34	-2.13	-2.57	-2.51
P3LC3	-3.51	-2.22	-3.20	-3.11	-3.63
P3LC4	1.07	-1.81	-2.26	-3.04	-3.19
P3LC5	-6.48	-2.53	-2.98	-3.43	-3.58
P1M1	-9.08	-1.22	-0.80	-0.25	-0.51
P1M2	-5.54	14.32	14.46	14.44	14.34
P1M3	-5.99	-12.53	-12.51	-12.20	-12.01
P1H1	-2.92	0.38	-3.58	-2.86	-3.82

sors. This does not occur with the usage of larger sensors, as casual flickering of the landmarks' positions is not enough to make the skin region mapped in one sensor to move and disengage its assigned sensor.

This circumstance suggests that there is an optimal value for the number and size of the sensors. The ideal sensor quantity emerges from a trade-off between sensor homogeneity and distortion induced by sensor noise and subject movement. For that reason, this balance depends on a variety of factors such as the camera's properties, its distance to the subject, the intensity of the subject movements, among others. As a result, the ideal number of sensors, which was found to be 9 for this dataset, may not be the same for other datasets and this parameter should therefore be adjusted when dealing with other environments such as the NICU's.

5.2.4 Signal Extraction

All the modules evaluated so far were optimized in order to facilitate the signal extraction process and maximize the quality of the extracted signal. Nonetheless, the signal itself and its extraction process represent a very important step in the determination of the HR. Distinct signals present unique properties and are therefore expected to behave contrastingly in divergent conditions, which will reflect different performances for the different challenges presented. As mentioned in Section 5.1, the three types of signals tested are the most commonly presented in the

literature: signal from the Green-Channel, BSS signal and Chrominance signal. As stated before, to the use of a BSS signal is associated a supplementary step, which is, for every sample, the choice of the component which best represents the BVP. In order to perform a fair comparison between the three signals, it is first mandatory to determine which method is the best to perform the said component selection.

5.2.4.1 ICA Component Selection

In order to independently test the several methods for ICA component selection, 300 random samples were stored from videos P1LC5, P2LC5, P3LC5, P1LC6, P1H1 and homemade videos which are not part of the used dataset. For each sample, ICA was applied and the 3 resulting components were once again stored as well as annotations of which component corresponded to the rPPG signal.

From the 300 samples used, 240 were randomly selected to train the classifiers (KNN and SVM) and the remaining 60 were used to test the four component selection methods described in Section 5.1.

Tables 5.4, 5.5, 5.7 and 5.6 show the confusion matrices of the prediction on the 60 test samples previously generated, performed by the SVM classifier, KNN classifier, highest FFT peak and correlation with the reference sine, respectively.

Table 5.4: Confusion Matrix SVM

		True		
		Comp. 1	Comp. 2	Comp. 3
Pred.	Comp 1.	16	0	0
	Comp. 2	2	24	0
	Comp. 3	2	0	16

Table 5.5: Confusion Matrix KNN

		True		
		Comp. 1	Comp. 2	Comp. 3
Pred.	Comp 1.	16	0	2
	Comp. 2	2	24	0
	Comp. 3	2	0	14

Table 5.6: Confusion Matrix PEAK

		True		
		Comp. 1	Comp. 2	Comp. 3
Pred.	Comp 1.	13	2	2
	Comp. 2	4	21	0
	Comp. 3	3	1	14

Table 5.7: Confusion Matrix SINE

		True		
		Comp. 1	Comp. 2	Comp. 3
Pred.	Comp 1.	13	2	2
	Comp. 2	3	21	1
	Comp. 3	4	1	13

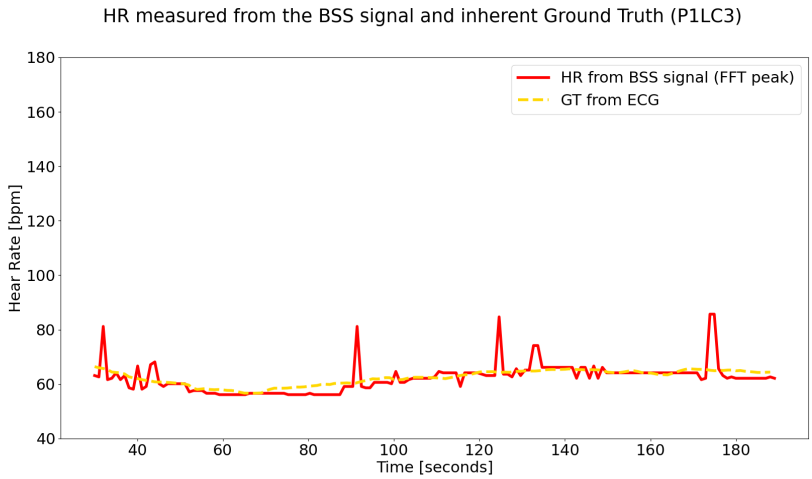
Although this poses as a multi-class classification problem, the three target classes have no particular and distinct meaning once the order of the components given by the ICA algorithm is arbitrary. In other words, there are no specific characteristics that make a signal belonging to a certain class, and therefore analysing each class separately has no significance in the context of this problem. That being said, the problem can be converted to a binary classification problem, in which it is only known if the classification method hits or misses the correct component. For that reason, the global accuracy was calculated for each of the four methods from its respective confusion matrix. From this calculation results a accuracy of 0.93 for the SVM classifier; 0.90 for the the KNN classifier; 0.80 for the sine correlation approach and 0.78 for the traditional highest FFT peak.

Figure 5.13 exhibits how poor component selection influences the final results. In particular, this figure shows the use of configuration C8 for HR extraction in video P1LC3. For the extraction presented on the top figure, the FFT peak method was used to select the rPPG signal from the three ICA components, while for the bottom figure, the method used was SVM. For both graphs, every visible peak in the solid lines, which represent the extracted HR curve, correspond to a failure in selecting the rPPG from the three independent components. In these cases, the chosen component does not reflect the pulse and, thus its predominant frequency is arbitrary, most probably being very different from the supposed value. It is also possible, though unlikely, that the wrongfully chosen component possesses a predominant frequency which is close to that expected of the rPPG signal. In such cases it is impossible to visually identify the failure in component selection. Comparing both figures also displays the differences in using the two mentioned methods and proves SVM superiority for this video, which can be widened for all other videos.

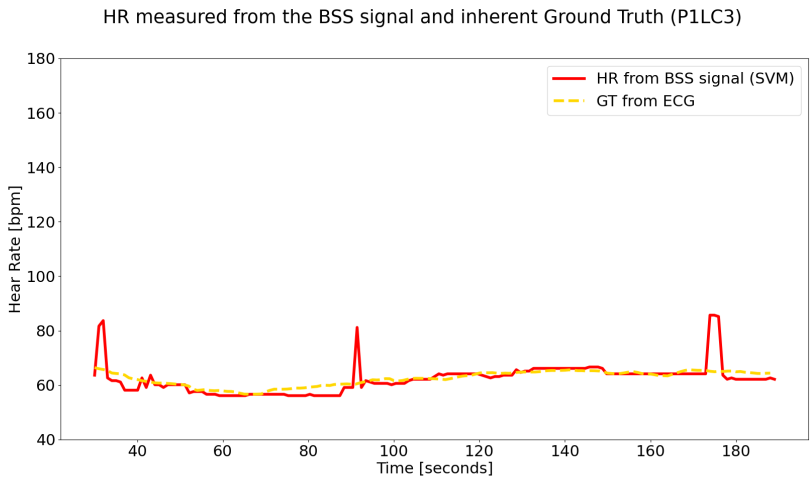
These results match those published by *Mokaresi et al.* [87] and, thus help establish the use of a classifier, and in particular SVMs, as the most accurate method for ICA component selection. However, these conclusions should be considered with caution, as the number of both training and validation samples was scarce. Moreover, the samples used were predominantly extracted from the *Public Benchmark Dataset for Testing rPPG Algorithm Performance*, therefore existing the possibility of overfitting. Once access is granted to a new dataset, preferably of newborn participants, these tests should be performed once again to confirm the results presented.

5.2.4.2 Signal Comparison

Once the best method for component selection in BSS signals is defined, it becomes possible to fairly compare the three different types of signals extracted. Table 5.8 displays once again the



(a)



(b)

Figure 5.13: Comparison best and worst (svm and highest peak, respectively) method for ICA component selection

average SNR for configurations C3, C7 and C8, which are in all aspects identical, except the signal type used.

It would be expected that both BSS and Chrominance signals would manifest an improvement in signal quality, particularly when dealing with non-white illumination (as portrayed in video P1LC6) and darker skin complexions. However, such did not occur and in fact the Green channel derived signal outperformed the other two methods for all conditions imposed by this dataset, which contradicts most literature on the topic. For that reason, all signals were once again tested once the optimal pipeline configuration was reached. Further testing should be performed on a more extensive dataset, in order to substantiate any signal type choice.

Table 5.8: Effect of different signal extraction methods on the rPPG signal's SNR

	C3 (Green)	C7 (BSS)	C8 (Chrom.)
P1LC1	3.69	2.24	-0.08
P1LC2	4.61	3.83	2.30
P1LC3	6.00	5.18	3.69
P1LC4	5.75	5.96	4.88
P1LC5	6.27	6.60	4.37
P1LC6	3.29	3.02	2.93
P1LC7	4.65	4.72	3.78
P2LC1	-7.32	-7.74	-8.25
P2LC2	-4.51	-5.64	-5.84
P2LC3	-5.25	-6.04	-6.43
P2LC4	-4.48	-5.99	-5.56
P2LC5	-4.51	-5.92	-5.73
P3LC1	-4.40	-4.18	-3.71
P3LC2	-2.34	-2.51	-2.34
P3LC3	-2.22	-1.67	-1.14
P3LC4	-1.81	-2.57	-2.45
P3LC5	-2.53	-2.60	-2.11
P1M1	-1.22	-1.25	-0.62
P1M2	14.32	12.09	4.11
P1M3	-12.53	-10.43	-1.74
P1H1	0.38	-1.47	-0.95

5.2.5 HR extraction and Post-Processing

Once all modules relative to signal extraction were evaluated and their best techniques set, the quality of the extracted rPPG signal can be considered as close to maximized as possible, for all the conditions tested. It then becomes essential to use the extracted signal to obtain the continuous HR - the final output of this portion of the framework.

As mentioned in Section 5.1, two distinct methods were employed and compared. Similarly to what has been performed for all other modules' evaluations, two identical pipeline configurations (C9 and C10) were established, the only difference between them being that the former used the Spectral Analysis to determine the HR from the rPPG signal while the latter used Peak Analysis. Since these techniques will in no way influence the quality of the rPPG signal, as they will make no alterations to it, new metrics have to be introduced to evaluate the performance of the module in question. These metrics address the final result and its relation with the ground truth values extracted from the ECG and thus, their use allows to analyse the performance of the overall pipeline. The metrics to be used are the *root mean squared error* (RMSE) and the Pearson's product-moment correlation.

The RMSE was calculated for every sample and averaged per video. As can be seen in Table 5.9, the error calculated was lower for all instances when using Peak Analysis rather than

Spectral Analysis. As had already been discussed when analysing previous modules, the higher presence of melanin severely degrades the quality of the extracted rPPG signal, which results in poor HR extraction. This phenomenon justifies the discrepancy presented in the RMSE between participants P2 and P3 and P1.

Table 5.9: RMSE calculated for the continuous HR curves extracted with configurations C9 and C10.

	C9	C10
P1LC1	4.38	1.54
P1LC2	2.44	1.52
P1LC3	2.38	0.75
P1LC4	1.93	0.94
P1LC5	1.21	0.74
P1LC6	15.65	5.58
P1LC7	3.63	2.05
P2LC1	38.69	13.38
P2LC2	12.77	3.98
P2LC3	29.35	6.11
P2LC4	12.99	6.40
P2LC5	10.09	6.80
P3LC1	38.56	8.92
P3LC2	27.05	7.01
P3LC3	22.12	7.76
P3LC4	20.15	5.79
P3LC5	32.00	9.56
P1M1	26.94	17.89
P1M2	3.51	3.49
P1M3	29.39	29.88
P1H1	1.90	1.23

Pearson product-moment correlation was also calculated between the ground truth HR curve and the HR curve extracted with configurations C9 and C10. This metric addresses the linear correlation between two continuous variables dividing the co-variance of the two variables by the product of their standard deviations. It is expressed as a value between 1 and -1, in which 1 means total positive linear correlation, 0 means no linear correlation, and -1 means total negative linear correlation. Table 5.10 compares the Pearson's product-moment correlation for all the videos when using configurations c9 and C10.

After analysing these two metrics, it becomes evident that Peak Analysis outperformed the use of the FFT. Spectral Analysis should in theory be more robust, once it is not disturbed by a moderate presence of spectral noise. In fact, when analysing the power spectrum of the rPPG signal, originated by the FFT, there should be no error in the extracted HR as long as the response of the unwanted frequencies does not surpass that of the desired frequency. On the other hand, as all local maxima (peaks) of the rPPG are used to determine the HR through Peak Analysis, any

Table 5.10: Correlation between the continuous HR curves extracted with configurations C9 and C10 and the ground-truth HR curve.

	C9	C10
P1LC1	0.02	0.77
P1LC2	0.45	0.76
P1LC3	0.80	0.98
P1LC4	0.69	0.93
P1LC5	0.89	0.95
P1LC6	0.49	0.72
P1LC7	0.64	0.81
P2LC1	-0.17	0.20
P2LC2	-0.34	0.24
P2LC3	0.50	0.66
P2LC4	0.21	0.11
P2LC5	0.46	0.40
P3LC1	-0.05	0.00
P3LC2	-0.11	0.34
P3LC3	-0.27	0.07
P3LC4	-0.03	0.41
P3LC5	-0.08	0.16
P1M1	-0.36	0.11
P1M2	-0.09	-0.20
P1M3	0.17	-0.05
P1H1	0.90	0.95

kind of signal noise may result in unwanted peaks to be inserted in between real peaks, severely affecting the interval between the adjacent beats and deteriorate the HR extraction.

However, there are also disadvantages associated to the use of the FFT, the main downside being its limitations in terms of frequency resolution. When using the FFT to compute the signals power spectrum, the number of frequency bins, and consequently the frequency resolution is defined as half of the number of instances which constitute the given signal. Once the sampling frequency of the extracted rPPG signal is low, in order to achieve a reasonable frequency resolution it is mandatory to use a signal with big enough length. However, increasing the length of each sample's signal comes with the cost of losing HR measures, as the entire length of the video has to be divided in bigger portions. This downside inherent to the FFT and the fact that all previous modules were employed so as to reduce the amount of noise in the extracted rPPG signal, justifies the superiority in accuracy of using Peak Analysis for HR extraction.

The last module to be analysed is the *Post-Processing* Module. Regarding this module, the tests performed were to compare the presence or absence of the technique above described in more detail. As can be derived from the analysis of Table 5.11, the use of the post-processing technique did not consistently improve the results for all videos. However, for those which were of most interest, meaning those which represent challenges which will likely be faced in a NICU environment such as P1LC6, P1LC7 and P1H1, this technique proved to be efficient in reducing

the RMSE.

Table 5.11: RMSE calculated for the continuous HR curves extracted with configurations C10 and C11.

	C10	C11
P1LC1	1.54	1.67
P1LC2	1.52	0.90
P1LC3	0.75	0.87
P1LC4	0.94	0.83
P1LC5	0.74	0.68
P1LC6	5.58	3.89
P1LC7	2.05	1.55
P2LC1	13.38	11.81
P2LC2	3.98	3.11
P2LC3	6.11	3.80
P2LC4	6.40	2.60
P2LC5	6.80	2.56
P3LC1	8.92	18.16
P3LC2	7.01	5.98
P3LC3	7.76	5.16
P3LC4	5.79	2.99
P3LC5	9.56	5.64
P1M1	17.89	21.83
P1M2	3.49	3.51
P1M3	29.88	29.87
P1H1	1.23	1.05

5.3 Overall Performance

With the evaluation of the techniques for the last two modules finished, the configuration which proved to be best, having in mind the challenges to be faced when dealing with newborn participants was C11. As expected, this pipeline's performance was not equal for all videos as each poses as a different challenge to test the tools accuracy. As has been noted in all modules' evaluations, the accuracy varies significantly across participants given that they have different skin tones. Although the accuracy of the framework for subjects with darker complexions has been lower than the accuracy for the participant with lighter complexion, it can be stated that the tool still managed to extract the continuous HR for these subjects with mild success. In fact, if the Lighting condition with less light intensity is discarded, the framework managed to always obtain RMSE below 4 bpm for participant P2 and below 6 bpm for participant P3. It is expected that, for neonatal participants the tool's performance resembles more that of participant P1 once newborn's have thinner skin and with less concentration of melanin for the first few weeks of life.

When it comes to *Challenge 3* proposed by the dataset, the tool did not present acceptable results which leads to the conclusion that it is not yet ready to deal with rhythmic and exaggerated

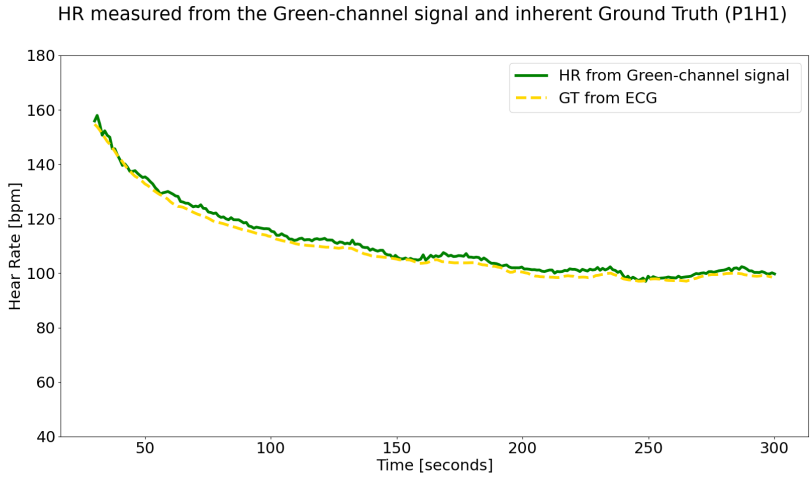
movement such as what is presented in videos P1M1, P1M2, P1M3. While the RMSE values for videos P1M1 and P1M3 clearly indicate the lack of robustness of the framework to deal with head motion, the RMSE calculated for video P1M2 may indicate otherwise. The reason why this particular value was so low is that the subject was nodding his head at rhythm of 60 bpms, which is extremely close to the 64 bpm of average HR which can be seen from the ground truth inherent to that video.

Regarding the challenging conditions which matter the most in the context of this problem, namely high and fluctuating heart rates (P1H1) and uneven lighting conditions (P1LC7) the results were convenient and support that the tool developed should have little or no difficulties when dealing with these specific challenges in a NICU environment. As can be seen in Figure 5.14, the extracted HR curve closely matches the continuous HR curve extracted from the ECG, being capable of detecting the slightest changes in HR even for uneven lighting conditions. Equivalent plots can be found for every video in the dataset in Appendix A.

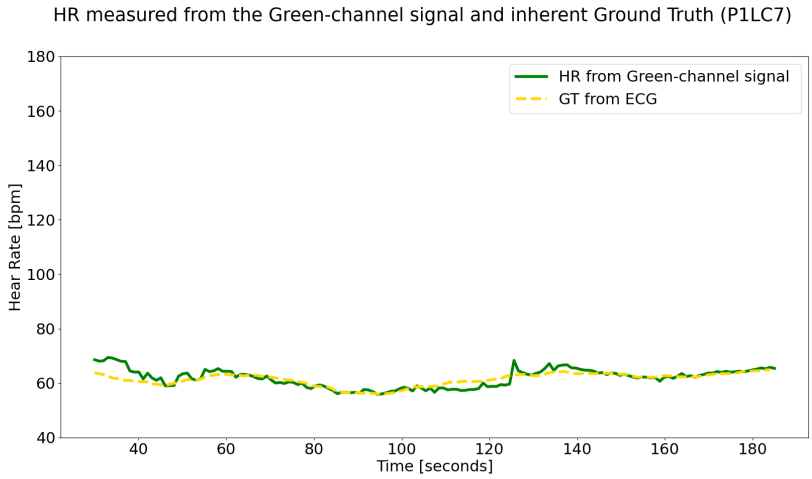
Bland-Altman analysis was also performed to assess the agreement between the HR extracted from the rPPG signal and the ECG signal. Once the mean of the distributions close to zero and few samples fall outside the range $[mean - 1.96\sigma, mean + 1.96\sigma]$ it can be concluded that the two methods for continuous HR extraction are correlated. Once again, Bland-Altman plots for every video of the dataset can be found in Appendix A.

