**Biomedical Engineering Program**

Doctoral thesis

Contributions to heart rate variability and respiratory rate analysis through video imaging techniques

Author: Angel Melchor Rodríguez

Thesis advisor: Juan José Ramos Castro
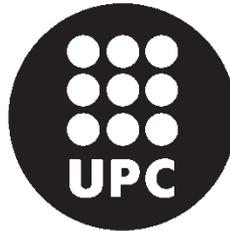
Biomedical Engineering Program

Contributions to heart rate variability and respiratory rate analysis through video imaging techniques

Author: Angel Melchor Rodríguez

Thesis advisor: Juan José Ramos Castro

Agradecimientos

Estas palabras de agradecimiento van dirigidas a todas las personas que, de una forma u otra, han sido parte fundamental para la realización de este proyecto.

Primeramente quiero agradecer a mis padres por apoyarme siempre, desde que era un niño, en los objetivos que me he propuesto. La educación y el ejemplo que me han dado han sido muy importantes para afrontar cada meta y que, sin ello, no hubiera sido posible. Ha sido muy difícil estar sin ustedes todos estos años en la distancia.

A mi esposa Elisabeth, por estar a mi lado cada día durante estos últimos años, por escucharme siempre, por tu comprensión, por tu apoyo y sobre todo por darme la dicha de ser padre por primera vez. Pablo y tú son mi gran motivación de cada día.

A Juan Ramos, mi director de tesis, por ser parte muy importante desde el inicio de este proyecto. Gracias por darme la oportunidad de trabajar juntos, por guiarme en cada paso de esta investigación y por todo el aprendizaje adquirido durante estos años. Muchas gracias Juan.

También quiero agradecer a todas las personas con las que compartí muchos momentos en esta universidad y, particularmente, a aquellas que trabajan en el

Departamento de Ingeniería Electrónica. A los compañeros de laboratorio, profesores, personal técnico y administrativo, así como también a las personas que participaron como voluntarios en esta investigación. A todos muchas gracias de corazón por su compañerismo, ayuda y colaboración en este trabajo.

Por último, también quiero dar las gracias al CONACYT por apoyarme para hacer realidad este proyecto. Espero que este tipo de iniciativas sigan existiendo para que cada vez más personas se formen académica y profesionalmente en lo que aspiran y, así, contribuir con trabajo y esfuerzo en el progreso de nuestro país.

Muchas gracias.

Abstract

The present doctoral thesis proposes a series of methods that have as a general objective the non-contact evaluation of heart rate variability (HRV) and respiratory rate (BR) through video imaging. The proposal of this technique has emerged in the last few years as an alternative to the traditional measurement systems, which assess these and other physiological parameters through direct contact with the person's body. Due to this condition, the traditional systems may cause some complications in people who present a sensitive or fragile skin, such as neonates or patients with burn injuries.

Furthermore, although the video imaging technique could be an excellent alternative, as in the cases mentioned above, it also aims at being a monitoring instrument for the general population in non-clinical environments. In fact, nowadays, the use of electronic devices such as the video cameras, smartphones, tablets and others, it starts being part of a healthy lifestyle. Thus, this promising technique, which could provide advantages such as the contactless measurement, portability, easy use and low-cost, it could be employed in the near future as part of the evaluation of the person's health status in everyday life.

On the other hand, despite the aforementioned advantages, the measurements obtained by this technique may be greatly affected by factors such as the movement of the person, lighting conditions, camera settings, among others. Hence, an analysis of the influence of various factors and conditions is carried out in this research study in order to obtain a better insight of the scope and limitations of the technique. For this purpose, a series of methods were

developed and implemented in which face detection and tracking algorithms are employed, as well as image and signal processing techniques.

The results obtained in each study are evaluated by using various statistical parameters and plots with the aim of measuring the agreement between the proposed methods and the reference systems. In general, the statistical analysis carried out shows a good level of agreement between the measurement systems. It is important to note the presence of an impact on the results, to a greater or lesser extent, by the factors and conditions analyzed in the different studies. In several cases, the results show a significant improvement in comparison with the data reported in related studies. By contrast, the results achieved in scenarios with a greater presence of artifacts show a decrease in the agreement of the measurements.

This contactless technique may eventually become an instrument to detect physical or psychological disorders in the future. Nevertheless, its use for this purpose will depend on the progress of the technique over the coming years, since it is still in the research and development phase. Therefore, more improvements are necessary to reach the reliability achieved by the current reference systems and, particularly, if its application in real-life scenarios is considered. The development of more robust algorithms is required in order to suppress, as much as possible, the contribution of artifacts present in a real environment. Consequently, the acquisition of measurements in several real-life scenarios, longer recordings, and the analysis of more factors that could influence the performance of the technique constitute some objectives for future work.

Abstract

En la presente tesis doctoral se proponen una serie de métodos los cuales tienen como objetivo general la evaluación sin contacto de la variabilidad de la frecuencia cardíaca y la frecuencia respiratoria mediante imagen de video. La propuesta de esta técnica ha surgido en los últimos años como una alternativa a los sistemas de medida tradicionales, los cuales evalúan estos y otros parámetros fisiológicos a través de contacto directo con el cuerpo de la persona. Debido a esta condición, los sistemas tradicionales podrían ocasionar algunas complicaciones en personas que presentan una piel sensible o frágil, tal como los recién nacidos o pacientes con lesiones por quemaduras.

Además, aunque la técnica de imagen de video podría ser una excelente alternativa, como en los casos mencionados previamente, ésta también tiene como objetivo ser un instrumento de monitorización para la población en general en entornos no clínicos. De hecho, hoy en día, el uso de dispositivos electrónicos tales como las cámaras de video, teléfonos inteligentes, tabletas y otros, empieza a ser parte del seguimiento de un estilo de vida saludable. Así, esta prometedora técnica, la cual podría proporcionar ventajas tales como la medición sin contacto, portabilidad, fácil uso y bajo costo, podría ser utilizada en un futuro cercano como parte de la evaluación del estado de salud de una persona en la vida cotidiana.

Por otra parte, a pesar de las ventajas mencionadas, las mediciones adquiridas mediante esta técnica pueden ser afectadas en gran medida por factores tales como el movimiento de la persona, las condiciones de iluminación, las ajustes de la cámara, entre otros. Por tanto, en este trabajo de investigación se

lleva a cabo un análisis de la influencia de varios factores y condiciones para obtener una mejor comprensión del alcance y las limitaciones de la técnica. Para este propósito, se han desarrollado e implementado una serie de métodos en los cuales se aplicaron algoritmos de detección facial y seguimiento, así como también técnicas de procesamiento de imágenes y señales.

Los resultados obtenidos en cada estudio son evaluados a través de distintos parámetros estadísticos y gráficos con el objetivo de medir el acuerdo entre los métodos propuestos y los sistemas de referencia. En general, el análisis estadístico llevado a cabo muestra un buen nivel de concordancia entre los sistemas de medida. Es importante tener en cuenta la presencia de un impacto en los resultados, en mayor o menor medida, por los factores y condiciones analizados en los distintos estudios. En varios casos, los resultados muestran una mejoría significativa en comparación con los datos reportados en estudios relacionados. Por el contrario, los resultados adquiridos en escenarios con una mayor presencia de artefactos muestran una disminución en el acuerdo de las medidas.

Esta técnica sin contacto podría eventualmente convertirse en un instrumento para detectar trastornos físicos o psicológicos en el futuro. No obstante, su uso para este objetivo dependerá del progreso de la técnica en los próximos años, ya que ésta se encuentra aún en fase de investigación y desarrollo. Por lo tanto, mayores mejoras son necesarias para alcanzar la fiabilidad que se consigue con los sistemas de referencia actuales y, particularmente, si se plantea su aplicación en escenarios reales. El desarrollo de algoritmos más robustos es requerido a fin de suprimir, tanto como sea posible, la contribución de artefactos presentes en un entorno real. Por consiguiente, la adquisición de medidas en varios escenarios de la vida real, registros de más larga duración, y el análisis de una mayor cantidad de factores que podrían influir al desempeño de la técnica constituyen algunos de los objetivos para trabajo futuro.

Contents

List of acronyms

ANS	Autonomic nervous system
BR	Breathing or respiratory rate
BRmean	Mean of all BR intervals
CVDs	Cardiovascular diseases
C-rPPG	Chrominance-based remote PPG
CWT	Continuous wavelet transform
ECG	Electrocardiogram
EMG	Electromyogram
EVM	Eulerian Video Magnification
FD	Frequency-domain
FEV1	Forced expiratory volume in the first second
FFT	Fast Fourier transform
fps	Frames per second
FVC	Forced vital capacity
Hb	Deoxyhemoglobin
HbO_2	Oxyhemoglobin
HCI	Human-computer interaction
HF	High-frequency component
HR	Heart rate
HRV	Heart rate variability
IBIs	Interbeat intervals

ICA	Independent Component Analysis
ICC	Intraclass correlation coefficient
iPPG	Video imaging photoplethysmography
kNN	k-nearest neighbor
LED	Light emitting diode
LF	Low-frequency component
LoA	Limits of agreement
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MBE	Mean bias error
MSAC	M-estimator sample consensus
N	Number of subjects
n.u.	Normalized units
NaN	Not-a-Number
NN	Sequence of RR intervals (normal-to-normal)
NN50	Number of pairs of adjacent NN intervals differing by more than 50 ms
NNmean	Mean of all NN intervals
NN-NN	Differences between adjacent NN measurements
PEF	Peak expiratory flow rate
pNN50	NN50 count divided by the total number of all NN intervals
PP	Pulse-to-pulse intervals
PPG	Photoplethysmography
PRV	Pulse rate variability
PSD	Power spectral density

PSNS	Parasympathetic nervous system
r	Pearson's correlation coefficient
RMSE	Root mean square error
RMSSD	Square root of the mean of the sum of the squares of differences between adjacent NN intervals
ROI	Region of interest
ROIe	ROI of the eyes
ROIf	ROI of the face
RR	Time interval between consecutive heartbeats
SDBR	Standard deviation of all BR intervals
SDNN	Standard deviation of all NN intervals
SDSD	Standard deviation of differences between adjacent NN intervals
SIDS	Sudden infant death syndrome
SNR	Signal-to-noise ratio
SNS	Sympathetic nervous system
TD	Time-domain
ULF	Ultra-low-frequency component
VLF	Very-low-frequency component
VT	Tidal volume
WHO	World Health Organization
ρ_c	Concordance correlation coefficient

Chapter 1

Introduction

1.1 Introduction

The measurement of physiological parameters is an important part of the evaluation of a person's health status in a medical examination. For instance, parameters such as heart rate (HR), heart rate variability (HRV), and breathing or respiratory rate (BR) can be used for assessing the physical and psychological condition of a person. The evaluation of these and other physiological parameters may provide relevant information on the health status, as well as a possible relation to disorders and diseases.