5.4 Summary

Considering the results explored above, the final and optimal configuration for contactless HR extraction WAS C11, whose description can be seen in Table 5.1. The tests performed in the *Public Benchmark Dataset for Testing rPPG Algorithm Performance* created at the *Eindhoven University of Technology* [25] led to conclude that the developed framework is capable of dealing with high and fluctuating heart rates as well as adverse lighting conditions. This should function as proof of concept that the developed tool would perform accordingly for neonatal subjects. However, extensive tests should be performed with newborn subjects and under real-world conditions to validate this assumption.

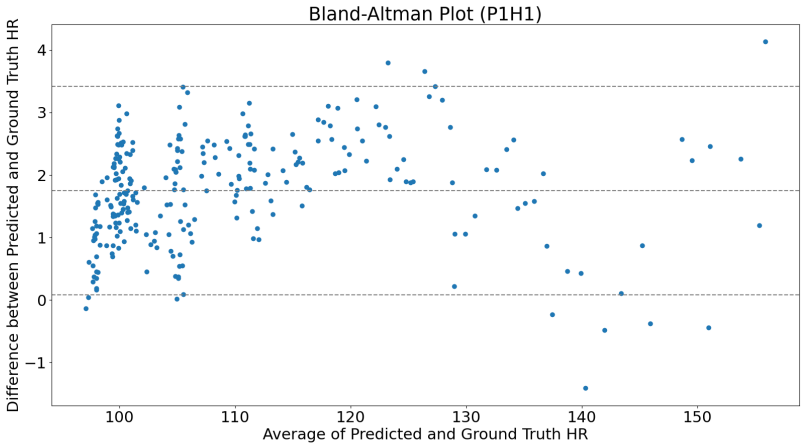


(a)

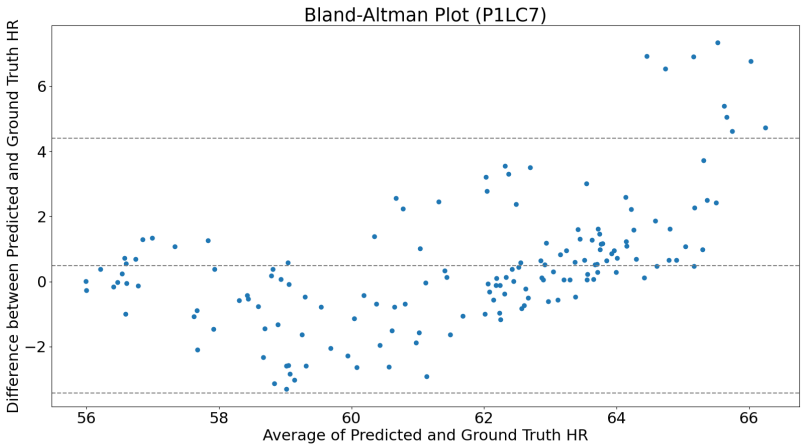


(b)

Figure 5.14: Continuous HR curve extracted using configuration C11 from videos (a) P1H1 and (b) P1LC7.



(a)



(b)

Figure 5.15: Bland-Altman plots for the extracted HR curve of videos: (a) P1H1; (b) P1LC7.

Chapter 6

Eulerian Video Magnification

Monitoring of vital signs in newborns, regardless of being contactless or not, is unquestionably valuable for clinical motives, in the sense that it allows healthcare professionals to assess the neonate's well-being and in case something is wrong, to quickly determine the cause. However, there is more to newborn monitoring other than its clinical purpose. Parents enjoy being able to often check-in on their newborn children. Since this is many times inconvenient for the parents, who have to walk to the infant's room, most buy baby-monitors. These are devices which consist of two components: a camera which is placed at the site of the newborn's crib and other which the parents carry to be able to watch their children regardless of their position. This is where video magnification comes in handy. When sleeping, infants are static most of the time, which may worry the parents when looking at the baby-monitor. For that reason, by magnifying either the infant's skin color variation due to the beating of the heart or chest movements due to breathing, the process of monitoring would also be made easier for the parents.

6.1 Methodology

In a renowned paper published in 2012, Wu et. al. [67] described a methodology they named Eulerian Video Magnification (EVM), which intends to reveal and display color or motion variations in videos over time, which are invisible to the naked eye. This technique has grown in popularity since its publishing date, as it can be integrated in a vast range of applications. EVM has been showcased as capable of amplifying crane movements caused by the wind, or to highlight the mechanical movements in a DSLR camera when auto-focusing, but most importantly to amplify color changes in skin caused by differences in blood volume on the most superficial capillaries, or magnifying chest movements due to breathing in videos of babies.

Figure 6.1 illustrates the general process through which magnification is achieved. Firstly the video sequence is decomposed into different spatial frequency bands. In order to do this, a full Laplacian pyramid is computed [94].

Temporal processing can then be applied. For this purpose, each pixel of each spatial frequency band is considered and its value extracted over time to create a 1D signal. A bandpass

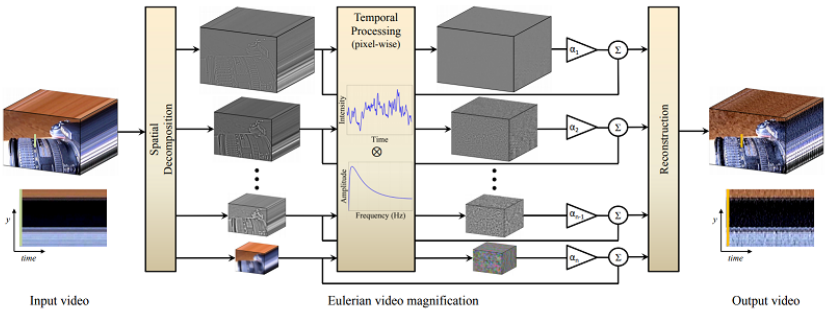


Figure 6.1: Overview of the Eulerian Video Magnification framework, extracted from [67].

filter is then applied to each pixel's signal in order to retain the frequency band of interest exclusively. The cut-off frequencies of this filter are set by the user according to the signal which they desire to magnify. It is important to disclose that the bandpass filter is the same for all spatial frequency bands and all of its pixels. All resulting signals are then multiplied by a magnifying factor α, which is given by the user. The magnified signal is then combined with the original and all spacial frequency bands (i.e. Pyramid levels) are collapsed resulting in the magnified video sequence. According to Wu et. al. [67], both increasing magnification factors and motion ($\delta(t)$) may introduce noise in the resulting video sequence. Therefore, it is beneficial to define different amplification factors for each spatial frequency band. For this purpose, the amplification factor is fixed to α for spatial bands that are within a derived bound, derived from the band's spatial frequency.

$$(1+\alpha)\times\delta(t) < \frac{\lambda}{8} \tag{6.1}$$

where λ is defined as $2\pi/w$. For higher spatial frequencies α is linearly decreased, or forced to zero, hence reducing any distortions.

When it comes to magnification of color induced by heart rate, it is intended to emphasize color changes in low spatial frequency bands, once the human skin is considerably homogeneous and therefore represented with more intensity in the low frequency bands. By applying a constant magnification factor, one would retain motion artifacts derived from pixel intensity changes caused by subtle movements of the subject's head. Although the participants of our dataset had their head supported by a rest, which should prevent rigid head motions. Nevertheless, nonrigid motions are still possible. To overcome this problem, for this type of applications, the magnifying factor α may be forced to 0 for spatial frequencies above a threshold (as explained in Equation 6.1). For motion magnification videos on the other hand, it would be advantageous to use a linear ramp transition for α.

6.2 Results and Discussion

Figure A.40 exhibits the effect of applying EVM to a patch of the P1LC5 video of the dataset in order to amplify color changes derived from pulse. The parameters were chosen heuristically and settled at $\alpha = 300$ and number of Laplacian Pyramid's levels equal to 6. As described by Wu et. al. [67], optimal results are achieved with more complete Pyramids, as the deepest levels enhance the change in regions with low spatial frequency, such as the skin. The temporal filter cutoff frequencies were purposely set to 0.7 and 3Hz, identically to what had been established for the rPPG tool. Although one would get better results with a narrower bandpass filter, such filter would be impossible to create without previous knowledge of the subject's heart rate. In Figure A.40 (a), four frames from the original video are displayed in the top row and the same four frames appear in the bottom row after magnification. The chosen frames are equally spaced and the temporal distance in between them was set so that consecutive images alternate between local maxima (second and fourth images) and and local minima (first and third) of the rPPG signal derived from the green channel alone. This was done to emphasize the difference made evident by EVM. Although all four images seem identical in the top row, after magnification, the difference in hue of light reflected by the skin is obvious. In Figure A.40 (b) one can see a vertical scan line extracted from the videos, which helps to better visualize the difference in color over time. Once again, the top image is relative to the original video and the bottom to its amplified version.

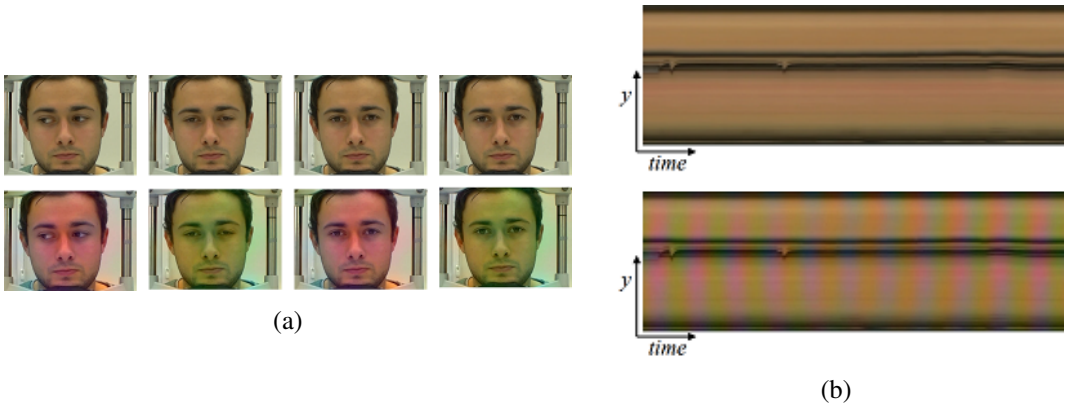


Figure 6.2: Example of the results of applying Eulerian Video Magnification to one of the dataset's videos. (a) In the top row, four frames from a patch extracted from the original video sequence and in the bottom row the same four frames after amplification. (b) A vertical scan line from the input video (top)and resulting video (bottom).

Although the use of EVM seems the most interesting for monitoring purposes, when it comes to contactless vital signs it would also be interesting to use as a pre-processing step to the heart and respiratory rate extraction algorithms previously developed. In order to test if this tool enhances rPPG signal quality, or even improves the final results at all, a comparison was performed between running the developed HR and RR extraction algorithm on the original and color magnified version of a few select videos.

For this analysis, only one video for each of the 3 participants was used (videos P1LC5, P2LC5, P3LC5). The chosen light condition was LC5. According to the results presented in Chapter 5, this Light Condition was the least challenging for all three skin-tones, in the sense that it was the Light Condition for which the rPPG signal quality was on average higher. The HR extraction configuration selected to test the effectiveness of EVM as a pre-processing step was configuration C11, which proved to be the one which better performs for the entirety of the dataset, as discussed in Chapter 5. This method uses the cheek area as the ROI, not taking advantage of its further segmentation. The rPPG signal used is chrominance-based and does not benefit from post processing of any kind. The filter used to obtain the signal is, nevertheless an adaptive filter.

At a first glance, one can notice that the rPPG signal extracted from the magnified video does not possess a higher SNR in comparison to is non-magnified equivalent. Figure A.40 shows one of the 155 one second samples used to extract the HR from the original P1LC5 video (on the left) and its magnified version (on the right). In this sample, which was chosen randomly, the SNR difference between the two methods is XXX dB. In fact, it is possible to conclude that in general EVM severely damaged the rPPG's signal quality, since the signal extracted from the original P1LC5 video had an average SNR of 5.54 dB and the signal from the magnified version revealed a SNR of -1.52 dB. One would expect that for the darker complexions, where the BVP phenomenon is less evident, EVM would come in handy. However, for both P2LC5 and P3LC5 the SNR dropped severely once again when extracting the rPPG signal from the magnified video. This decrease in signal quality may result from the influence of the first levels of the Laplacian Pyramid, i.e. the high spatial frequency regions. This regions which simultaneously fall inside the ROI, such as nose contours or even skin imperfections or shadow limits, might flicker in intensity or color due to unwanted factors rather than the beat of the heart, most likely due to non-rigid movements, which despite being small are not insignificant in the context of this problem. The signal caused by this variation possesses a dominant frequency which, probably does not correspond to the frequency of the heart beat, hence decreasing SNR. By definition, the SNR is the ratio of the energy of the desired frequency (in this case corresponds to its peak) over the sum of the energies of all other frequencies. As the more predominant the frequency of interest is, the higher the signal's SNR is going to be, the presence other frequencies different from the interest one, will decrease the SNR.

Although the influence of the lower levels of the Laplacian Pyramid could be reduced by previously smoothing the image or lowering the threshold λ, this would most certainly not translate to a sufficient increase in the SNR of the magnified signal that would make it preferable over the signal of the original video. Besides the influence of the lower levels of the Pyramid, there are also artifacts induced by low spatial frequency regions which degrade the signal extracted. These artifacts come in the form of irregular color patterns across the face , which may result from the fact that the blood travels gradually up the face and away from the heart, increasing the blood volume with different intensities for distinct facial regions at a time. These artifacts have particular influence when the ROI is evaluated as a whole, instead of being divided into smaller and less heterogeneous independent sensors. Not surprisingly, the poor signal quality in magnified videos

results in a flawed HR extraction, most certainly potentiated by a defective peak detection, as can bee seen in Figure 6.3. As explained in the previous chapter, RR extraction is a much more fragile problem which is completely dependent on signal quality. For this reason, the decrease in the algorithm's performance is even more significant for this vital sign.

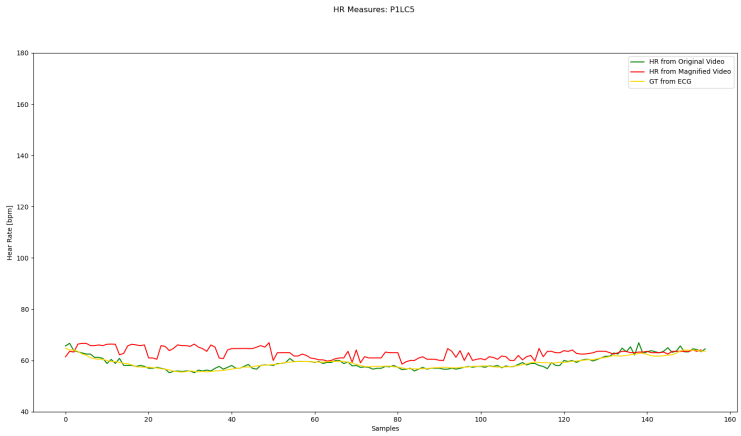


Figure 6.3: Comparison of the continuous HR curve extracted from a magnified video and its original (P1LC5). The solid curve represents the HR curve obtained from the magnified video, while the dashed curve represnts the HR curve from the original video. The ground truth can be seen as the dotted line.

6.3 Summary

To sum up, EVM is undoubtedly a fascinating technique, whose interest relies mostly on monitoring processes and not on accurate vital sign extraction itself, since its use does not improve the results for both HR and RR. Furthermore, the high computation time associated to this technique makes it difficult for it to ever integrate a real-time application, until general hardware capabilities improve enough to accommodate more computationally expensive processes.

Chapter 7

Respiratory Rate

One other goal established for this master's thesis was the extraction of RR from videos. From the method described in Chapter 5, one can extract not only the final HR measures but also the complete rPPG signal as a intermediate product. This signal contains useful information, which will be the basis for the RR extraction.

In order to comprehend the process of extracting the RR from the rPPG signal, one must grasp the concepts of Heart Rate Variability (HRV) and respiratory sinus arrhythmia (RSA). It is well known that any healthy individual's HR is non-stationary and its variability may contain indicators of disease, general well-being or impending cardiac diseases. Heart rate variability (HRV) is defined as the variation over time of the period between consecutive heartbeats, being a reflection of the many physiological factors which modulate the normal rhythm of the heart such as the interplay between the sympathetic and parasympathetic nervous systems [95]. One of the most important causes of HRV is the breathing cycle. RSA is the component of heart rate variability derived from respiration, characterized by shortening of R-R intervals on an ECG during inspiration and prolongation during expiration. Although RSA has been used as an index of cardiac vagal function, it is also a physiologic phenomenon reflecting respiratory-circulatory interactions universally observed among vertebrates [96].

7.1 Methodolgy

Chen et al. [97] developed a method, which exploits the phenomenon of SRA in favor of contactless extraction of the RR. This rPPG-based method was chosen instead of methods which rely on chest movements (described in Chapter 3), since the latter are more influenced by subject movement. Furthermore, it was beneficial to use a rPPG-based method since the rPPG signal had already been extracted for the HR component of this thesis and with reasonable success. The framework of the method which was used is outlined in Figure 7.1.

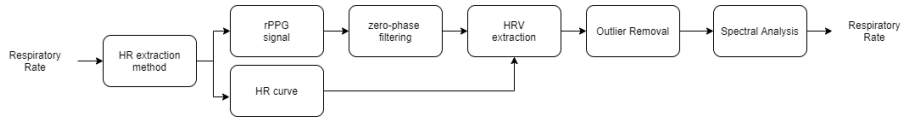


Figure 7.1: Framework used for contactless RR measurement from face videos.

7.1.1 Filtering and Peak Refinement

That being said, the previously extracted rPPG signal, described in detail in Chapter 5, was used as an input for the RR extraction portion of the framework developed. This rPPG signal had already been filtered with a bandpass filter to exclude frequencies outside the possible HR frequency band. However, this signal does not have enough quality for HRV obtainment, as this is a much more challenging task than that of HR's. Identically to what had been described by *Chen et al.* [97], the solution to this problem was the design of a second filter. This is a infinite impulse response (IIR) filter with a much narrower band, whose cut-off frequencies are calculated from the dynamic range of the HR curve previously extracted for the time interval in analysis. The frequency interval (in bpm) of the bandpass was defined as $[HR_min - 30, HR_max + 30]$. The offset value was set to 30, as this was the optimal value considered to preserve HRV and exclude as much noise as possible, once larger offsets would have increased the noise presence and smaller offsets could degrade HRV. Since peak location is key for the success of this task, the previous filter was applied in a zero-phase filtering process. This process consists of filtering the signal forward and then backwards, resulting in no phase distortion, which facilitates HRV extraction. Figure 7.2 displays how the bandwidth of the narrow filter is defined.

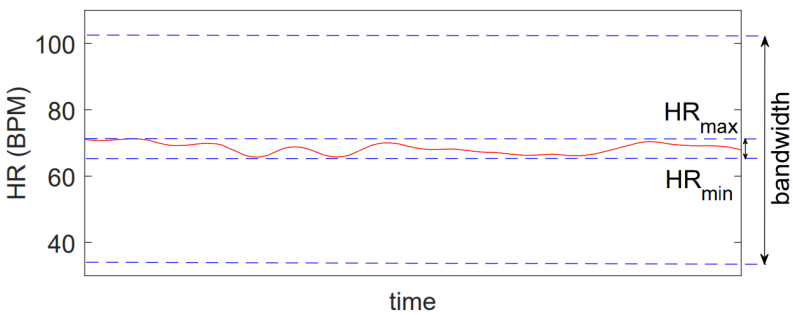


Figure 7.2: Definition of the second bandpass filter's bandwidth, extracted from [97]. The solid curve indicates the continuous heart rate curve. The selected bandwidth of the second filter is wider than the dynamic range of HR curve in the video segment.

7.1.2 HRV Extraction

Peak analysis is then performed on the filtered signal, in order to detect the signals values which correspond to heart beats. Due to the relatively low sampling frequency of the rPPG signal (30Hz), a peak refinement step is added, in a repeated effort to maximize the accuracy of the beat's location. This consists of interpolating the filtered rPPG signal quadratically around the peaks' location. Figure 7.3 stresses the improvement in beat location induced by the peak refinement step.

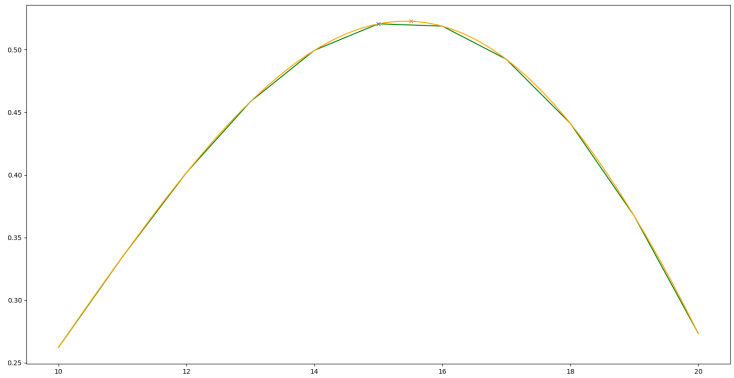


Figure 7.3: Influence of peak refinement in improving the position of the heart beats. The solid curve represents the portion of the rPPG signal and its peak is represented by a cross. The dashed line is the same portion after peak refinement. The new peak location can be identified by the triangle.