On the other hand, nowadays, the interest of the population in following a healthy lifestyle has been growing with the aim of preventing diseases and improving the quality of life. The use of technology has become part of this lifestyle as a tool for assessing the health status. Certainly, the progress in technology in the last years has facilitated that commercial electronic devices such as video cameras, smartphones, tablets and others can be used as instruments to measure physiological parameters in our daily lives.

1.2   Problem statement

The measurement of physiological parameters such as HR, HRV, and BR by using traditional measurement systems is performed in direct contact with the body, for example, through the use of transducers and electrodes. Nevertheless, this monitoring condition may cause some complications in people who present a sensitive or fragile skin, such as neonates or patients with burn injuries. Hence, in cases like these, the patient monitoring without physical contact could be an excellent alternative.

Furthermore, the interest in measuring physiological parameters in non-clinical environments by using commercial electronic devices has increased in the last few years. In this respect, the use of a video camera as a measuring instrument has gained attraction. The proposal of the video imaging technique, which could provide advantages such as the contactless measurement, portability, easy use and low-cost, it emerged in recent years as an alternative to the traditional systems. However, on the other hand, the measurements obtained by this technique may be greatly affected by factors such as the movement of the person, lighting conditions, camera settings, among others.

1.3   Objectives

Because the video imaging technique is a recent field of study, an analysis of the influence of various factors and conditions is carried out in this research study, with the aim of obtaining a better insight of the scope and limitations of the technique. For this purpose, a series of methods are developed and implemented as detailed throughout the chapters of this thesis.

Based on the problem statement, several objectives are raised in this research study, as summarized below:

- Development of algorithms to measure physiological parameters such as HR, HRV, and BR by employing the video imaging technique. The corresponding measurements will be evaluated by using various statistical parameters and plots in order to measure the agreement between the proposed methods and the reference systems.

- Implementation and evaluation of image and signal processing techniques in order to improve the acquisition of signals by video imaging. These techniques should be aimed at improving the quality of the signals by focusing on the regions of the body that could provide more information on the physiological parameters and on the characteristics of the signals, respectively.

- Comparative analysis of video cameras and settings with the aim of assessing their performance to measure physiological parameters. Different commercial cameras and settings such as the video resolution and frame rate will be evaluated in order to improve the measurements achieved with the proposed methods.

- Implementation and evaluation of face detection and tracking algorithms in order to measure physiological parameters both in stationary and motion conditions. The application of these algorithms would enable the automatic detection of the regions and reference marks of interest along the video sequences and, particularly, in the motion scenarios.

- A real-time implementation of the developed algorithms with the aim of performing a quick analysis of the physiological parameters and, also, the feasibility of real-time monitoring of the signals under influences such as the motion and the lighting conditions.

- Comparative analysis of natural and artificial lights in order to assess their influence on the measurements. Some of the most common light sources will be evaluated to find out which light option achieves the best results among them.

- Implementation and evaluation of the developed algorithms in order to be applied in a real-life scenario. The application of the video imaging technique in such scenarios could provide the possibility in the future of the non-contact measurement of physiological parameters in a real-life environment.

Chapter 2

Theoretical framework

2.1   Official statistics of cardiovascular diseases

Nowadays, the mortality statistics show that cardiovascular diseases (CVDs), which are defined as disorders of the heart and blood vessels, are the leading cause of death in the world. According to recent data reported by the World Health Organization (WHO), 15 million of deaths occurred in 2015, 8.76 million of which were from ischemic heart diseases and 6.24 million from strokes [1]. These diseases were the leading causes of death in the 2000-2015 period (Fig. 2.1). It is estimated that around 23.4 million people will die from CVDs in 2030 and it is also expected that they remain as the leading cause of death.

It is well known that risk factors such as bad diet habits, a lack of physical activity, stress, and a high consumption of alcohol and tobacco increase the chances of developing a CVD. Thus, most of these diseases could be prevented if actions are taken over these risk factors following a healthy lifestyle.
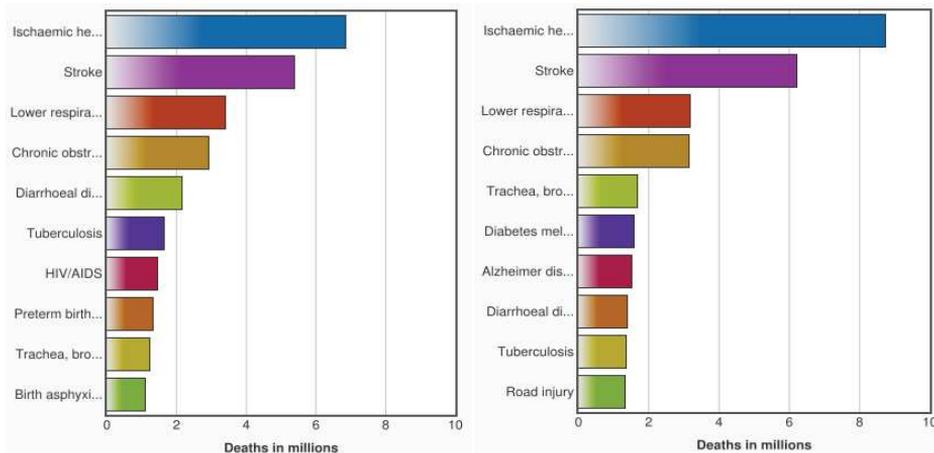
Fig. 2.1   The 10 leading causes of death in the world: (left) in 2000; (right) in 2015. (Source: "The top 10 causes of death". *World Health Organization (WHO)*, [1]).