Once the refined peaks are determined and their timestamps known, the time (in seconds) in-between consecutive peaks, also known as Inter-beat Interval (IBI), is calculated. The time for each IBI sample is set to the middle time of the interval, resulting in an unevenly sampled signal. The HRV signal (in bpm) is calculated by dividing 60 by each sample, since this measure is simply the reciprocal of the IBI. The HRV signal is then detrended by subtraction of the previously calculated HR curve. This is done to map the variability in relation to the HR, meaning that the closest to zero a sample is, the nearer that sample is to the measured HR. Figure 7.4 shows an example of the filtered rPPG signal with its peaks outlined, as well as the corresponding HRV and detrended HRV.

7.1.3 Outlier Removal and Respiratory Rate Extraction

The detrended HRV will be used in the final step to extract the RR. However, it is important to remember that its samples were derived from an rPPG signal instead of a PPG signal or even an ECG. That being said, the detrended HRV may contain samples which represent outliers as a consequence of sudden subject movement or lighting complications at the time of video acquisition. To overcome this, an outlier removal step is enforced. This process relies on the assumption that

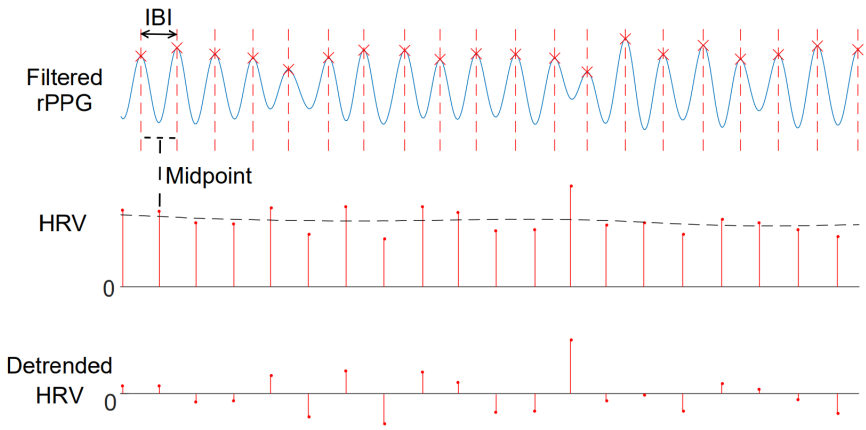


Figure 7.4: HRV and detrended HRV calculated from the filtered rPPG signal, extracted from [97].

the detrended HRV samples follow a Gaussian distribution $N(\mu, \sigma^2)$. The distribution's parameters (μ and σ) are then estimated by maximum likelihood estimation (MLE), and any samples that fall outside the range $[\mu - \alpha\sigma, \mu + \alpha\sigma]$ are considered outliers and discarded. Identically to what was done in *Chen et al.'s* [97] paper, α was set to 3 according to the three-sigma rule [98].

Since the detrended HRV signal is an unevenly sampled signal, it is impossible to perform spectral analysis, by using the standard FFT. For that reason, the *Lomb-Scargle periodogram*, a method based on a least squares fit of sinusoids to the data samples [99], was used. From the spectrum, the RR is extracted as the frequency with the maximal energy response inside the range of 5 to 30 breaths per minute, the normal breathing rate range for human adults. Since the aim of this application is to work for preterm infants rather than adult subjects, once adaptations are made, the selection range has to be shifted to approximately 30 t0 60 breaths per minute. This choice is justified by the fact that the normal breathing rate of children in their first months of life is much higher as that of an adult, as explained in Chapter 1.

7.2 Results and Discussion

As the dataset did not provide any physiological data that directly reflects the respiratory rate, the ground truth had to be extracted from the only valid physiological measure provided: the ECG. *ECG-Derived Respiration* (EDR) is a technique based on the fact that, as the lungs fill and empty and the chest rises and falls, the positions of ECG electrodes on the chest surface move relative to the heart, thus varying trans-thoracic impedance. This results in variations of the mean cardiac electrical axis, that are correlated with respiration [100]. However, given the quality of the ECG included in the dataset's files, in particular the fact that only one lead was provided, the use of this technique to determine the ground truth was impossible. As an alternative, we calculated the RR from the ECG by taking advantage of the SRA phenomenon, similarly to what had been done to extract RR from the rPPG signal. To do so, every R peak (from the ECG's QRS complexes) were detected and the intervals between R peaks (R-R intervals) were calculated. The R-R intervals

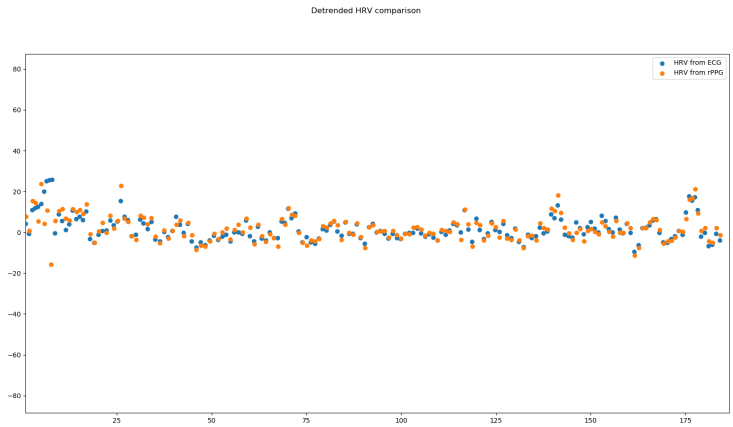
were then used to calculate the HRV signal from which the RR was extracted based on the principal of SRA. In contrast to what was necessary for the rPPG signal, both peak refinement and outlier removal steps were not included, since the sampling frequency is extremely high (1024Hz) and the R peaks detected do not include outliers. This ground truth method will be referred to as *GT1*. As a complement, respiratory rate was also derived from the video by visually counting chest movements. To do so, a sequence of one inhalation and one exhalation was considered as one breath.This ground truth method will be referred to as *GT2*. It is important to recall that neither method used as ground truth compares to standard and clinically accepted methods, which provide a calibrated respiration signal, such as spirometry, measurements from nasal thermistors, and plethysmography.

As a first and more immediate analysis, for each video the detrended HRV signal was analysed and the *Pearson's correlation coefficient* was calculated between the said signal and the detrended HRV extracted from the R-R intervals of the ECG. The most evident finding for this analysis is the discrepancy which can bee seen between videos of corresponding lighting conditions for the patients with lighter and darker skin complexion, as demonstrated in Figure 7.5.

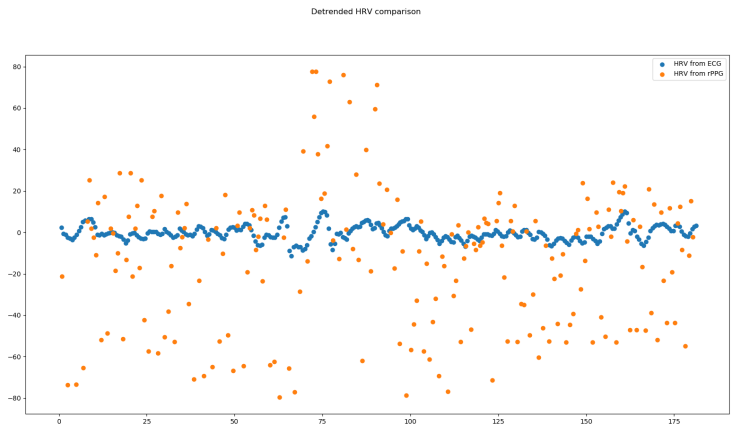
It can be concluded that, when considering videos recorded under the same lighting conditions (LC5 in the case represented in Figure 7.5), the detrended HRV signal extracted for participant P1 is evidently more precise than for participant P2 and presents a much stronger correlation with the detrended HRV from the ECG (Pearson product-moment correlations of 0.85 and 0.0650., respectively). Through this analysis is possible to predict the probable failure of the described algorithm for darker skin complexions, which is once again substantiated by the limited rPPG signal quality associated with higher concentrations of melanin in subjects' skins.

Furthermore, as described by *Chen et al* [97], the inadequate use of narrow bandpass filters can degrade the rPPG signal and unintentionally eliminate the traces of HRV. The pipeline configuration for HR extraction which was chosen as the best (configuration C11, described in table 5.1), makes use of an adaptive filter whose frequency band is centered around the frequency of the previous estimation, allowing for a much narrower filter. Although this step may improve results for HR extraction, it can impair the process of RR extraction. In order to evaluate the influence of of narrower bandpass filters, the RMSE of the final RR results and Pearson's correlation was calculated for all videos using the rPPG signal extracted with configurations C8 (wide bandpass filter) and C11 (narrow bandpass filter). So as to remove the disregard the accuracy of HR extraction portion of the framework, the HRV signal was also detrended by subtracting the HR curve from the ECG, instead of the obtained HR curve. This will provide information about the exclusive influence of the quality of the signal and the RR extraction methodology alone, without being biased by the performance of the portion of the framework described in Chapter 5. The three videos inherent to *Challenge 2* of the dataset (P1M1, P1M2 and P1M3) were not addressed as the rPPG signal and consequent extracted HR curve were found to not have sufficient quality to enable RR extraction.

Table 7.1 shows the differences in RMSE for the final RR extracted from rPPG signals filtered with both types of filters. This metric was calculated for every video between the extracted RR



(a)



(b)

Figure 7.5: Overlapped detrended HRV from rPGG and ECG signals of participants with distinct skin complexions: (a) From video P1LC5; (b)From video P1LC5.

and the ground truth obtained from both methods. Since it is impossible to obtain the detrended HRV from the ground truth method of visually counting the breaths, Table 7.2 only exhibits the Pearson's correlation coefficient calculated betwen the detrended HRV calculated from the rPPG and ECG signals.

By analysing the presented results it can be confirmed that the algorithm developed did not reliably extract the RR for participants with darker skin tones. Furthermore, a increase in RMSE and decrease in Pearson's correlation with progressively worse lighting conditions suggests that the accuracy of the developed tool is strongly influenced by the quality of the rPPG signal extracted as worse lighting conditions, such as LC1 and LC2 consistently induced both low SNR and correlation between the calculated and true detrended HRV signals. Both metrics were found acceptable for video P1H1, in which tests high and fluctuating HR, proving the framework's potential for a newborn specific application.

Careful analysis of the peak refinement step supported what had been described in *Chen's*

Table 7.1: RMSE calculated between extracted and both ground truth HRV signals, using both fixed and adaptive filtering. Detrending was performed with both the extracted HR curve and the HR curve calculated from the ECG.

	Adaptive + ECG		Fixed + ECG		Adaptive + rPPG		Fixed + rPPG	
	GT1	GT2	GT1	GT2	GT1	GT2	GT1	GT2
P1LC1	5.34	8.78	4.86	5.67	5.38	9.23	4.93	6.08
P1LC2	3.78	7.56	3.02	4.24	3.82	4.56	4.24	4.42
P1LC3	3.36	7.58	3.40	4.50	4.20	7.63	3.67	7.16
P1LC4	3.02	6.25	2.87	3.99	3.21	6.35	3.56	5.34
P1LC5	2.46	5.98	2.23	3.54	2.03	6.22	1.98	6.67
P1LC6	3.56	8.69	5.34	7.01	3.79	7.33	5.82	4.02
P1LC7	3.37	6.28	2.78	3.87	3.40	3.91	3.97	4.34
P2LC1	30.49	35.56	28.11	29.15	30.63	29.22	32.62	29.66
P2LC2	29.23	35.04	27.32	28.97	29.87	28.45	30.25	28.87
P2LC3	28.28	34.74	25.78	26.37	28.73	26.40	28.54	26.90
P2LC4	29.43	39.45	33.54	34.75	29.65	35.62	32.32	35.86
P2LC5	28.11	34.28	25.26	26.86	28.56	25.38	28.73	26.65
P3LC1	30.16	39.87	32.15	32.97	30.35	32.73	31.78	34.83
P3LC2	33.07	34.56	32.19	32.80	33.44	32.67	33.58	33.51
P3LC3	30.92	34.03	30.78	31.55	31.35	31.32	31.05	32.47
P3LC4	30.26	33.79	29.98	30.22	30.78	30.50	30.46	30.60
P3LC5	29.96	33.59	29.21	30.32	30.23	30.79	30.08	30.24
P1H1	3.56	7.34	3.22	3.87	3.77	3.93	3.91	4.02

et al. [97] paper. This step is essential, once it artificially recreates resolution which will be valuable for the HRV signal and would not have been possible to recover otherwise. The method in question, changed, on average, the peaks location by 0.018 seconds, reaching 0.03 seconds in the most extreme cases. This resulted in an increase in the average Pearson's correlation coefficient between the IBIs extracted from the rPPG signal and the R-R intervals extracted from the ECG for videos of participant P1. The same metric was not calculated for the other two patients once the signal was so degraded that the position of the peaks did not reflect HRV in any way.

7.3 Summary

In conclusion, despite the results seeming promising for patients with lighter complexions, further tests should be performed, given the fact that the two ground truth measures used are not as reliable as required in a clinical context. More extensive testing should include, participation of neonatal subjects, on whom this technique has never been applied. After analysing the results, it can be concluded that there are two main factors which define the success of the RR extraction, those being the quality of the extracted rPPG signal and the robustness of the peak detection tool. For that reason, the evolution of accurate RR determination based on RSA would benefit from the development of a tool to robustly detect peaks specifically for rPPG problems. Methods for improvement of the signal quality would not only exponentially improve the results of RR extraction,

Table 7.2: Pearson's product-moment correlation calculated between ground truth and extracted HRV signals using both fixed and adaptive filtering. Detrending was performed with both the extracted HR curve and the HR curve calculated from the ECG.

	Adaptive + ECG	Fixed + ECG	Adaptive + rPPG	Fixed + rPPG
P1H1	0.85	0.82	0.79	0.80
P1LC1	0.34	0.77	0.75	0.60
P1LC2	0.52	0.63	0.34	0.60
P1LC3	0.75	0.80	0.65	0.78
P1LC4	0.82	0.98	0.92	0.95
P1LC5	0.85	0.89	2,03	1,98
P1LC6	0.65	0.78	0.83	0.85
P1LC7	0.72	0.68	0.58	0.67
P2LC1	-0.17	0.10	0.05	-0.18
P2LC2	-0.36	0.35	0.28	0.30
P2LC3	0.15	0.12	0.10	0.20
P2LC4	0.2	0.18	0.16	0.22
P2LC5	0.065	0.089	-0.034	0.040
P3LC1	0.012	0.008	0.010	0.012
P3LC2	0.089	0.0093	0.082	0.09
P3LC3	0.14	0.10	-0.10	0.012
P3LC4	0.37	-0.22	0.032	0.32
P3LC5	-0.12	0.10	0.08	-0.12

but would also be beneficial for HR extraction as demonstrated in Chapter 5.

Chapter 8

Validation in Neonatal Subjects

As mentioned in Chapter 4, the late acquisition performed at CMIN allowed the validation of the developed method in neonatal subjects. The built dataset resulted in a challenging dataset, which incorporates not only the challenges integrated in the development database but also some unexpected challenges. These include variations of lighting conditions over time, face obstructions resulting from movement from the neonates and artifacts caused by the glass of the incubator or crib, which can come in the form of distortions, reflections or scratches, which may occlude the region of interest. Furthermore the videos for this dataset integrate more than one challenge as the acquisition conditions were impossible to control and therefore it was impossible to isolate individual challenges.

The developed tool was tested as had been developed with few alterations other than the adaptation of the cut-off frequencies of the passband filter to include the normal ranges for the Heart Rate of newborn subjects. Immediate analysis of the results, unveiled the fact that the framework was unable to detect the face and its landmarks, which compromises the success of the whole tool. This may have been due to the distinct facial proportions of the newborn subjects when compared to adult individuals. As the classifiers for both face and landmark detection were trained on adult subjects, this may justify the algorithms inability to detect such structures. For that reason, these modules had to be overlooked and the ROI defined manually. This will evidently introduce a lot of error for those videos in which there are broad head movements, and thus the framework was unable to extract valid measures of Heart Rate for those videos (two out of the seven recorded). Two other videos included significant damage in the crib's glass, which covered the ROI. For these reasons, these four videos were excluded from analysis. For the remaining videos, the algorithm proved to perform with significant success, specially when having in mind that it was employed as had been assembled, never having been trained on data of neonatal subjects. It was also concluded that the reduced signal quality (addressed in the form of low or even negative values of SNR) made it impossible for *Peak Analysis* to be used to determine the final value of HR and thus, the second best technique tested for this module was employed - *Spectral Analysis* in the form of the FFT. Table 8.1 shows the average RMSE and the Pearson's correlation coefficient for the HR curves of the three remaining videos, those used to draw conclusions.

Table 8.1: Results for the HR extraction in the videos acquired at CMIN.

	RMSE (bpm)	Pearson's Correlation
Participant 2	7.65	0.66
Participant 4	9.26	0.63
Participant 5	3.89	0.72

Figure 8.1 shows the HR curve for the video relative to *Participant 5*, for which the HR extraction was particularly accurate. This video presented considerable lighting variations over time, as its the main illumination source was natural sunlight through a window. As can be seen in the Figure, this effect did not influence the extraction of the said vital sign. It can however be concluded that despite the success in following the overall trend of evolution of the HR, the algorithm lacked ability to detect rapid change in this vital sign, as becomes evident when analysing Figure 8.1. Such inability may be overcome by reducing the window's size at the cost of resolution in frequency, in case the FFT is being used to determine the final value of the HR.

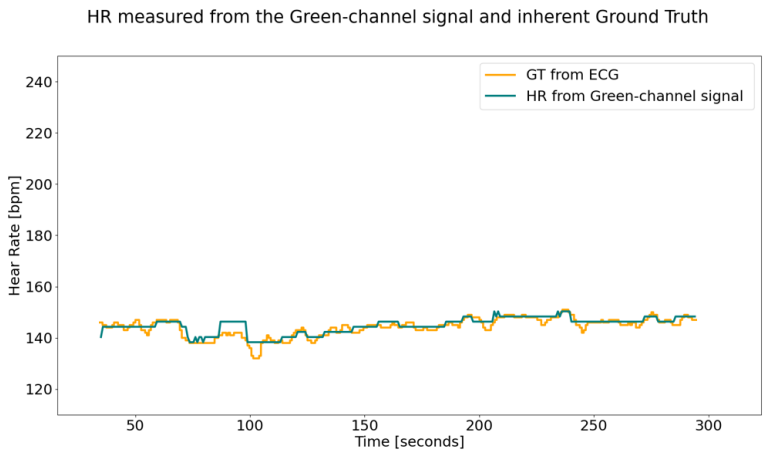


Figure 8.1: Example of continuous Heart Rate curve extracted from one of the videos in the neonatal database and corresponding Ground Truth

8.1 Summary

The method developed displayed extreme potential for a newborn specific application as it was able to continuously extract the HR with relative accuracy for the videos which did not convey the more complex level of challenges. Regardless, a few improvements in specific modules of the framework would most probably translate in successfully extracting the HR even for those challenges, which were considered extremely difficult. In that line of thought, the modules which would require the most effort for improvement would be the modules of Face Detection and Tracking and ROI definition. This framework would benefit from a more robust landmark detection which would be invariant to face obstructions as the infants often move their arms in a way that

partially cover the face. It would also be beneficial to develop an adaptive method for ROI selection which would combine different body structures for an optimal rPPG signal. Nevertheless, this results were considerably satisfactory and should be backed up with more extensive testing.

Chapter 9

Conclusion and Future Work

More than 11% of all births worldwide occur after an incomplete gestational period, which results in 15 million babies being born preterm each year. In fact, preterm birth and its complications are the second leading cause of death among children under 5 years of age, responsible for approximately 1 million deaths. Since preterm birth is a synonym of underdevelopment of the organs and body functions, an infant who does not spend sufficient time in the womb is born as a fragile being who is highly susceptible to the conditions of the environment around him. That being said, it is of extreme importance that preterm infants are carefully monitored to assure their health status does not deviate from the desired, while they are being assisted in NICUs for their first weeks of life. Conventionally, monitoring vital signs in preterm infant's, such as HR and RR, is performed via probes affixed to their skin. However, such instruments may cause damage to the epidermis and increase the risk of infection. Therefore, contactless monitoring solutions appeared as a potential replacement for the current methods used in NICUs. These approaches rely on recent advances in image capturing methods as well as Computer Vision techniques.