The CVDs are the leading cause of death in countries with diverse economic status. Ischemic heart disease and stroke were the top leading causes of death in lower-middle-income, upper-middle-income, and high-income economies in 2015 (Fig. 2.2). Based on these data, the chances of developing a CVD is not related to the economic status of the countries, which makes them a global problem.

The awareness-raising of these statistics represents a great interest to the health system of the countries in decision-making and resources management. The high mortality rate from particular diseases can help to allocate more economic resources in order to combat them. The creation of preventive health

care programs may help to decrease the high mortality rates and the health costs caused by CVDs.
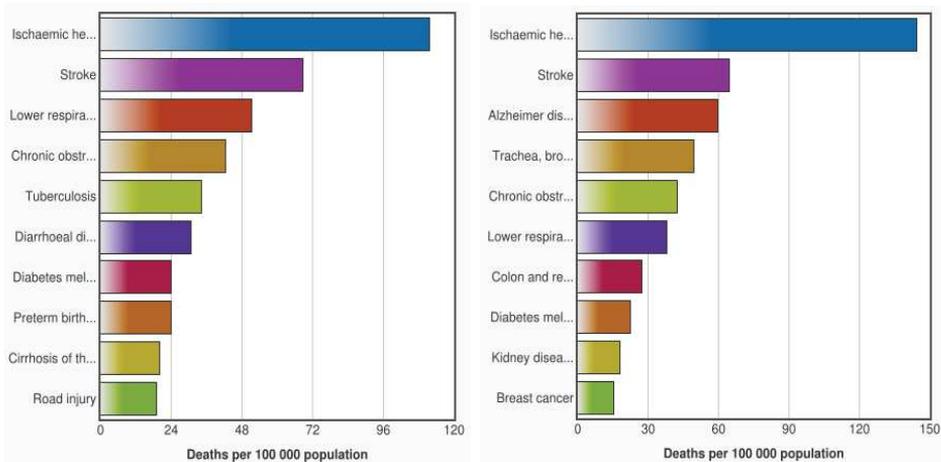
Fig. 2.2   The 10 leading causes of death in 2015: (left) lower-middle-income economies; (right) high-income economies. (Source: "The top 10 causes of death". *World Health Organization (WHO)*, [1]).

Being aware of the problem, the interest of the population in following a healthy lifestyle has notably increased in the last years. Therefore, research in novel methods to measure physiological parameters such as heart rate, HRV or respiratory rate is important, with the aim of performing an early detection of disorders and diseases.

2.2   Heart rate and HRV

Heart rate and HRV are two closely related parameters which are considered excellent indicators of the cardiovascular health status. Heart rate is a widely known parameter that represents the number of times the heart beats per minute,

7

and HRV is defined as the oscillation in the time interval between consecutive heartbeats (RR) [2] (Fig. 2.3). Variations in heart rate make it a nonstationary parameter whose regulation influences the HRV [3].

HRV is a physiological parameter that has gained importance due to its relations with the autonomic nervous system (ANS) and CVDs. It is a noninvasive measurement, reliable, and easy to obtain [3]. In general, a high HRV is related to a good health status, wellness, and optimal adaptation to physical activity performance [4]. Conversely, a low HRV is related to CVDs, poor fitness, and a non-adaptation to physical and psychological stress [5, 6].
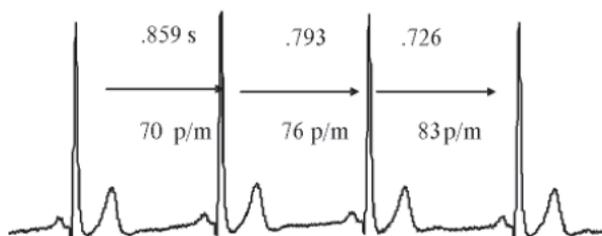
Fig. 2.3   Representation of HRV in seconds (s) and as instantaneous heart rate in beats per minute (bpm). (Source: G. Rodas et al., "Variabilidad de la frecuencia cardíaca: concepto, medidas y relación con aspectos clínicos (I)" [7]).

HRV has a close relation with the ANS, which is responsible for the heart activity through its branches: the sympathetic nervous system (SNS) and the parasympathetic nervous system (PSNS). These two branches have opposite actions in the heart rate activity. On the one hand, the SNS increases the heart rate during physical and psychological stress, and on the contrary, the PSNS decreases the heart rate in a resting state (Fig. 2.4) [8]. Therefore, the predominance of stimulation of these two divisions will depend on the physical and psychological status of the person.