When it comes to HR and RR extraction, segmentation can be made into two distinct pathways: (1) color-based methods, which rely on the temporal fluctuations in light reflected by the skin due to variation in blood volume caused by cardiac pumping, phenomenon known as BVP, to extract the so called rPPG signal and (2) motion-based methods, which are able to track individual pixels and extrapolate the HR and RR from the periodic motion of that individual pixels caused by the flux of blood which enters the neonate's head. Once color-based methods tend to be more robust to subject motion and typically neonate's have few restrictions to head movement when lying in an incubator, the former group of techniques was studied in more depth. Although this type of methods is well established and abundantly reported for adult subjects in controlled environments, the same does not apply for neonatal infants, particularly in a real-world scenario. The aim of the work developed was hence to develop and validate an rPPG tool for HR and RR extraction capable of performing for neonatal subjects in a NICU environment.

However, the global pandemic which arose concurrently with the development of this study, did not permit the acquisition of a dataset specific for neonatal subjects in uncontrolled conditions, which was to be built in partnership with CMIN. As an alternative, the *Public Benchmark Dataset*

for Testing rPPG Algorithm Performance created at the *Eindhoven University of Technology* [25] was used. This dataset, thought for testing and benchmark purposes as the name suggests, consists solely of adult subjects and therefore only allowed for the development of the mentioned technique. Although the developed method serves as proof of concept, its analysis should be complemented with tests on a more extensive dataset, which mimics the NICU's environment and with neonatal subjects: the target conditions and population.

The framework developed is capable of simultaneously preform continuous HR, HRV and RR extraction and has proved its worth in challenging conditions. Despite its weaknesses, namely the inability to cope with severe motion and the discrepancy in results when leading with different complexions, it has managed to outcome the challenges of high and fluctuating HR and challenging lighting conditions, proposed in the used dataset. These two challenges had particular significance and were those whose conquest mattered the most, as they are the ones which better reflect the challenges more likely to be faced in a NICU environment. Since homogeneous lighting conditions cannot be assured in a real-word NICU environment and even light temperature and intensity can vary from unit to unit, the framework should be robust enough to withstand a wide range of lighting conditions.

The modularity of the framework developed for contactless extraction of the HR allowed for each of its components to be evaluated individually, while having in mind the overall performance of the algorithm. It became clear that detection of facial landmarks for tracking allows for a much more precise ROI definition than the use of a tracker method such as Median Flow. When testing for newborns, it must be considered that these infants frown and cry with intensity, which accounts for non-rigid motions, not tested in this dataset. It will thus be beneficial to use a landmark detection method which can accurately identify more keypoints in order to get a more specific ROI (more complex shapes rather than the hexagon used), contributing for robustness against non-rigid motions. Furthermore, it was concluded that by focusing on the cheeks region rather than the forehead, not only was the discrepancy between performance for different skin tones attenuated but also it is presumably easier to consistently find patches of uncovered skin in the newborns.

Although the perks associated with the use of either BSS and Chrominance signals should be evident, both methods under-performed for all lighting conditions tested when in comparison with the use of the Green Channel signal, specially for non-white illumination (represented in video P1LC6) where these methods should be far superior. For that reason, no signal was discarded and all should be tested in a dataset which uncovers more lighting conditions and counts with the participation of more subjects.

The variety of pipeline configurations tested led to believe that perks of rPPG signal peak analysis are evident and not only make RR extraction possible, but also can significantly improve HR measures. Advantages of using Peak Analysis over the FFT reside on the surprising increase in resolution associated with the frequency resolution limit imposed by the number of samples when using the FFT. However, the frequency extraction method which is not power spectral analysis is not as robust, since this method is much more sensitive to outliers and low SNR. It would be therefore valuable to develop a more robust peak detection algorithm which is capable of dealing

with outliers as well as preform peak refinement in order to artificially recreate a higher sampling frequency.

Regarding the tests performed and their validity, it is important to understand the limitations imposed. Although the dataset is very useful in the sense that it can help to get a notion of the algorithm's performance beyond the more simple and controlled conditions, this dataset lacks depth, which makes it difficult to assess repeatability. The fact that each video addresses one specific condition, whether it is the combination of a particular lighting condition with a specific skin tone, the presence of movement or high heart rates, makes it impractical to split the dataset into training and test videos, since each condition would only be covered in one of the subsets. Due to the relatively short length of the video, as well as the buffer period associated with extracting the first rPGG sample, it is also impossible to use a few samples from all videos for training and the other samples for testing. Reducing the buffer period, and consequently the sample size was prejudicial approach as well since it severely affected the frequency resolution of the FFT and hence the results. Besides, by doing so, any temporal relationship between samples would be lost and thus, techniques such as adaptive filtering and post-processing would be impossible to apply. In the particular case of video P1H1 (which contains high heart rates at the beginning) its division would generate an unbalanced dataset which would lack high HR samples in one of the subsets. This is particularly concerning as P1H1 is arguably the most important video because it reflects the high heart rates which are to be found in neonatal participants.

One other downside to the used dataset is that the three videos assigned to *Challenge 3* contain exaggerated and rhythmic head movement which in no way reflects the natural head motion which one would expect from a neonate resting inside an incubator. Although the algorithm under performed for these videos, it is believed that the tool would better handle the slight head movements of the infants.

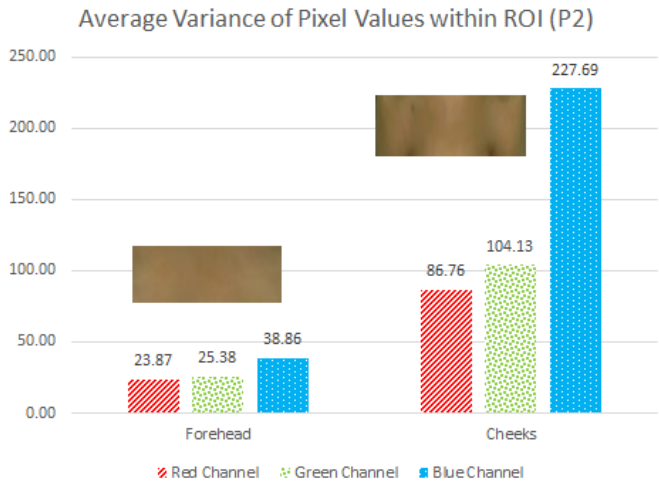
In addition to continuous vital sign extraction framework, an algorithm for amplification of subtle color changes in video was also employed in order to emphasise the BVP phenomenon. Despite being a fascinating method, which has a vast range of applicabilities even outside the biomedical field, EVM applied to contactless vital sign extraction serves mostly a monitoring and visualization purpose rather than improving the extraction of the vital sign itself.

With regard to future work, the most urgent task to be carried out is to test the described framework for neonatal subjects, in order to validate the work developed. Such tests will most certainly unveil minor tweaks or even possible structural changes required for optimizing the tool for its target subjects, such as parameter optimization, namely the number of independent sensors used, the length and displacement of each samples, among others. Other interesting extensions to this study, which would complement the work developed so far would be the adaptation of the described framework for Near Infrared (NIR) imagery. NIR light is often captured with the use of active cameras. This type of cameras differ from passive cameras by possessing an emitter, which casts light with specific wavelengths to illuminate an area of interest, and having a sensor sensitive to that same wavelengths, which captures the radiation reflected back to the camera and interprets it to generate an image. The use of NIR passive cameras would allow for the developed method

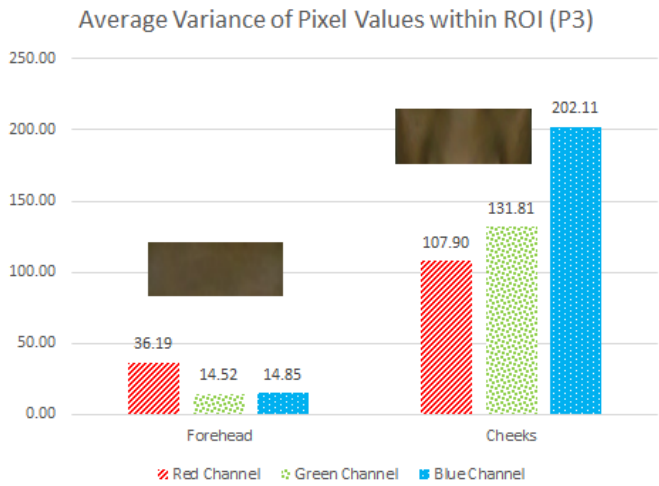
to be used inside the NICU even in the absence of natural and artificial light, as happens during the night, without disturbing the infants and the normal routine of the practitioners. Furthermore, rPPG methods for both HR and RR extraction would benefit from improvements that address the quality of the rPPG signal, once high and consistent signal quality will open new doors for the capabilities of such technologies. Regarding this topic, it would be of great interest to exploit the recent advances in Artificial Intelligence and Signal Processing to facilitate the reconstruction of the ECG from the rPPG signal extracted. Such would be extremely beneficial once the ECG, regardless of how many leads are extracted, contains much more valuable information beyond the HR and RR. FC

Appendix A

Additional Plots for Heart Rate Extraction

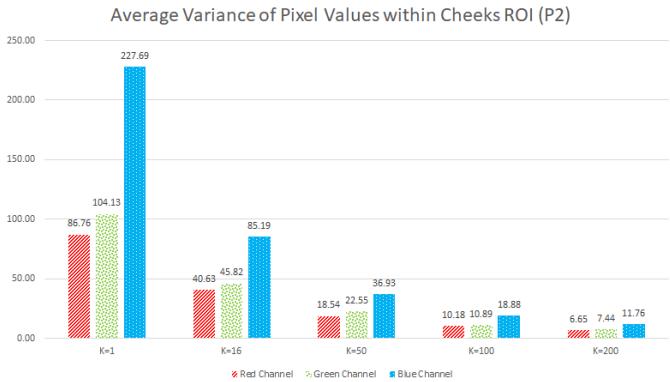


(a)

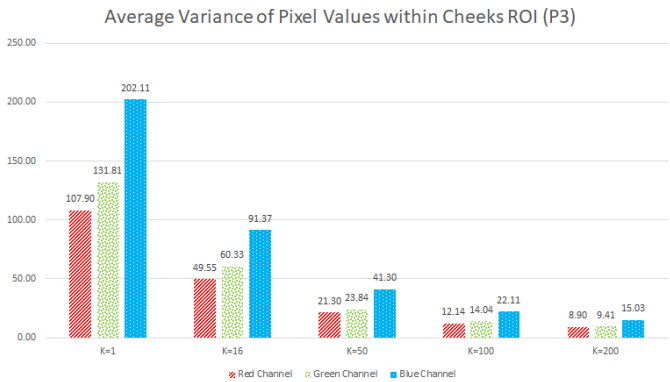


(b)

Figure A.1: Differences in pixel Value Variance imposed by using distinct skin regions as ROI: (a) P2; (b) P3.



(a)



(b)

Figure A.2: Average pixel value variance for increasingly smaller sensors within the ROI: (a) P2; (b) P3.

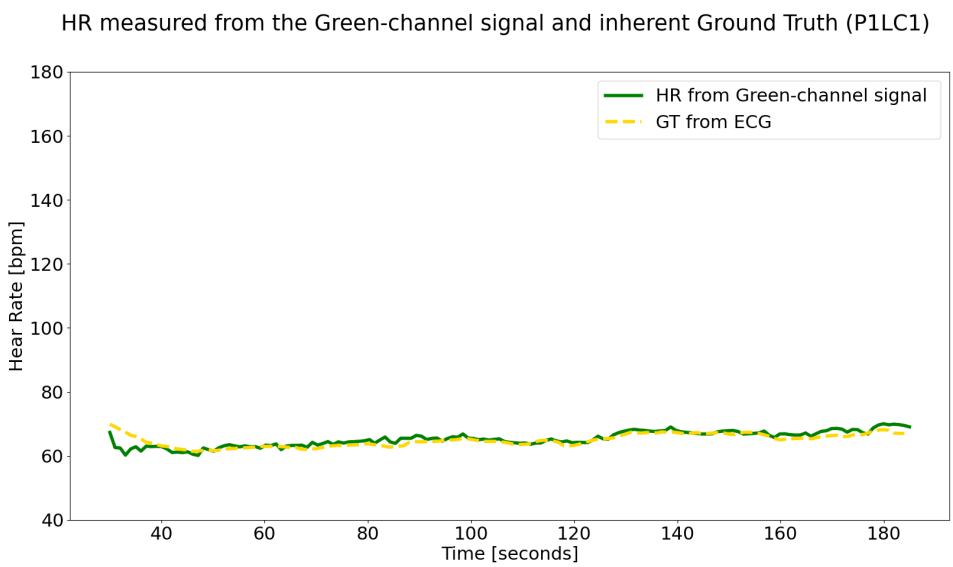


Figure A.3: Continuous HR curve extracted from video P1LC1 using configuration C11.

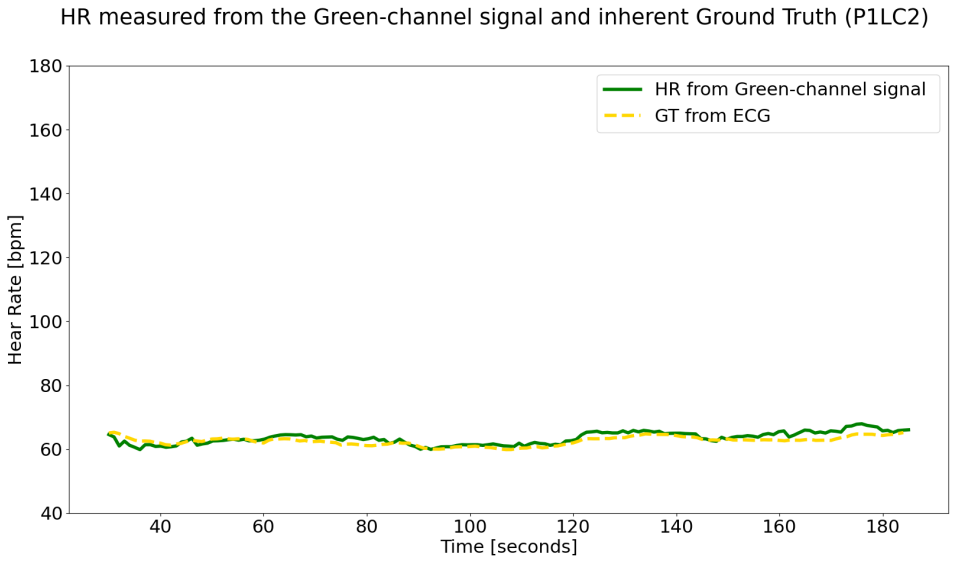


Figure A.4: Continuous HR curve extracted from video P1LC2 using configuration C11.

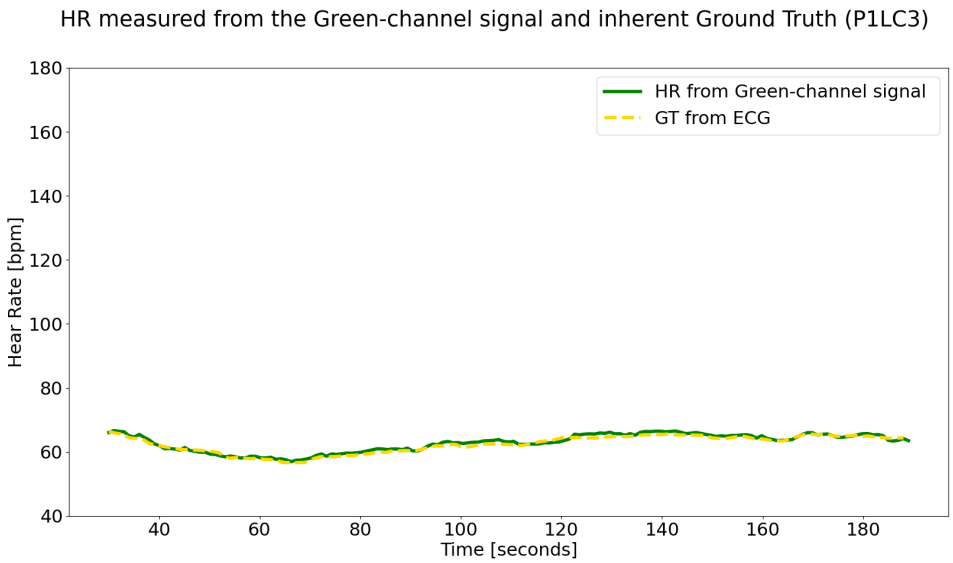


Figure A.5: Continuous HR curve extracted from video P1LC3 using configuration C11.

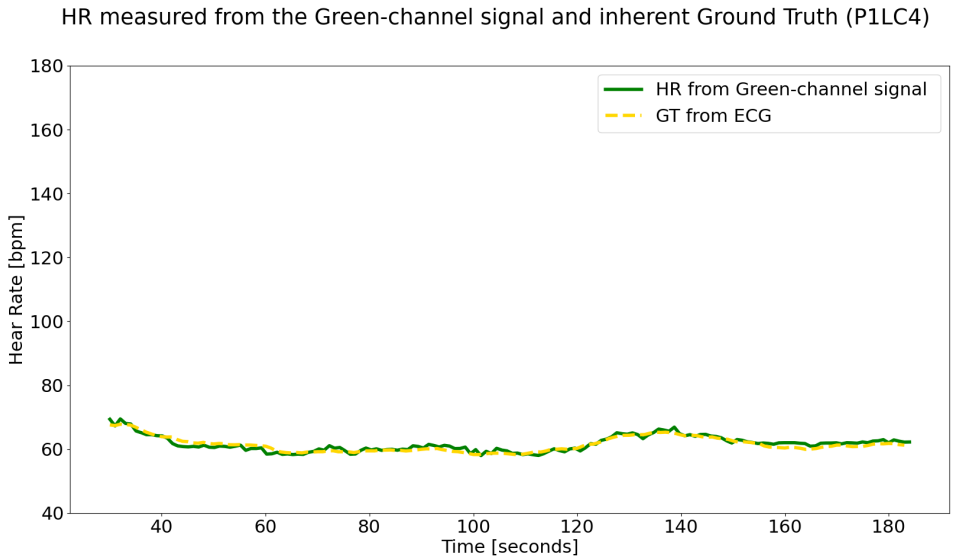


Figure A.6: Continuous HR curve extracted from video P1LC4 using configuration C11.

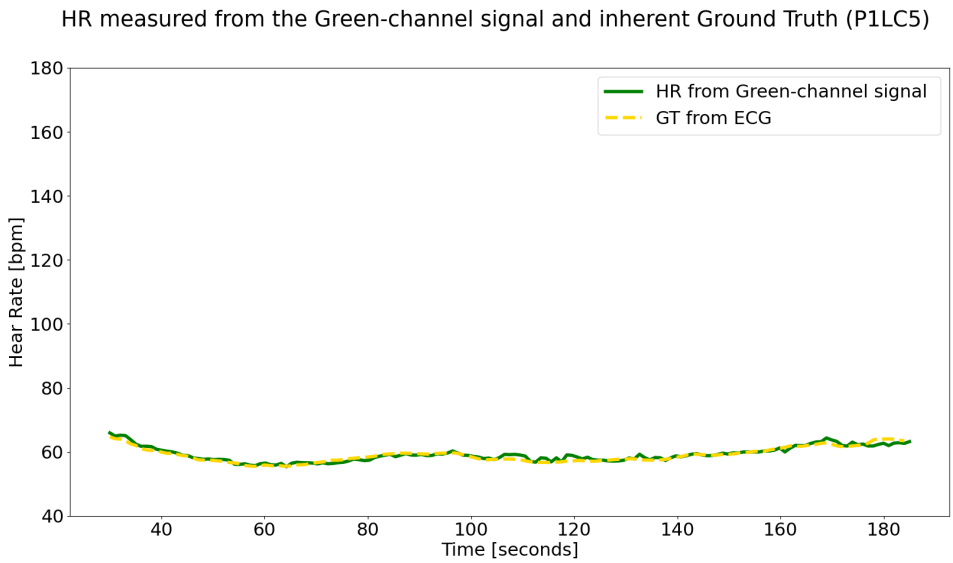


Figure A.7: Continuous HR curve extracted from video P1LC5 using configuration C11.