HRV is inversely proportional in relation to the heart rate and workload, thus, an increase in them would cause a decrease in HRV. There are several factors that can influence the HRV analysis such as the thermoregulation, the respiratory system or the baroreceptor reflex. Moreover, there are other factors that affect the heart rate such as the age, the gender, the posture, the temperature, the alcohol consumption, among others [7].
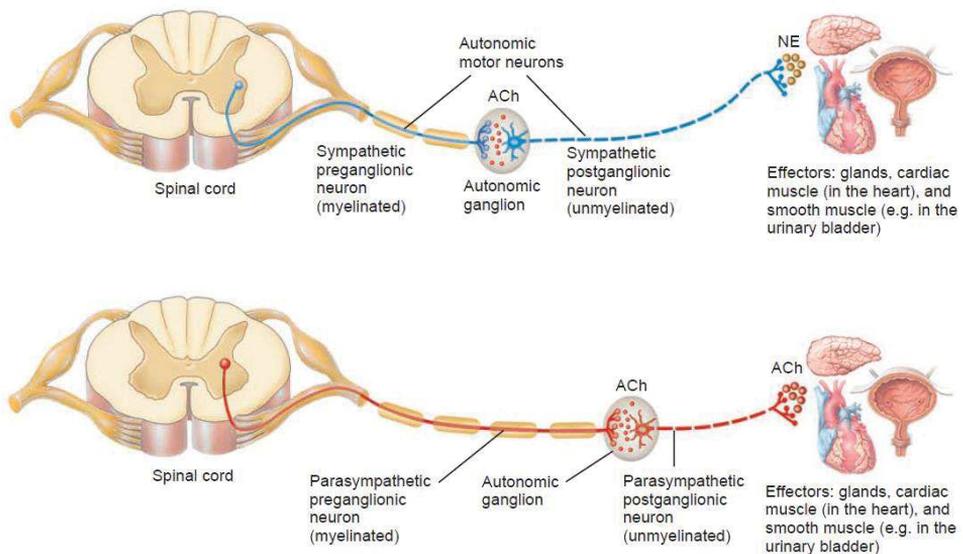
Fig. 2.4   Motor neuron pathways in the autonomic nervous system (ANS): (top) the sympathetic division is responsible for increasing the heart rate during physical and psychological stress; (bottom) the parasympathetic division is responsible for decreasing the heart rate in a resting state. (Source: Tortora, G. and Derrickson, B. *Principles of anatomy and physiology* [8]).

HRV analysis can provide relevant information on the cardiovascular and psychological health. For example, the HRV parameters can give information about the prediction or evolution of a CVD, as well as more knowledge of the adaptation to physical or psychological stress in a person.

2.2.1   Time-domain parameters of HRV

Time-domain parameters are possibly the simplest assessment to perform a HRV analysis. These components can be divided into two types: the parameters that are obtained directly from the sequence of RR intervals (or normal-to-normal (NN) intervals), and those that are calculated from the differences between adjacent RR intervals [2]. Some of the most common time-domain parameters are presented in Table 2.1.

Table 2.1   Time-domain parameters of HRV [2].

Parameter	Units	Description
NNmean	ms	Mean of all NN intervals.
SDNN	ms	Standard deviation of all NN intervals.
RMSSD	ms	Square root of the mean of the sum of the squares of differences between adjacent NN intervals.
SDSD	ms	Standard deviation of differences between adjacent NN intervals.
NN50	count	Number of pairs of adjacent NN intervals differing by more than 50 ms.
pNN50	%	NN50 count divided by the total number of all NN intervals.

The estimation of each parameter provides important information of a HRV analysis. In the case of the parameters obtained from the sequence of NN intervals, the NNmean parameter is the average of all NN intervals and the SDNN represents the total variability of the recording [7]. Due to the total variance of HRV increases with the record length [2], it is important to compare the SDNN values obtained from recordings of the same length.

Regarding the second type of time-domain parameters, the RMSSD component provides information on short-term variations of the NN intervals and, it is also used to analyze the influence of the PSNS over the cardiovascular system [7]. A high pNN50 value describes high spontaneous variations in heart rate [7], which is related to a high HRV. These last three parameters of short-term variations measure high-frequency changes in heart rate, therefore, they present a strong correlation (Fig. 2.5) [2].
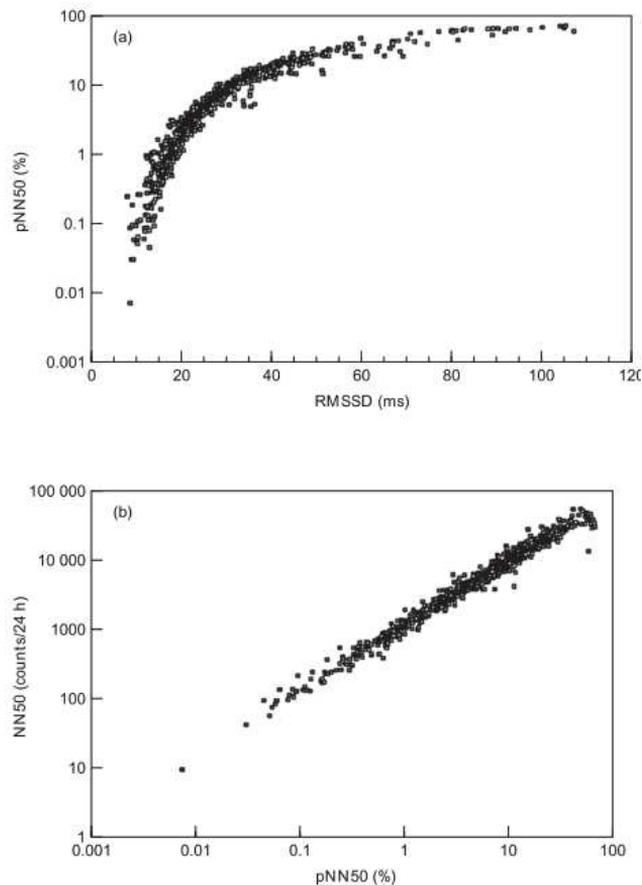
Fig. 2.5   HRV measurements from 857 nominal 24-h Holter records in survivors of acute myocardial infarction: (top) Relationship between RMSSD and pNN50; (bottom) relationship between pNN50 and NN50. (Source: Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use" [2]).