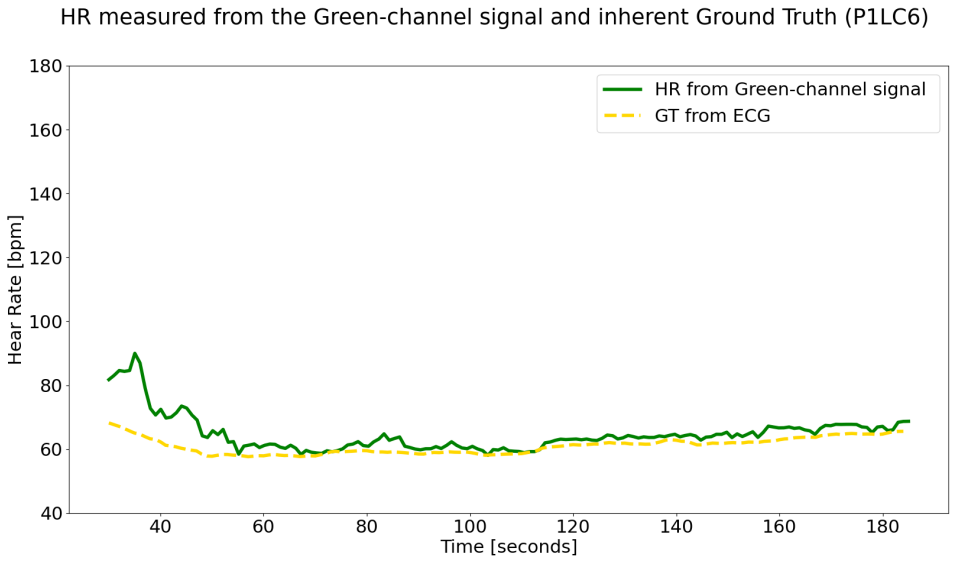


Figure A.8: Continuous HR curve extracted from video P1LC6 using configuration C11.

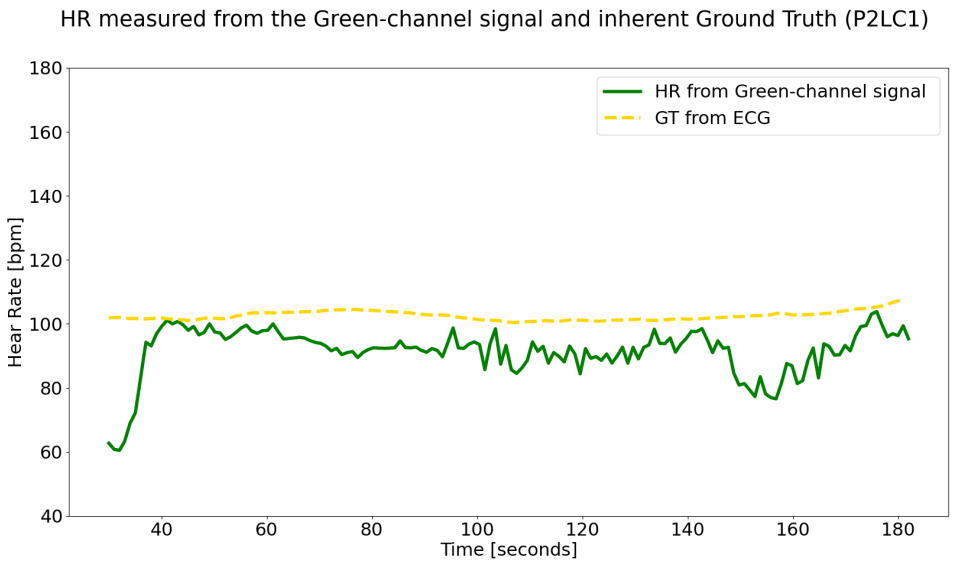


Figure A.9: Continuous HR curve extracted from video P2LC1 using configuration C11.

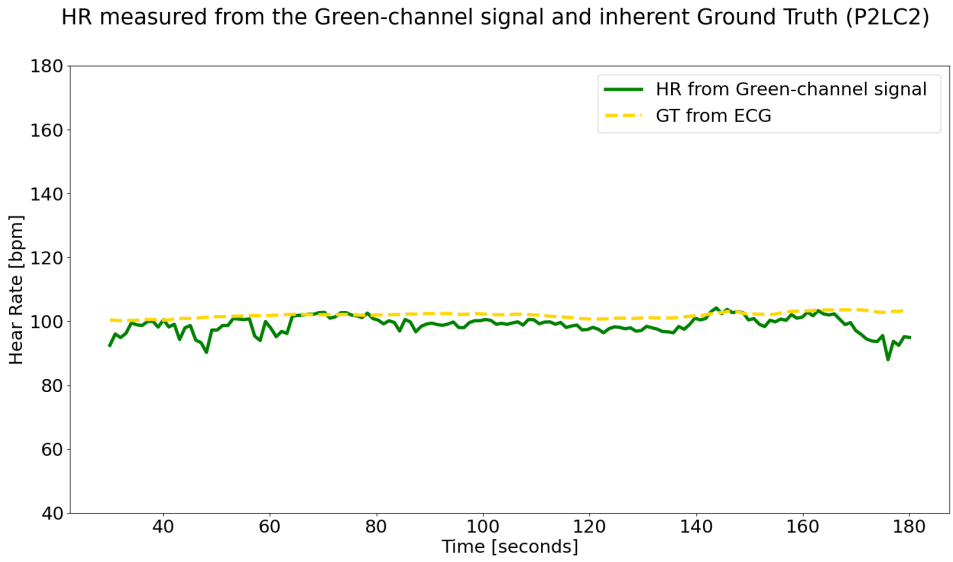


Figure A.10: Continuous HR curve extracted from video P2LC2 using configuration C11.

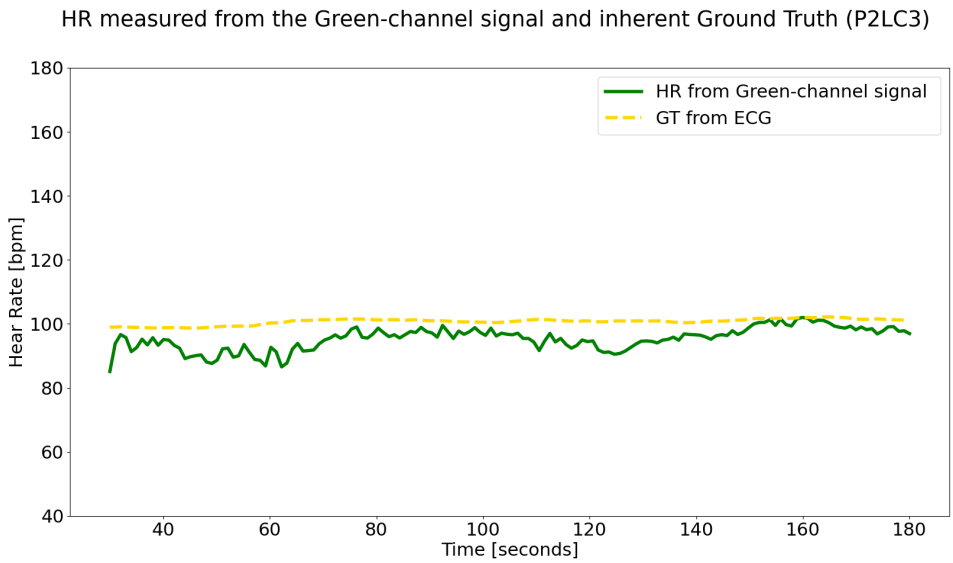


Figure A.11: Continuous HR curve extracted from video P2LC3 using configuration C11.

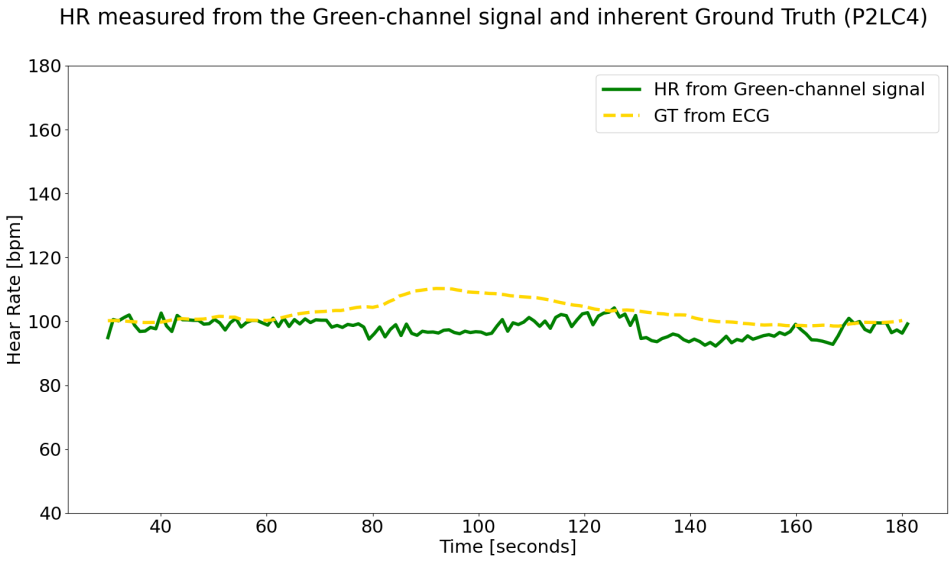


Figure A.12: Continuous HR curve extracted from video P2LC4 using configuration C11.

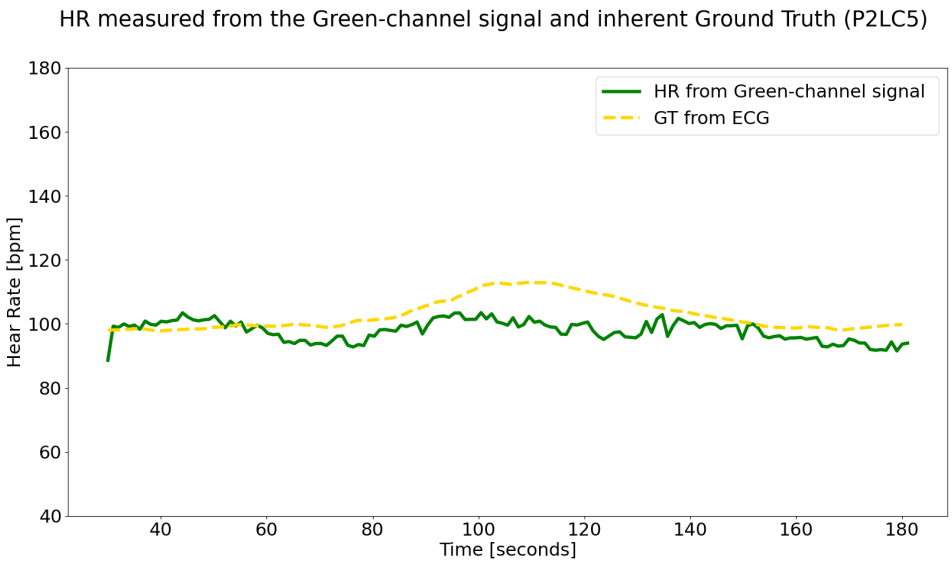


Figure A.13: Continuous HR curve extracted from video P2LC5 using configuration C11.

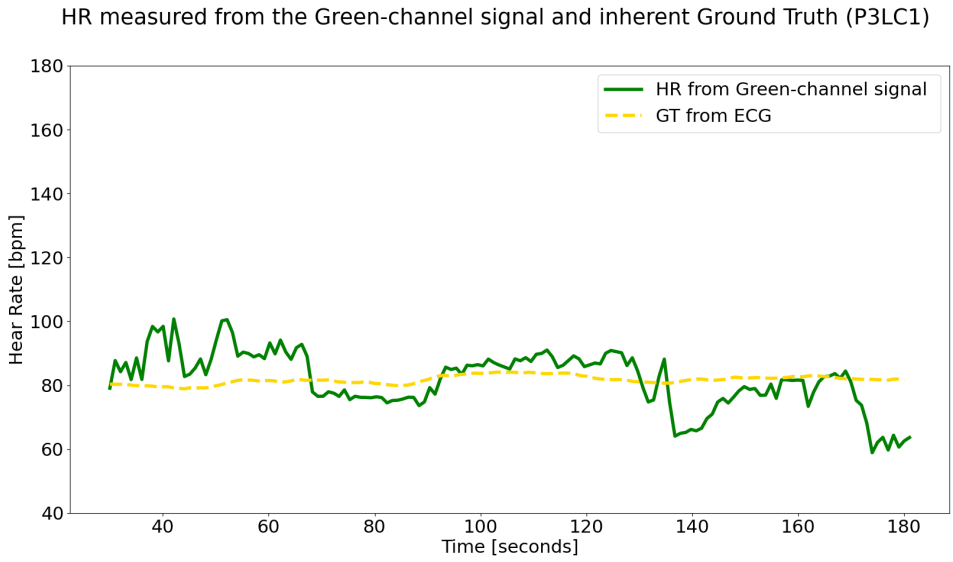


Figure A.14: Continuous HR curve extracted from video P3LC1 using configuration C11.

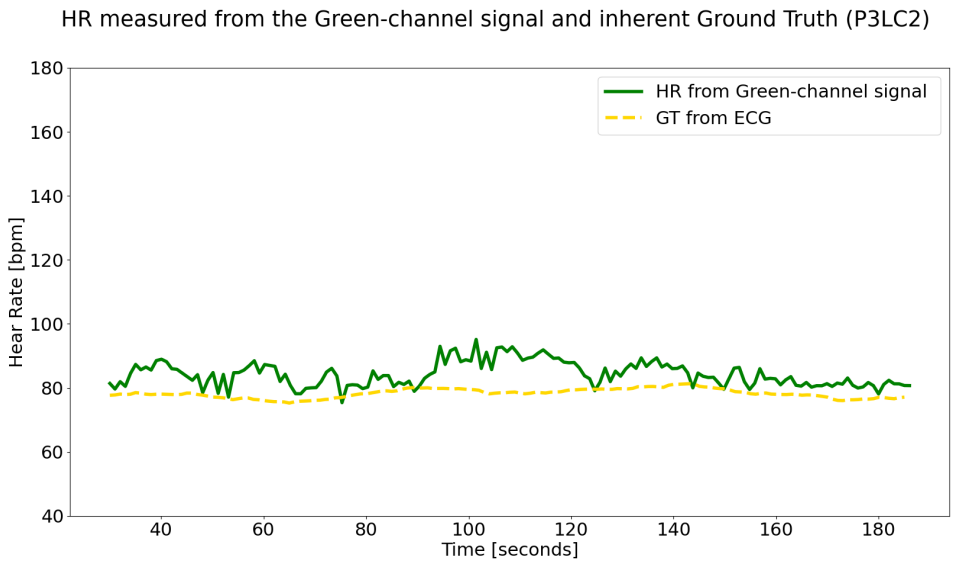


Figure A.15: Continuous HR curve extracted from video P3LC2 using configuration C11.

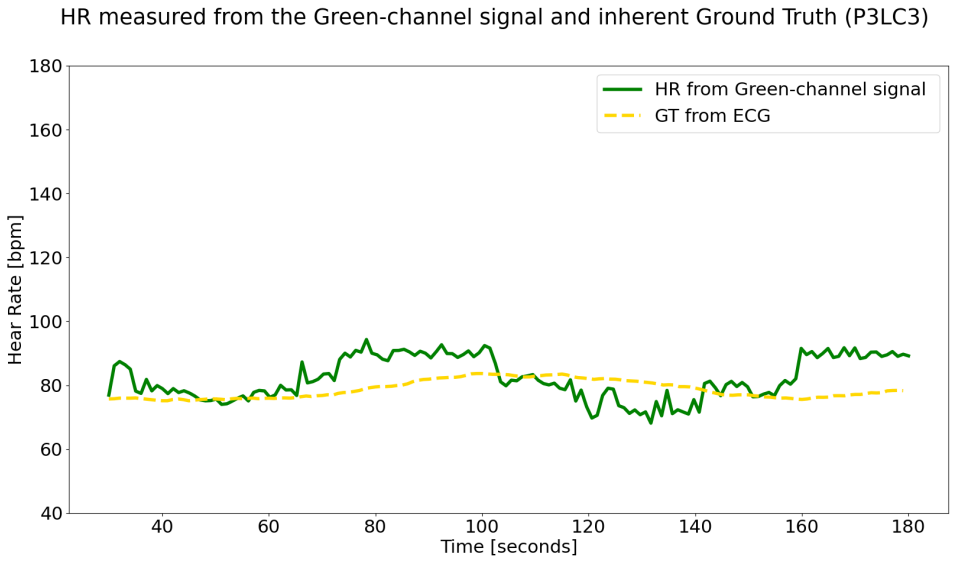


Figure A.16: Continuous HR curve extracted from video P3LC3 using configuration C11.

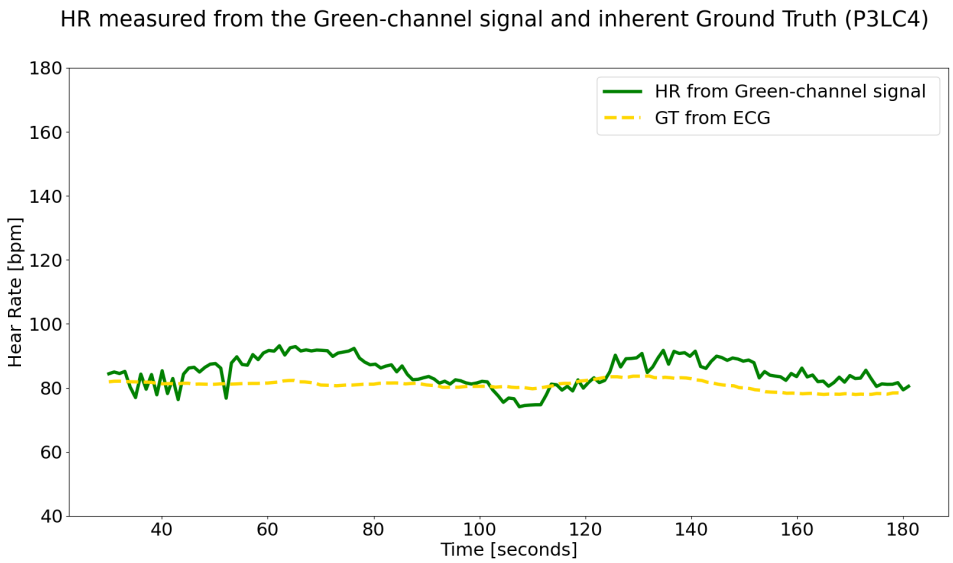


Figure A.17: Continuous HR curve extracted from video P3LC4 using configuration C11.

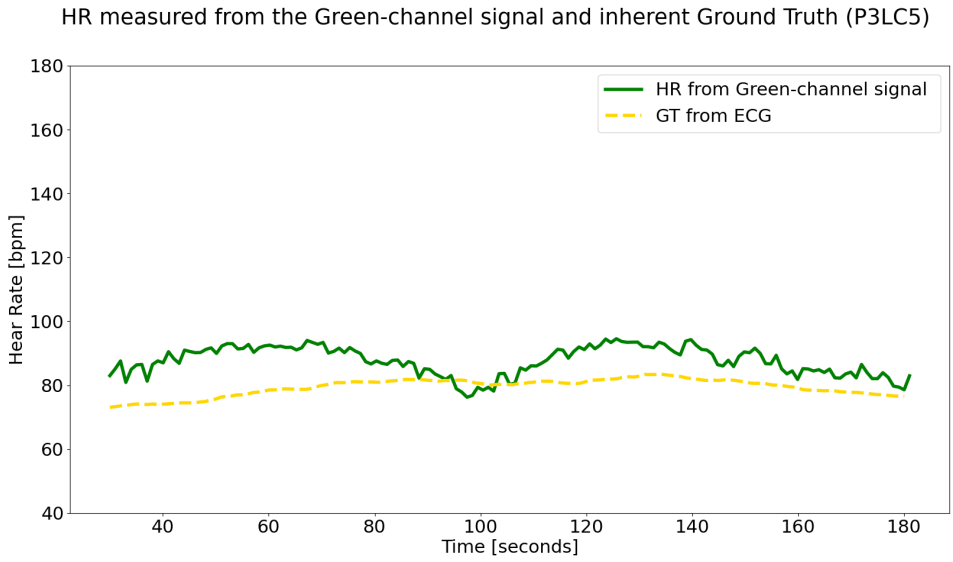


Figure A.18: Continuous HR curve extracted from video P3LC5 using configuration C11.

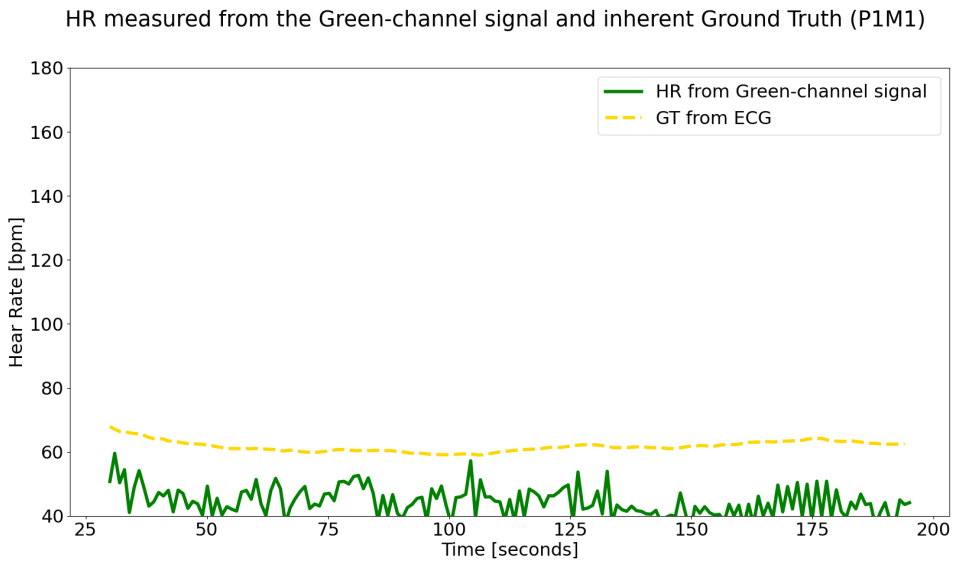


Figure A.19: Continuous HR curve extracted from video P1M1 using configuration C11.

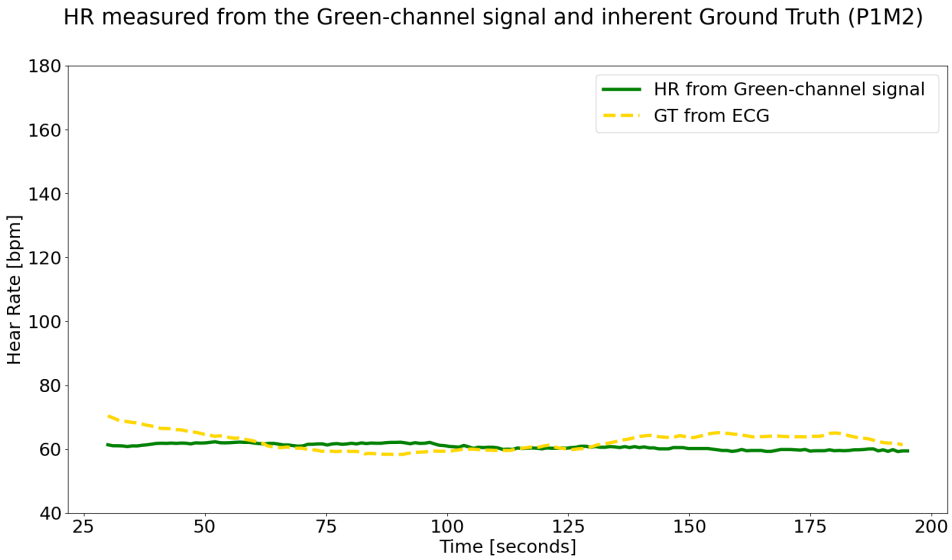


Figure A.20: Continuous HR curve extracted from video P1M2 using configuration C11.

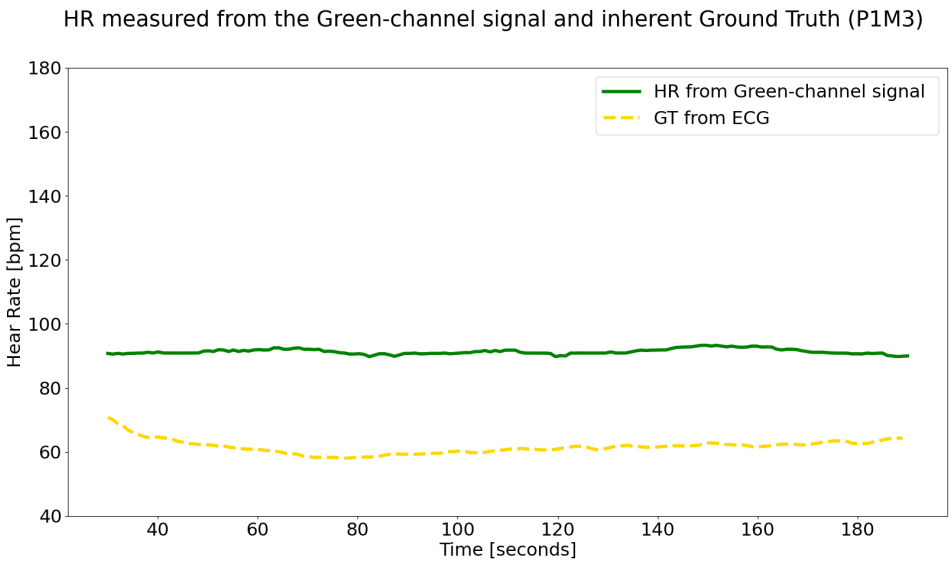


Figure A.21: Continuous HR curve extracted from video P1M3 using configuration C11.

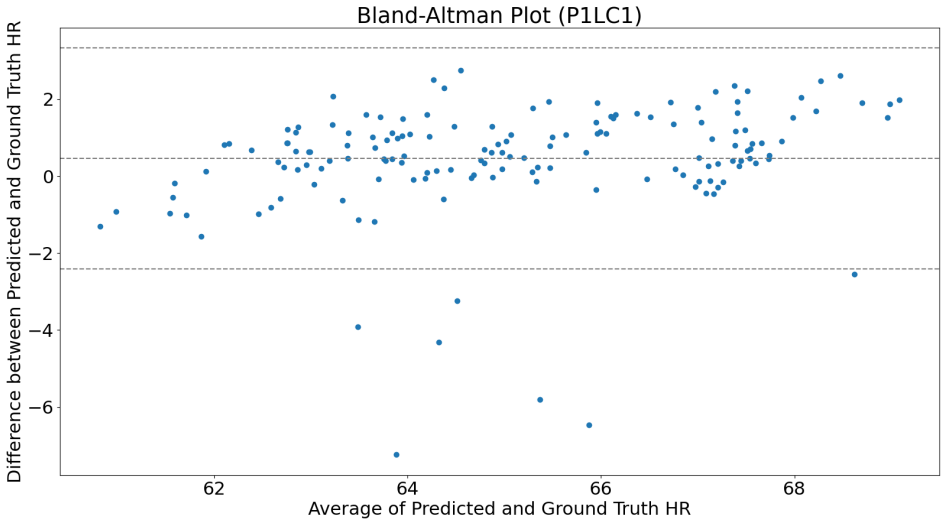


Figure A.22: Bland-Altman plot for the HR samples extracted from video P1LC1 using configuration C11.

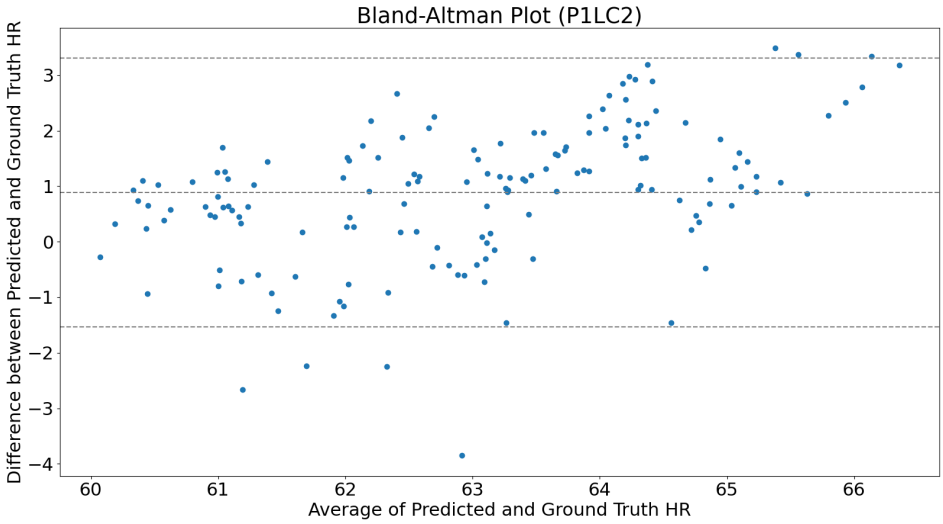


Figure A.23: Bland-Altman plot for the HR samples extracted from video P1LC2 using configuration C11.

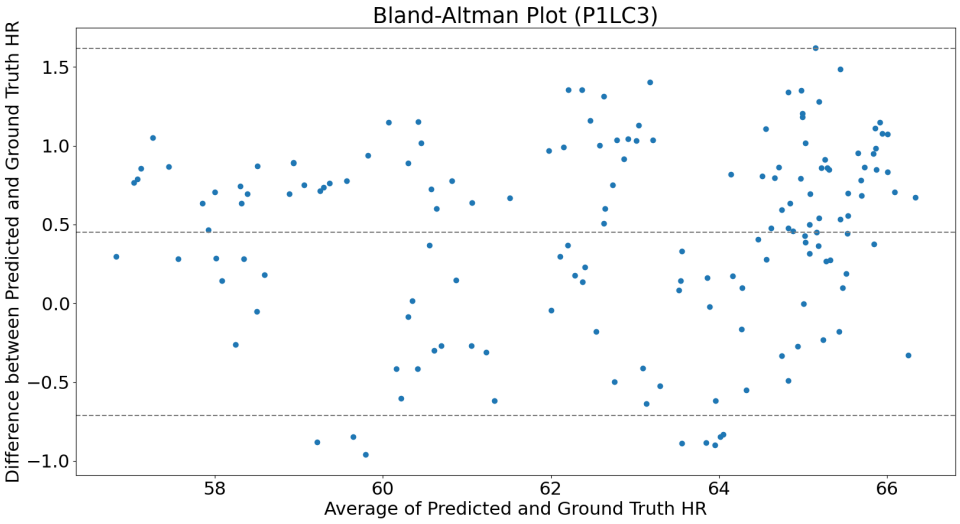


Figure A.24: Bland-Altman plot for the HR samples extracted from video P1LC3 using configuration C11.

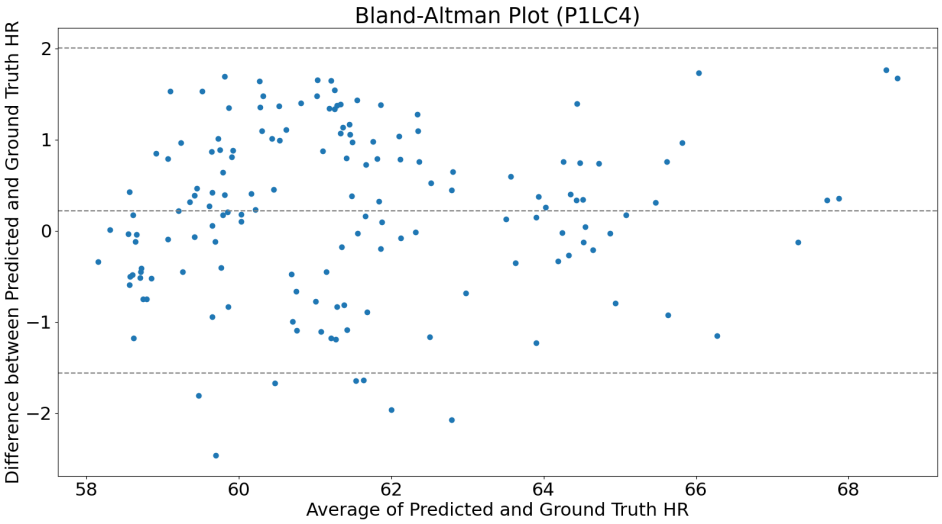


Figure A.25: Bland-Altman plot for the HR samples extracted from video P1LC4 using configuration C11.

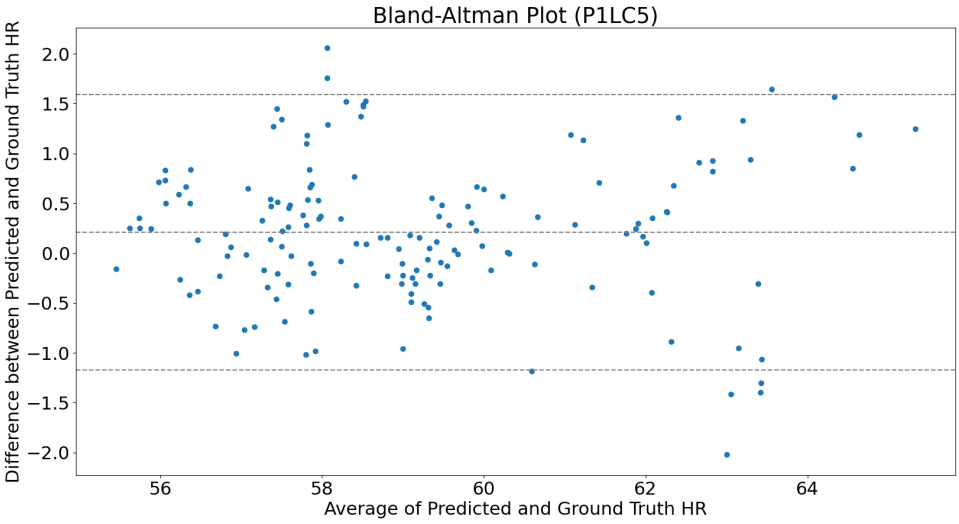


Figure A.26: Bland-Altman plot for the HR samples extracted from video P1LC5 using configuration C11.

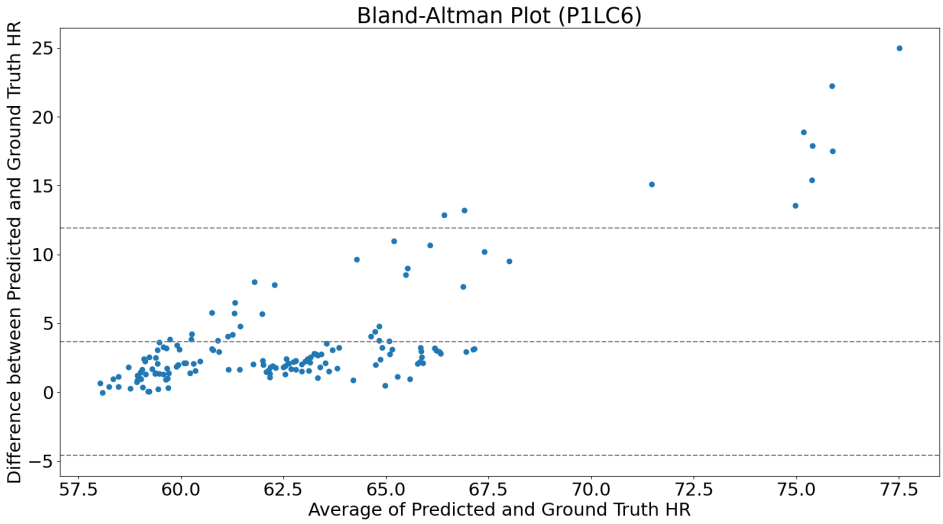


Figure A.27: Bland-Altman plot for the HR samples extracted from video P1LC6 using configuration C11.

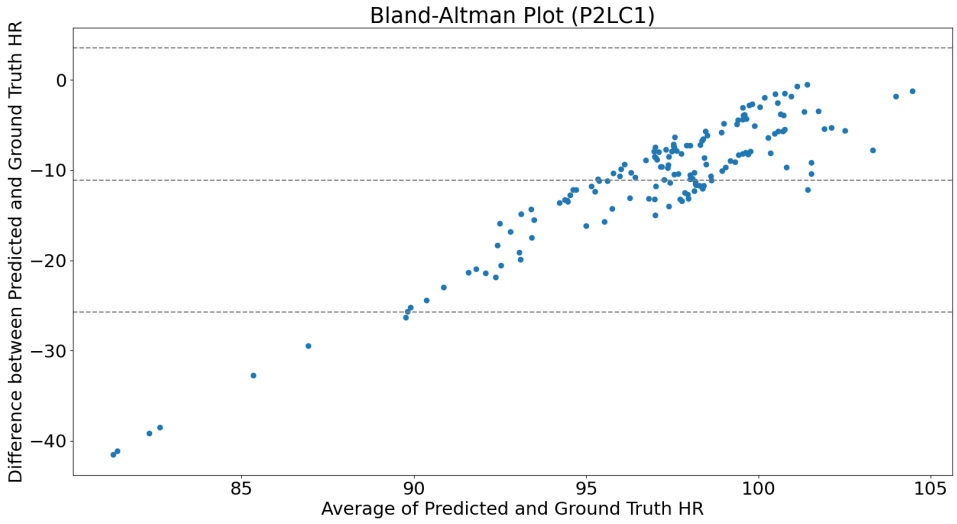


Figure A.28: Bland-Altman plot for the HR samples extracted from video P2LC1 using configuration C11.

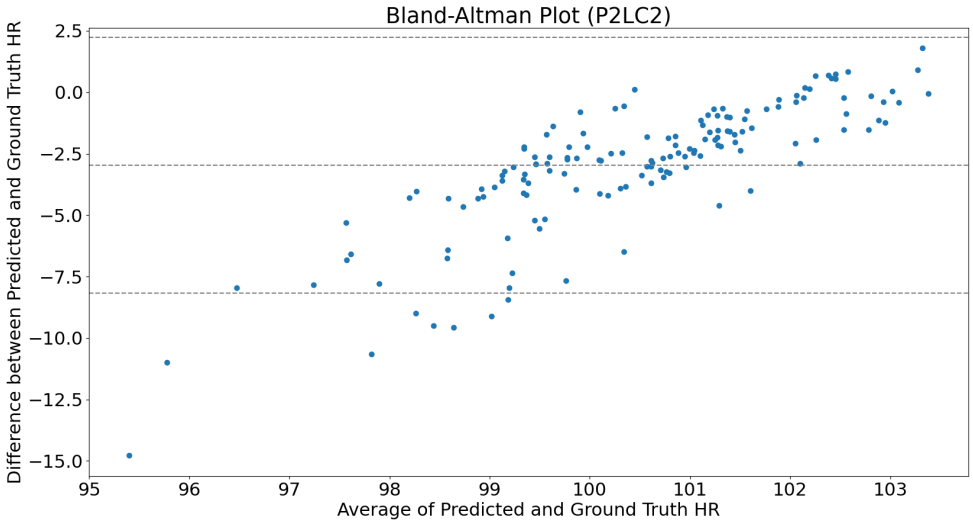


Figure A.29: Bland-Altman plot for the HR samples extracted from video P2LC2 using configuration C11.

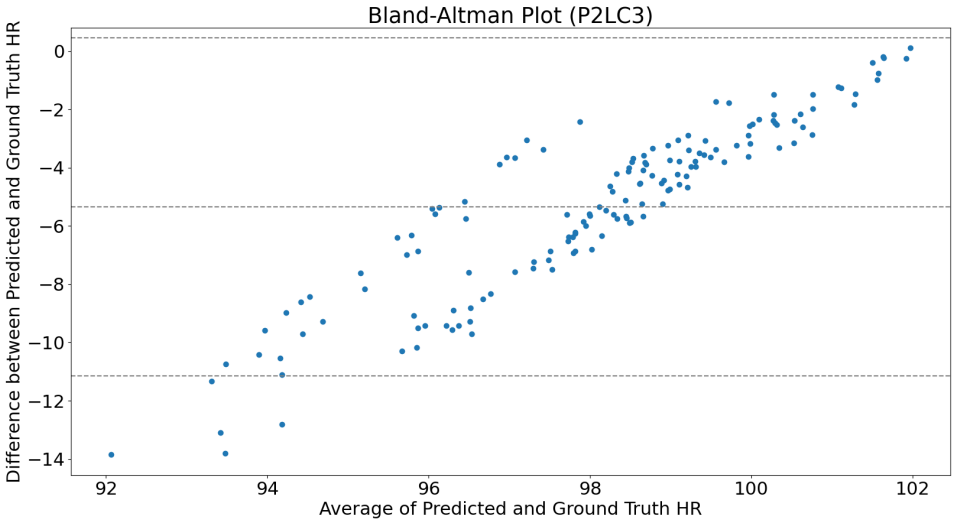


Figure A.30: Bland-Altman plot for the HR samples extracted from video P2LC3 using configuration C11.

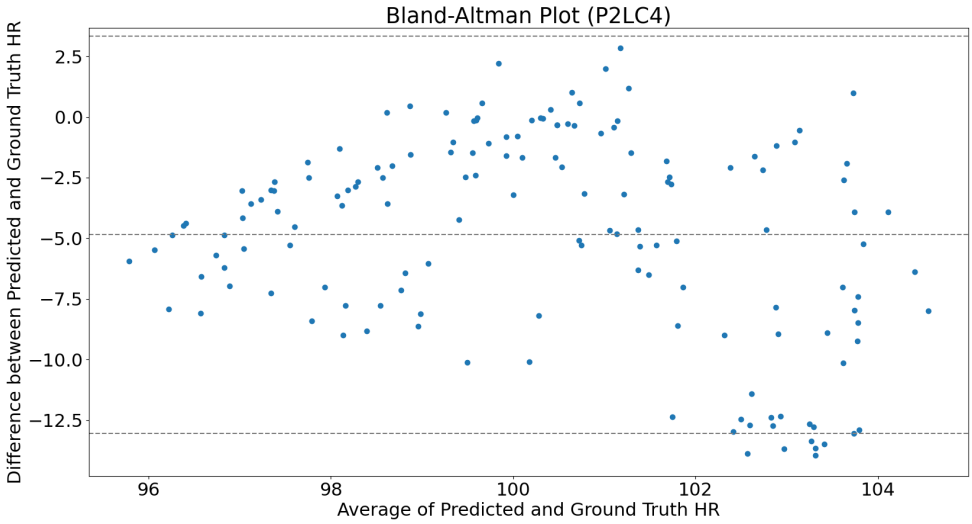


Figure A.31: Bland-Altman plot for the HR samples extracted from video P2LC4 using configuration C11.