2.2.2 Frequency-domain parameters of HRV

The power spectral density (PSD) of HRV allows the analysis of power distribution in accordance with frequency. The calculation of the PSD can be obtained by parametric and non-parametric methods. The parametric methods present some advantages such as smoother spectral components, easy detection of the central frequency of the components, and accurate estimation of PSD even on a small number of samples in which the stationarity of the signal is assumed. On the other hand, the non-parametric methods have advantages such as the simplicity of the algorithm utilized (fast Fourier transform (FFT), in most of the cases) and the high processing speed [2].

The spectrum of HRV is mostly distributed between the 0.003 to 0.4 Hz range, in which different parameters can be distinguished as shown in Table 2.2.

Table 2.2   Frequency-domain parameters of HRV (analysis of short-term recordings) [2].

Parameter	Units	Description	Frequency range
VLF	ms^2	Power in very-low-frequency range.	≤ 0.04 Hz
LF	ms^2	Power in low-frequency range.	0.04 - 0.15 Hz
LF norm	n.u.	LF power in normalized units LF/(Total power - VLF) x 100	
HF	ms^2	Power in high-frequency range.	0.15 - 0.4 Hz
HF norm	n.u.	HF power in normalized units HF/(Total power - VLF) x 100	
LF/HF		Ratio LF [ms^2]/HF [ms^2]	

n.u.: normalized units.

These components provide relevant physiological information [7]:

- VLF (very-low-frequency): it represents hormonal, vasomotor, and thermoregulatory influences.

- LF (low-frequency): it mainly represents the influence of the SNS, which is the branch that controls an increase in heart rate, though, it could have influences of both divisions of the ANS. It also represents the baroreceptor activity, which has a frequency of 0.1 Hz approximately.

- HF (high-frequency): it represents the influence of the PSNS, which is the division that controls a decrease in heart rate. The changes in the respiratory rate have an effect on this component by modifying the HF peak.

- LF/HF ratio: it provides information about the balance between the SNS and PSNS divisions. Thus, it is possible to know the predominance of activity of the divisions of the ANS and, therefore, the interpretation of a low or high HRV of a person.

The VLF, LF and HF components are distinguished in short-term recordings (2 to 5 min) and, additionally, it is possible to identify an ultra-low-frequency component (ULF) in long-term recordings (24 h), which is below 0.003 Hz [2].

The VLF, LF, and HF can be measured in absolute values of power (ms^2) and the last two components also in normalized units (n.u.). The parameters in n.u. represent their relative value in proportion to the total power (minus the VLF component). The representation of LF and HF parameters in n.u. accentuates the balanced behavior of the two divisions of the ANS.

The effect of the units in a spectral analysis is represented in Fig. 2.6 using HRV data from a healthy subject at rest and 90º head-up tilt [2]. At rest, the LF and HF components were 310 ms^2 (48.95 n.u.) and 302 ms^2 (47.78 n.u.) respectively with an LF/HF ratio of 1.02. At 90º head-up tilt, the values were

308 ms^2 (75.96 n.u.) and 95 ms^2 (23.48 n.u.) respectively with an LF/HF ratio of 3.34. It is noted that the LF component had a slight decrease from rest to tilt in absolute values of power (from 310 ms^2 to 308 ms^2), and conversely, it had a notable increase in normalized units (from 48.95 n.u. to 75.96 n.u.).
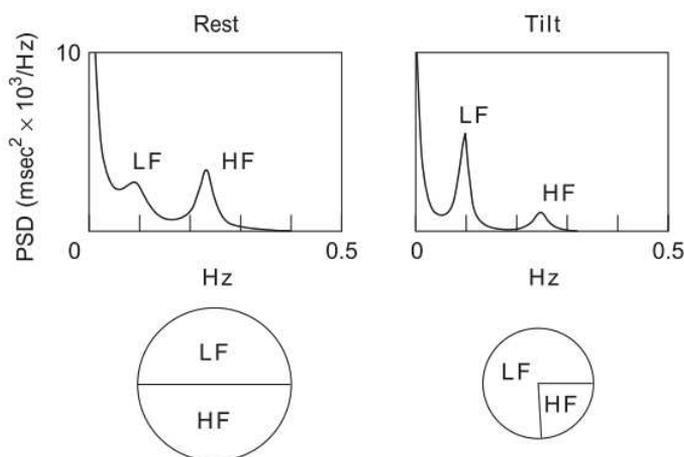
Fig. 2.6   Spectral analysis of HRV in a healthy subject at rest and 90º head-up tilt. (Source: Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use" [2]).