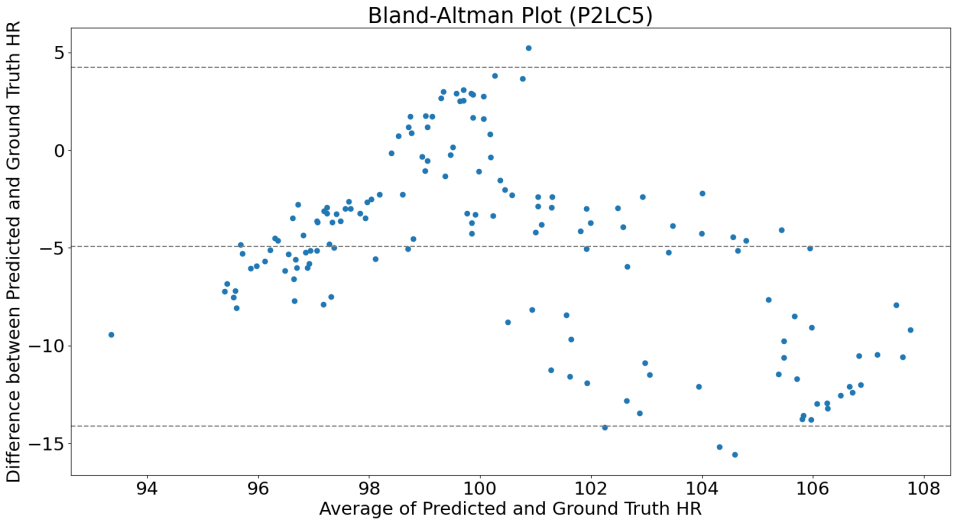


Figure A.32: Bland-Altman plot for the HR samples extracted from video P2LC5 using configuration C11.

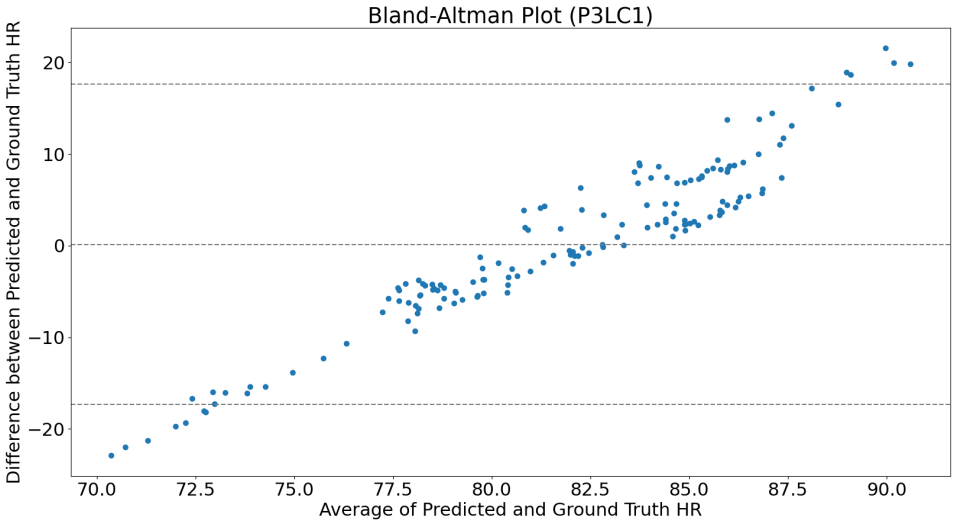


Figure A.33: Bland-Altman plot for the HR samples extracted from video P3LC1 using configuration C11.

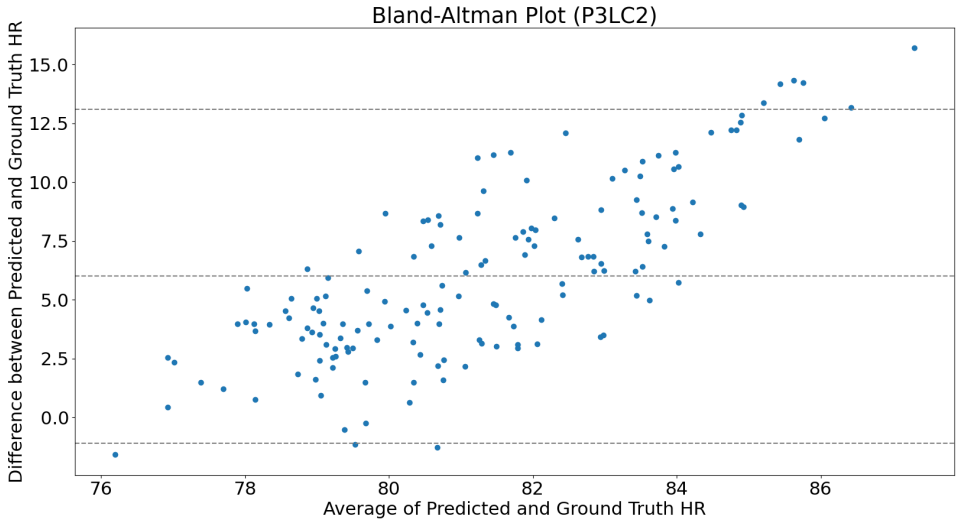


Figure A.34: Bland-Altman plot for the HR samples extracted from video P3LC2 using configuration C11.

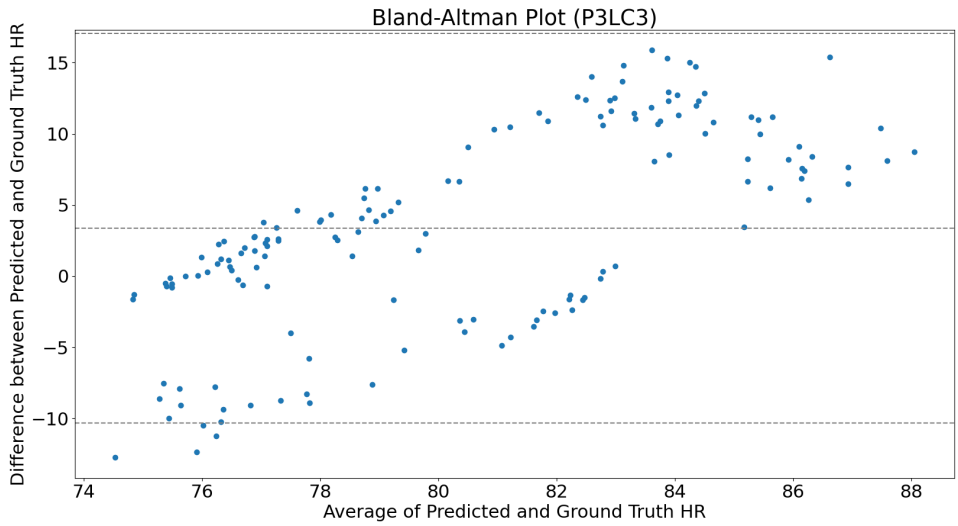


Figure A.35: Bland-Altman plot for the HR samples extracted from video P3LC3 using configuration C11.

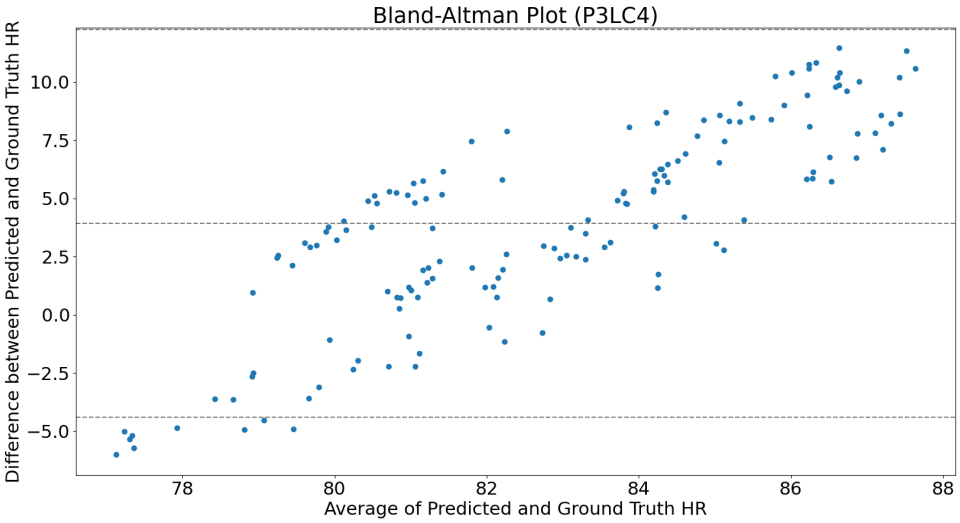


Figure A.36: Bland-Altman plot for the HR samples extracted from video P3LC4 using configuration C11.

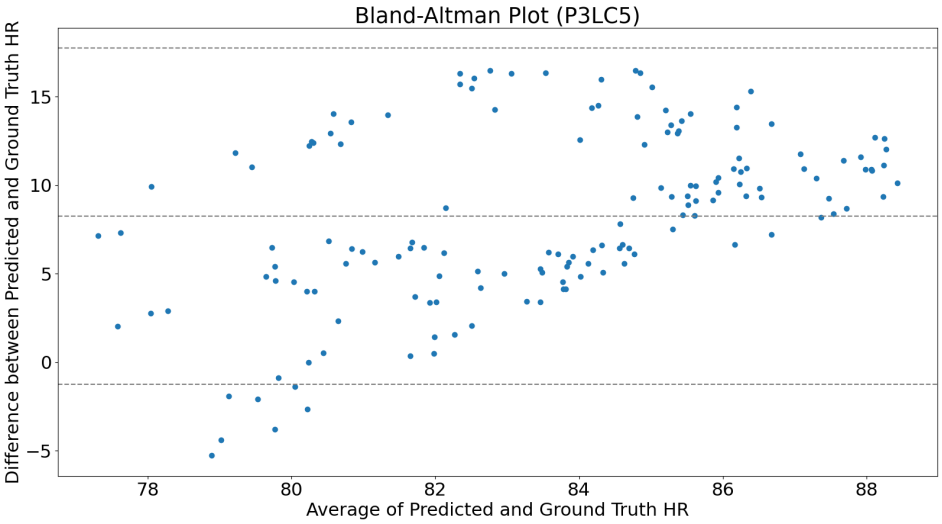


Figure A.37: Bland-Altman plot for the HR samples extracted from video P3LC5 using configuration C11.

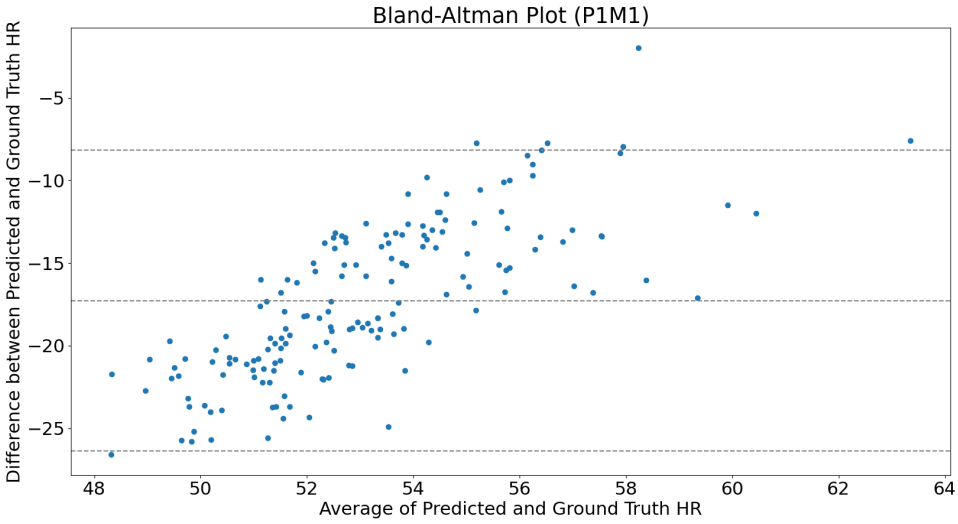


Figure A.38: Bland-Altman plot for the HR samples extracted from video P1M1 using configuration C11.

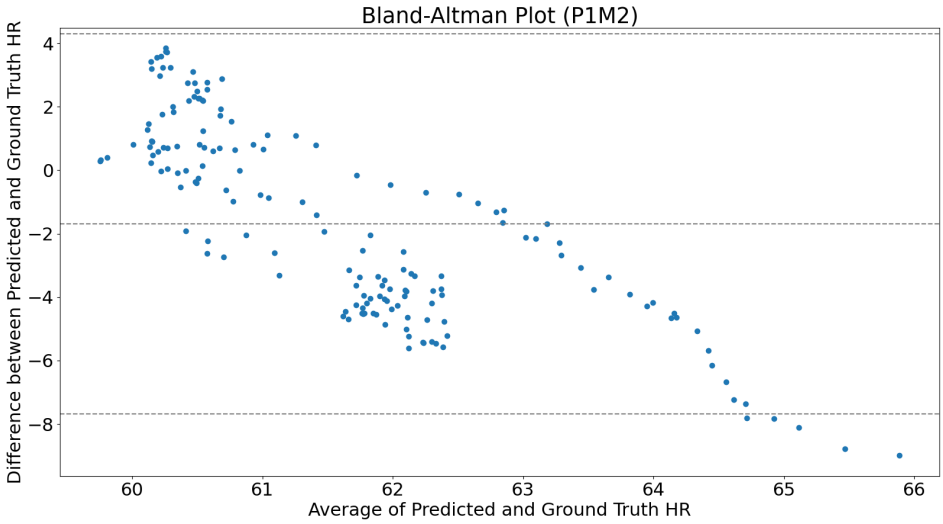


Figure A.39: Bland-Altman plot for the HR samples extracted from video P1M2 using configuration C11.

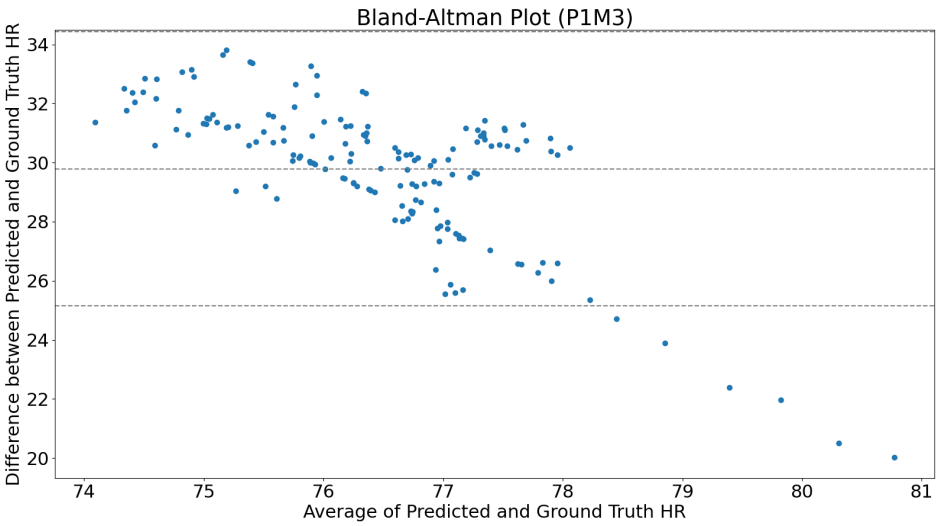


Figure A.40: Bland-Altman plot for the HR samples extracted from video P1M3 using configuration C11.

Appendix B

Leaflet

MONITORIZAÇÃO SEM CONTACTO DE BEBÉS PRÉ-TERMO

Photo by Sharon McCutcheon on Unsplash

Este estudo visa desenvolver e validar um método capaz de monitorizar o batimento cardíaco e ritmo respiratório em recém-nascidos pré-termo, recorrendo a captura de vídeo, em vez dos usuais sensores colocados na pele da criança. Para além disso, também visa a criação de um método para quantificar dor sentida pelo recém-nascido, baseado na imagem e nos sinais vitais previamente extraídos. Para esse propósito, para cada participante serão gravados vídeos com duração não superior a 5 minutos e serão armazenados os sinais vitais que são normalmente extraídos. A participação neste estudo, não prejudicará o bem-estar do recém-nascido, uma vez que a recolha dos vídeos e dos sinais vitais não implica um acesso extraordinário ao interior da incubadora. Todos os dados relativos à identificação dos Participantes neste estudo são confidenciais e será mantido o anonimato. Os dados recolhidos serão mantidos pelo período de um ano após a data de recolha. Muito obrigado pela sua contribuição.

INVESTIGADOR PRINCIPAL

Hélder Oliveira Filipe Pinto de Oliveira, PhD | helder.f.oliveira@inesctec.pt

EQUIPA INTEGRANTE

Sara Campos Monteiro Sabino Domingues, MD| saradomingues@hotmail.com
Diogo Terleira Malafaya Baptista | diogo.t.baptista@inesctec.pt

INESCTEC

INESCTEC
Campus da Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, 4200-465 Porto
+351 222 094 000 | info@inesctec.pt | https://www.inesctec.pt

CMIN

CENTRO MATERNO INFANTIL DO NORTE
Largo da Maternidade, 4050-371 Porto
22 207 75 00

Appendix C

Acquisition Protocol

NON-CONTACT MONITORING OF PRETERM INFANTS

1. Introduction

Monitoring of new-borns is a challenging task, which is carried daily at every Neonatal Intensive Care Unit (NICU). Due to the delicate state of equilibrium in neonates' health, vital sign monitoring is important, as it allows for early detection of medical issues and therefore actively contributes for the infant's well-being and health. Conventionally, new-borns are monitored via probes affixed to their skin. However, such instruments may cause damage to the epidermis and increase the risk of infection.

Non-contact imaging methods represent an alternative to record physiological signals such as heart rate, respiratory rate and body temperature. Advantages of contactless monitoring methods include lack of contact with skin (reduces skin breakdown) and minimizing the number of probes and monitors used (leaves more body surface-area for other care).

In order to assure that the maximum amount of relevant medical information can be extracted with the minimum amount of time and resources, variability due to extraneous factors must be reduced or eliminated. In order to do so, it is essential that the images conform to a standard, which also ensures that the acquisition meets with the research goals of the project.

The primary goal of this project is to develop an accurate visual based method to monitor the heart rate, respiratory rate and body temperature of preterm infants in Neonatal Intensive Care Units. Secondary goals include monitoring other vital signs as well as building an easy to use digital interface. Validating the method developed dictates that not only we acquire different types of images (described as "modalities" in this document) but also that we record the desired vital signs through the current methods used clinically, for ground truth purposes.

2. Goals

The primary goals of this project are:

- Developing an accurate visual based method to monitor the heart rate and respiratory rate of preterm infants in NICUs.
- Another entry in the list

Secondary goals include:

- Monitoring other relevant vital signs.
- Building an easy to use digital interface, which...

3. Methods

Two optical image modalities from one device will be investigated as described below. Simultaneously, vital signs should be extracted for validation purposes.

3.1. Microsoft Kinect

Conceived firstly for computer gaming and home entertainment applications, RGB-D cameras, such as the Microsoft Kinect are sensing systems that capture RGB images along with per-pixel depth information. This device has one RGB camera and one Infra-Red camera which functions with an Infra-Red emitter.

3.1.1. RGB Video

Microsoft Kinect V1 is capable of recording RGB video at a medium-resolution (640x480 pixels) and 30 frames per second.

3.1.2. IR video

The IR camera on this device records video with a resolution of 320x240 pixels at 30fps. It captures the light which is emitted by the Infra-red emitter and partially reflected by the subject. Therefore, the wavelength of the light captured is restricted to approximately 830 nm (i.e. the wavelength of the light emitted).

4. Acquisition Factors

4.1. Imaging Time

Videos of approximately 5 minutes will be recorded for each subject/incubator with one still devices.

4.2. Camera Mount

Both cameras will be mounted on a tripod or rig at an appropriate distance from the incubator (~20cm) and at the same height as the new-born, so that the system is perpendicular to the incubator glass and hence reducing glares and distortions.