2.2.3   Technical requirements and recommendations

In order to perform an appropriate HRV analysis, it is important the accomplishment of technical requirements and recommendations established by the Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology [2]. According to the recommendations, it is necessary to choose a minimum sampling rate to record the ECG signal. An optimal range is established between 250-500 Hz or even higher. If the signal is recorded at a lower sampling rate (in any case $\geq$ 100 Hz), it is necessary to carry out an interpolation to refine the R wave fiducial point.

It is required the use of a well-tested algorithm to perform the detection of the R wave fiducial point in order to calculate the RR intervals. Some issues such as an incorrect detection, the presence of ectopic beats, arrhythmias, missing data or noise artifacts would generate errors in the RR tachogram (RR time series). In these cases, an appropriate detection and correction of outliers is an option to reduce the error in the HRV analysis.

In the frequency domain, an important recommendation is to check the frequency response of the applied filter in order not to affect the components of interest. Moreover, it is recommended 5 min recordings with the aim of standardizing short-term studies of HRV, unless the study has other objectives.

Traditionally, a HRV analysis has been carried out using the ECG signal, though some authors have investigated the possibility of measuring pulse rate variability (PRV) as a surrogate of HRV [9, 10]. In summary, these studies have shown that PRV is sufficiently accurate as a calculation of HRV, especially, in healthy subjects at rest. Moderate physical activity and mental stress are some factors that may decrease the agreement between these two measurements. A more intense physical activity such as walking or physical exercising could greatly affect the agreement due to motion artifacts.

2.3   Heart rate and HRV analysis by video imaging

In recent years, the evaluation of physiological parameters by video imaging has attracted great interest because it is a low-cost method that is contactless and easy to implement. Certainly, the technological advances in electronics and optics fields have allowed the possibility of using video cameras as monitoring devices of physiological parameters.

2.3.1 Functional principle

In the last few years, some studies have measured heart rate or HRV parameters using a video camera [11-14]. This technique may be known as video imaging photoplethysmography (iPPG) and focuses on the measurement of the small changes in skin color caused by blood perfusion. Diverse factors may influence this technique such as the anatomical and physiological characteristics of the person involved, the motion, the lighting conditions, the camera settings, among others.

Similarly to the photoplethysmography (PPG) method, the iPPG technique is based on the interaction of light through the tissues and the changes in blood perfusion. The tissues and the blood, and particularly the hemoglobin, absorb part of the light that passes through the skin and other parts can be scattered and reflected towards the skin surface [15, 16]. Consequently, the reflected light can be detected by the optical sensors of the cameras. Thus, mainly due to the pulsatile component of the arterial blood, which causes changes in light absorption, variations in reflected light occur which can be measured by the optical sensors.

Fig. 2.7 shows the absorption spectrum of oxyhemoglobin (HbO_2) and deoxyhemoglobin (Hb) within the visible region of the electromagnetic spectrum [15]. The wavelength range within the visible region is from 380 to 750 nm approximately in which the red (620-750 nm), green (495-570 nm), and blue (450-495 nm) colors are present. As noted in the absorption spectrum, the wavelengths corresponding to the green and blue colors present higher absorption levels of the Hb and HbO_2 in comparison with the red light. Also, the green and blue colors show larger differences of absorption between the Hb and HbO_2 at some specific wavelengths.
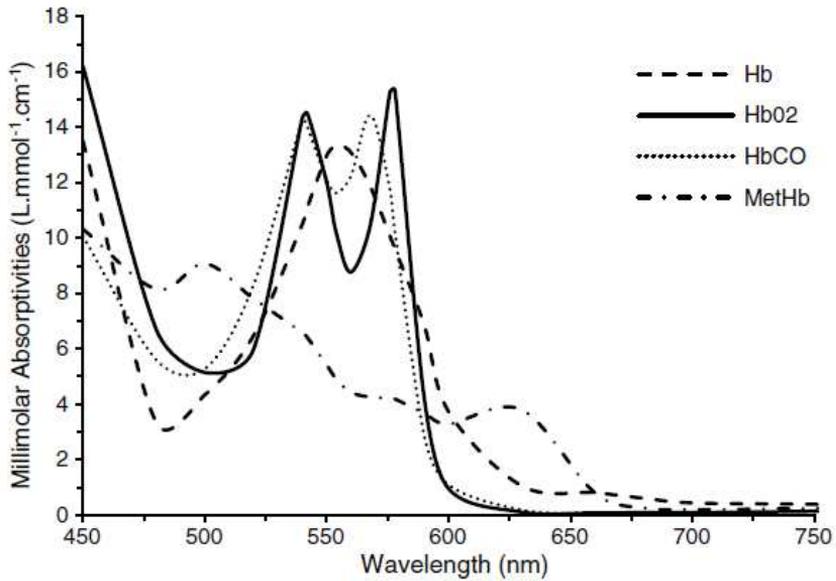
Fig. 2.7   Absorption spectrum of deoxyhemoglobin (Hb) and oxyhemoglobin (HbO$_2$) within the visible region of the electromagnetic spectrum. (Source: T. Lister et al., "Optical properties of human skin" [15]).