4.3. Camera Positioning

The camera should be positioned facing one of the lateral walls of the incubator and the field of view of the image should be such that the infant fills the frame but is not cropped by it. It is essential that the neonate’s face isn’t obstructed and can clearly be seen in its entirety.

4.4. Image Integrity

All images will be assigned a unique patient identifier to preserve confidentiality, in accordance with data protection rules.

5. Appendix

5.1. Specifications of Microsoft Kinect V1

Resolution RGB: 640 x 480 pixels

Resolution Near IR: 320 x 240 pixels

Depth Type Sensor: Structured light

Framerate: 30fps

Field of view (FOV) RGB: 62º x 48.6º

Field of view (FOV) IR: 57º x 43º

Depth-Range: 40cm (Near Mode)- 4m

Data connection: USB 2.0

5.2. Specifications of Microsoft Kinect V2

Resolution RGB:

Resolution Near IR:

Depth Type Sensor: Time of Flight

Framerate: 30 fps

Field of view (FOV) RGB: 84.1º x 53.8º

Field of view (FOV) IR: 70º x 60º

Depth Range: 50cm- 4.5m

Connection: USB 3.0

References

[1] The American College of Obstetricians and Gynecologists and The American Academy of Pediatrics. *Giudelines for Perinatal Care 7th Edition*. 2012. URL: http://www.jointcommission.org/perinatal{_}care/default.aspx?print=y.

[2] Janet Tucker and William McGuire. Epidemiology of preterm birth. *ABC of preterm birth*, 95(1), 2009.

[3] Li Liu, Shefali Oza, Dan Hogan, Yue Chu, Jamie Perin, Jun Zhu, Joy E. Lawn, Simon Cousens, Colin Mathers, and Robert E. Black. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, 388(10063):3027–3035, 2016. URL: http://dx.doi.org/10.1016/S0140-6736(16)31593-8, doi:10.1016/S0140-6736(16)31593-8.

[4] Rebecca B Russell, Nancy S Green, Claudia A Steiner, Susan Meikle, Jennifer L Howse, Karalee Poschman, Todd Dias, Lisa Potetz, Michael J Davidoff, Karla Damus, et al. Cost of hospitalization for preterm and low birth weight infants in the united states. *Pediatrics*, 120(1):e1–e9, 2007.

[5] Gillian Lim, Jacinth Tracey, Nicole Boom, Sunita Karmakar, Joy Wang, Jean-Marie Berthelot, and Caroline Heick. Hospital costs for preterm and small-for-gestational age babies in canada. *Birth*, 750(999):1–000, 2009.

[6] Hannah Blencowe, Simon Cousens, Mikkel Z. Oestergaard, Doris Chou, Ann Beth Moller, Rajesh Narwal, Alma Adler, Claudia Vera Garcia, Sarah Rohde, Lale Say, and Joy E. Lawn. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *The Lancet*, 379(9832):2162–2172, 2012. URL: http://dx.doi.org/10.1016/S0140-6736(12)60820-4, doi:10.1016/S0140-6736(12)60820-4.

[7] La Vone E. Simmons, Craig E. Rubens, Gary L. Darmstadt, and Michael G. Gravett. Preventing Preterm Birth and Neonatal Mortality: Exploring the Epidemiology, Causes, and Interventions. *Seminars in Perinatology*, 34(6):408–415, 2010. URL: http://dx.doi.org/10.1053/j.semperi.2010.09.005, doi:10.1053/j.semperi.2010.09.005.

[8] Marilee C Allen, Pamela K Donohue, and Amy E Dusman. The limit of viability–neonatal outcome of infants born at 22 to 25 weeks' gestation. *New England Journal of Medicine*, 329(22):1597–1601, 1993.

[9] Ying Dong and Jia-Lin Yu. An overview of morbidity, mortality and long-term outcome of late preterm birth. *World Journal of Pediatrics*, 7(3):199, 2011.

[10] Nina B. Kyrklund-Blomberg, Fredrik Granath, and Sven Cnattingius. Maternal smoking and causes of very preterm birth. *Acta Obstetricia et Gynecologica Scandinavica*, 84(6):572–577, 2005. doi:10.1111/j.0001-6349.2005.00848.x.

[11] Richard E Behrman and Adrienne Stith Butler. *Causes of Preterm Birth*. 2007.

[12] B. M. Mercer, R. L. Goldenberg, A. H. Moawad, P. J. Meis, J. D. Ianis, A. F. Das, S. N. Caritis, M. Miodovnik, M. K. Menard, G. R. Thurnau, M. P. Dombrowski, J. M. Roberts, and D. McNellis. The Preterm Prediction Study: Effect of gestational age and cause of preterm birth on subsequent obstetric outcome. *American Journal of Obstetrics and Gynecology*, 181(5 I):1216–1221, 1999. doi:10.1016/S0002-9378(99)70111-0.

[13] Saroj Saigal and Lex W. Doyle. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, 371(9608):261–269, 2008. doi:10.1016/S0140-6736(08)60136-1.

[14] Craig Lockwood, R N Bn, Graddipnsc Clinnurs, Tamara Page, R N Bn, Hyperbaricnurscert Graddipnsc, and Highdep Mnsc. Vital signs. pages 207–230, 2004.

[15] David Evans, Brent Hodgkinson, and Judith Berry. Vital signs in hospital patients: A systematic review. *International Journal of Nursing Studies*, 38(6):643–650, 2001. doi:10.1016/S0020-7489(00)00119-X.

[16] Maureen Lynch. Pain as the fifth vital sign. *Journal of Infusion Nursing*, 24(2):85–94, 2001.

[17] M Balakrishnan Malarvili and Mostefa Mesbah. Newborn seizure detection based on heart rate variability. *IEEE transactions on biomedical engineering*, 56(11):2594–2603, 2009.

[18] Susannah Fleming, Matthew Thompson, Richard Stevens, Carl Heneghan, Annette Plüddemann, Ian MacOnochie, Lionel Tarassenko, and David Mant. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: A systematic review of observational studies. *The Lancet*, 377(9770):1011–1018, 2011. URL: http://dx.doi.org/10.1016/S0140-6736(10)62226-X, doi:10.1016/S0140-6736(10)62226-X.

[19] D Leduc, S Woods, and Community Paediatrics Committee. Temperature measurement in paediatrics. *Paediatrics Child Health*, 5(5):273–276, 01 2000. URL: https://doi.org/10.1093/pch/5.5.273, arXiv:http://oup.prod.sis.lan, doi:10.1093/pch/5.5.273.

[20] Anna Lubkowska, Sławomir Szymański, and Monika Chudecka. Surface body temperature of full-term healthy newborns immediately after Birth—Pilot study. *International Journal of Environmental Research and Public Health*, 16(8), 2019. doi:10.3390/ijerph16081312.

[21] Sotirios Fouzas, Kostas N. Priftis, and Michael B. Anthracopoulos. Pulse oximetry in pediatric practice. *Pediatrics*, 128(4):740–752, 2011. doi:10.1542/peds.2011-0271.

[22] Carolyn Houska Lund and Joseph A Tucker. Adhesion and newborn skin. *Neonatal Skin: Structure and Function. 2nd ed. New York, NY: Marcel Dekker*, pages 299–324, 2003.

[23] Abbas AlZubaidi, Yahya Ethawi, Georg Schmölzer, Sherif Sherif, Michael Narvey, and Molly Seshia. Review of biomedical applications of contactless imaging of neonates using infrared thermography and beyond. *Methods and protocols*, 1(4):39, 2018.

[24] Carina Barbosa Pereira, Xinchi Yu, Tom Goos, Irwin Reiss, Thorsten Orlikowsky, Konrad Heimann, Boudewijn Venema, Vladimir Blazek, Steffen Leonhardt, and Daniel Teichmann. Noncontact monitoring of respiratory rate in newborn infants using thermal imaging. *IEEE Transactions on Biomedical Engineering*, 66(4):1105–1114, 2018.

[25] Wouter F.C. Hoffman and Daniël Lakens. Public benchmark dataset for testing rppg algorithm performance. 2019. doi:https://doi.org/10.4121/uuid:2ac74fbd-2276-44ad-aff1-2f68972b7b51.

[26] Sarah J Kilpatrick, Lu-Ann Papile, George A Macones, et al. *Guidelines for perinatal care*. Am Acad Pediatrics, 2017.

[27] Irfan Ahmad, Dan Nemet, Alon Eliakim, Robin Koeppel, Donna Grochow, Maria Coussens, Susan Gallitto, Julia Rich, Andria Pontello, Szu-Yun Leu, et al. Body composition and its components in preterm and term newborns: A cross-sectional, multimodal investigation. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 22(1):69–75, 2010.

[28] Susan Blackburn. Environmental impact of the nicu on developmental outcomes. *Journal of pediatric nursing*, 13(5):279–289, 1998.

[29] Timothy JM Moss. Respiratory consequences of preterm birth. *Clinical and Experimental Pharmacology and Physiology*, 33(3):280–284, 2006.

[30] Ashish Arunkumar Sharma, Roger Jen, Alison Butler, and Pascal M Lavoie. The developing human preterm neonatal immune system: a case for more research in this area. *Clinical Immunology*, 145(1):61–68, 2012.

[31] Jacqueline M Melville and Timothy JM Moss. The immune consequences of preterm birth. *Frontiers in neuroscience*, 7:79, 2013.

[32] Robert D White. The newborn intensive care unit environment of care: how we got here, where we're headed, and why. In *Seminars in perinatology*, volume 35, pages 2–7. Elsevier, 2011.

[33] Edward F Bell, Marie R Weinstein, and William Oh. Heat balance in premature infants: comparative effects of convectively heated incubator and radiant warmer, with and without plastic heat shield. *The journal of pediatrics*, 96(3):460–465, 1980.

[34] Robin B Knobel. Thermal stability of the premature infant in neonatal intensive care. *Newborn and Infant Nursing Reviews*, 14(2):72–76, 2014.

[35] William A Silverman, John W Fertig, and Agnes P Berger. The influence of the thermal environment upon the survival of newly born premature infants. *Pediatrics*, 22(5):876–886, 1958.

[36] Michael H LeBlanc. Relative efficacy of an incubator and an open warmer in producing thermoneutrality for the small premature infant. *Pediatrics*, 69(4):439–445, 1982.

[37] Kimio Yashiro, Forrest H Adams, George C Emmanouilides, and M Ray Mickey. Preliminary studies on the thermal environment of low-birth-weight infants. *The Journal of pediatrics*, 82(6):991–994, 1973.

[38] C Deguines, P Décima, A Pelletier, L Dégrugilliers, L Ghyselen, and P Tourneux. Variations in incubator temperature and humidity management: a survey of current practice. *Acta paediatrica (Oslo, Norway: 1992)*, 101(3):230–235, 2012.

[39] Karen A Thomas and Annie Uran. How the nicu environment sounds to a preterm infant: update. *MCN: The American Journal of Maternal/Child Nursing*, 32(4):250–253, 2007.

[40] Alex Robertson, Celeste Cooper-Peel, and Paul Vos. Sound transmission into incubators in the neonatal intensive care unit. *Journal of Perinatology*, 19(7):494–497, 1999.

[41] Edward F Bell and Gladys R Rios. Air versus skin temperature servocontrol of infant incubators. *The journal of pediatrics*, 103(6):954–959, 1983.

[42] Colin J Morley, Peter G Davis, Lex W Doyle, Luc P Brion, Jean-Michel Hascoet, and John B Carlin. Nasal cpap or intubation at birth for very preterm infants. *New England Journal of Medicine*, 358(7):700–708, 2008.

[43] Ole K Hejlesen, Simon Lebech Cichosz, Steffen Vangsgaard, Mikkel Frank Andresen, and Lars Peter Madsen. Clinical implications of a quality assessment of transcutaneous co2 monitoring in preterm infants in neonatal intensive care. *Stud Health Technol Inform*, 150:490–494, 2009.

[44] Patrick Eberhard. The design, use, and results of transcutaneous carbon dioxide analysis: current and future directions. *Anesthesia & Analgesia*, 105(6):S48–S52, 2007.

[45] Manon Ranger, C Céleste Johnston, and KJS Anand. Current controversies regarding pain assessment in neonates. In *Seminars in perinatology*, volume 31, pages 283–288. Elsevier, 2007.

[46] C Celeste Johnston, Bonnie J Stevens, Fang Yang, and Linda Horton. Differential response to pain by very premature neonates. *Pain*, 61(3):471–479, 1995.

[47] Ana Maria Gallo. The fifth vital sign: implementation of the Neonatal Infant Pain Scale. *Journal of obstetric, gynecologic, and neonatal nursing : JOGNN / NAACOG*, 32(2):199–206, 2003. doi:10.1177/0884217503251745.

[48] Diane Hudson-Barr, Beverly Capper-Michel, Sally Lambert, Tonya Mizell Palermo, Kristen Morbeto, and Stephanie Lombardo. Validation of the pain assessment in neonates (pain) scale with the neonatal infant pain scale (nips). *Neonatal Network*, 21(6):15–22, 2002.

[49] Daiva Bieri, Robert A Reeve, G David Champion, Louise Addicoat, and John B Ziegler. The faces pain scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. *Pain*, 41(2):139–150, 1990.

[50] T Debillon, V Zupan, N Ravault, JF Magny, and M Dehan. Development and initial validation of the edin scale, a new tool for assessing prolonged pain in preterm infants. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 85(1):F36–F41, 2001.

[51] Arindam Sikdar, Santosh Kumar Behera, and Debi Prosad Dogra. Computer-vision-guided human pulse rate estimation: a review. *IEEE reviews in biomedical engineering*, 9:91–105, 2016.

[52] Ming Yang, Qiong Liu, Thea Turner, and Ying Wu. Vital sign estimation from passive thermal video. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[53] Jin Fei and Ioannis Pavlidis. Analysis of breathing air flow patterns in thermal imaging. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 946–952. IEEE, 2006.

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

[55] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007.

[56] Philipp V Rouast, Marc TP Adam, Raymond Chiong, David Cornforth, and Ewa Lux. Remote heart rate measurement using low-cost rgb face video: a technical literature review. *Frontiers of Computer Science*, 12(5):858–872, 2018.

[57] H Emrah Tasli, Amogh Gudi, and Marten den Uyl. Remote ppg based vital sign measurement using adaptive facial regions. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1410–1414. IEEE, 2014.

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

[59] Alberto Fernández, Juan Luis Carús, Rubén Usamentiaga, Eduardo Alvarez, and Rubén Casado. Unobtrusive health monitoring system using video-based physiological information and activity measurements. In *2015 International Conference on Computer, Information and Telecommunication Systems (CITS)*, pages 1–5. IEEE, 2015.

[60] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[61] Michal Uricár, Vojtech Franc, and Václav Hlavác. Detector of facial landmarks learned by the structured output svm. In *VISAPP (1)*, pages 547–556, 2012.

[62] Gerard De Haan and Arno Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.

[63] Wenjin Wang, Sander Stuijk, and Gerard De Haan. Exploiting spatial redundancy of image sensor for motion robust rppg. *IEEE transactions on Biomedical Engineering*, 62(2):415–425, 2014.

[64] Min-Che Li and Yuan-Hsiang Lin. A real-time non-contact pulse rate detector based on smartphone. In *2015 International Symposium on Next-Generation Electronics (ISNE)*, pages 1–3. IEEE, 2015.

[65] Duc Nhan Tran, Hyukzae Lee, and Changick Kim. A robust real time system for remote heart rate measurement via camera. In *2015 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2015.

[66] Bin Han, Kamen Ivanov, Lei Wang, and Yan Yan. Exploration of the optimal skin-camera distance for facial photoplethysmographic imaging measurement using cameras of different types. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, pages 186–189, 2015.

[67] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)*, 31(4):1–8, 2012.

[68] Ramin Irani, Kamal Nasrollahi, and Thomas B Moeslund. Improved pulse detection from head motions using dct. In *2014 international conference on computer vision theory and applications (VISAPP)*, volume 3, pages 118–124. IEEE, 2014.

[69] Jean-Yves Bouguet et al. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation*, 5(1-10):4, 2001.

[70] Luca Iozza, Jesús Lázaro, Luca Cerina, Davide Silvestri, Luca Mainardi, Pablo Laguna, and Eduardo Gil. Monitoring breathing rate by fusing the physiological impact of respiration on video-photoplethysmogram with head movements. *Physiological measurement*, 40(9):094002, 2019.

[71] Lorenzo Scalise, Natascia Bernacchia, Ilaria Ercoli, and Paolo Marchionni. Heart rate measurement in neonatal patients using a webcamera. In *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*, pages 1–4. IEEE, 2012.

[72] Lonneke AM Aarts, Vincent Jeanne, John P Cleary, C Lieber, J Stuart Nelson, Sidarto Bambang Oetomo, and Wim Verkruysse. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study. *Early human development*, 89(12):943–948, 2013.

[73] Shakith Fernando, Wenjin Wang, Ihor Kirenko, Gerard de Haan, Sidarto Bambang Oetomo, Henk Corporaal, and Jan van Dalfsen. Feasibility of contactless pulse rate monitoring of neonates using google glass. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, pages 198–201, 2015.

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

[75] Luca Antognoli, Paolo Marchionni, Susanna Spinsante, Stefano Nobile, Virgilio Paolo Carnielli, and Lorenzo Scalise. Enanced video heart rate and respiratory rate evaluation: standard multiparameter monitor vs clinical confrontation in newborn patients. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–5. IEEE, 2019.

[76] Ce Liu, Antonio Torralba, William T Freeman, Frédo Durand, and Edward H Adelson. Motion magnification. *ACM transactions on graphics (TOG)*, 24(3):519–526, 2005.

[77] Jue Wang, Steven M Drucker, Maneesh Agrawala, and Michael F Cohen. The cartoon animation filter. *ACM Transactions on Graphics (TOG)*, 25(3):1169–1173, 2006.

[78] Le Liu, Le Lu, Jingjing Luo, Jun Zhang, and Xiuhong Chen. Enhanced eulerian video magnification. In *2014 7th International Congress on Image and Signal Processing*, pages 50–54. IEEE, 2014.

[79] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[80] S Brearley, CP Shearman, and MH Simms. Peripheral pulse palpation: an unreliable physical sign. *Annals of the Royal College of Surgeons of England*, 74(3):169, 1992.

[81] Stephen V Rice, Frank R Jenkins, and Thomas A Nartker. The fourth annual test of ocr accuracy. Technical report, Technical Report 95, 1995.

[82] R Grande, E Gutierrez, E Latorre, and F Arguelles. Physiological variations in the pigmentation of newborn infants. *Human biology*, pages 495–507, 1994.

[83] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. Forward-backward error: Automatic detection of tracking failures. In *2010 20th International Conference on Pattern Recognition*, pages 2756–2759. IEEE, 2010.

[84] Ville Lehtola, Heikki Huttunen, Francois Christophe, and Tommi Mikkonen. Evaluation of visual tracking algorithms for embedded devices. In *Scandinavian Conference on Image Analysis*, pages 88–97. Springer, 2017.

[85] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[86] Christos Sagonas, Epameinondas Antonakos, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: Database and results. *Image and vision computing*, 47:3–18, 2016.

[87] Hamed Monkaresi, Rafael A Calvo, and Hong Yan. A machine learning approach to improve contactless heart rate monitoring using a webcam. *IEEE journal of biomedical and health informatics*, 18(4):1153–1160, 2013.

[88] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[89] Litong Feng, Lai-Man Po, Xuyuan Xu, and Yuming Li. Motion artifacts suppression for remote imaging photoplethysmography. In *2014 19th International Conference on Digital Signal Processing*, pages 18–23. IEEE, 2014.

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

[91] Philip E Bickler, John R Feiner, and John W Severinghaus. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 102(4):715–719, 2005.

[92] John R Feiner, John W Severinghaus, and Philip E Bickler. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. *Anesthesia & Analgesia*, 105(6):S18–S23, 2007.

[93] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Continuous heart rate measurement from face: A robust rppg approach with distribution learning. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 642–650. IEEE, 2017.

[94] Peter Burt and Edward Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.

[95] U Rajendra Acharya, K Paul Joseph, Natarajan Kannathal, Choo Min Lim, and Jasjit S Suri. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006.

[96] Fumihiko Yasuma and Jun-ichiro Hayano. Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? *Chest*, 125(2):683–690, 2004.

[97] Mingliang Chen, Qiang Zhu, Harrison Zhang, Min Wu, and Quanzeng Wang. Respiratory rate estimation from face videos. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 1–4. IEEE, 2019.

[98] Friedrich Pukelsheim. The three sigma rule. *The American Statistician*, 48(2):88–91, 1994.

[99] Nicholas R Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and space science*, 39(2):447–462, 1976.

[100] George B Moody, Roger G Mark, Marjorie A Bump, Joseph S Weinstein, Aaron D Berman, Joseph E Mietus, and Ary L Goldberger. Clinical validation of the ecg-derived respiration (edr) technique. *Computers in cardiology*, 13(1):507–510, 1986.