The penetration depth of light into the tissues is related to its wavelength, where longer wavelengths reach deeper penetrations into the skin than shorter ones [16]. For this reason, the green light has a minor penetration into the skin in comparison with the red wavelength. Thus, the signals obtained from the red light source may have significant part of their origin from deeper blood vessels, while the signals acquired from the green light could have a more superficial influence. Based on their results, Cui et al. states that the pulsatile response of the green light PPG is improved for lower blood fractional volumes in tissue in comparison with the red PPG. They concluded that the reduced penetration of the green light source results in a better signal-to-noise ratio (SNR), as well as the pulsatile blood volume changes can be more reliably detected with the use of light within the range of 510-590 nm, when compared with longer wavelengths.

Other authors have shown through a comparative analysis that the green light PPG achieved the best results for heart rate monitoring [17, 18]. Thus, these studies are some evidence that may explain why the green channel of the video gives the best iPPG signal in comparison with the other channels.

2.3.2 State of the art

Although this research area has attracted great interest, the number of studies that have measured HRV parameters has so far been very limited. One of the first and highly cited works in the field is the study presented by Poh et al. [19]. In this work, the method of blind source separation by Independent Component Analysis (ICA) was proposed to obtain the iPPG signals. The algorithm based on work by Viola and Jones [20] and Lienhart and Maydt [21] was used to detect the region of interest (ROI). HR was measured in 30-s moving windows with 1-s of overlap by applying the FFT. The study was conducted in stationary and motion conditions.

Along the same lines, Monkaresi et al. [22] proposed an algorithm for measuring HR in sitting still and naturalistic human-computer interaction (HCI) scenarios. The aim of the study was to improve the HR estimation by employing the Poh et al. method [19] in combination with machine learning techniques: linear regression and k-nearest neighbor (kNN). HR measurements were also extracted from 30-s moving windows with 1-s of overlap by a power spectral analysis.

Other works that also used the video imaging technique have measured the interbeat intervals (IBIs) or HRV parameters [23-32]. Poh et al. [23] presented a multiparameter physiological measurement work in which HR and HRV parameters were obtained by using the ICA method. Unlike in their previous work, HR measurements were estimated by calculating the mean of the IBIs.

The frequency-domain components were estimated by a PSD using the Lomb periodogram. Alternatively, Sun et al. [24] conducted a PRV analysis focused on the palm of the subject's hand. Since the measurements obtained from this body part are also affected by the motion, the study was performed in stationary conditions by using a cushion placed under the hand to minimize artifacts. However, a procedure of reduced frames by pixel averaging was carried out to attenuate small motion artifacts. The iPPG signals were obtained from each averaged position across the sequence of reduced frames. Subsequently, a technique based on wavelet transforms was employed to detect the pulse-to-pulse (PP) intervals.

McDuff et al. [25, 26] have proposed the use of a five band digital camera with the aim of evaluating alternate combinations of frequency bands that yield better results in the measurement of physiological parameters. Correlations for all combinations of the color channels were calculated in [25]. The measurements of both works were obtained at rest and under stress conditions. In another study, Moreno et al. [27] conducted a HRV analysis in supine and sitting postures with controlled illumination, synchronized breathing, and eyes closed. This work presented a cross-correlation analysis with the aim of finding the face areas on averaged frames that could provide more information on HR. Some of the most common time and frequency domain parameters of HRV were reported in the study. Alghoul et al. [28] presented a comparison between two approaches to measure HRV parameters from the face in stationary conditions. These approaches are based on the ICA and Eulerian Video Magnification (EVM) methods, respectively.

In addition to stationary conditions, some authors also have conducted the acquisition of the IBIs in motion scenarios. Bousefsaf et al. [29] proposed a motion-tolerant method to measure the instantaneous HR. This method employs a skin detection filter and the $u*$ component of the CIE $L*u*v*$ color space to

improve its robustness in presence of motion or illumination changes. A wavelet-based filtering is applied in order to remove noise components from the raw iPPG signals. The recordings were obtained at rest and with predefined head movements. The study of Kumar et al. [30] proposed a method of combining skin-color change signals from different regions of the face. This method uses a weighted average to improve the SNR in which the weights depend on the blood perfusion and the incident light intensity in the region. PP interval estimations were carried out in stationary, reading, watching video, and talking scenarios.

In another study, Antink et al. [31] performed a beat-to-beat estimation by using different signals and their fusion. The assessed signals were obtained from the changes of the skin color and the head motion, both by video and from a ballistocardiographic mat sensor, which were fused using a Bayesian approach. Evaluations of each signal and their fusion are presented in the paper. The trials performed in this study were: sitting still, reading without motion, and reading without further instructions.

Huang and Dung [32] proposed the application of the chrominance-based remote PPG (C-rPPG) method followed by a continuous wavelet transform (CWT)-based denoising technique. A data acquisition procedure was carried out before employing C-rPPG and CWT. The procedure included face and skin color detection, the computing of the averaged RGB values within the ROI, and the upsampling of the signals. The face tracking was performed by means of nose detection for purposes of robustness. HRV parameters were obtained in static postures and occasional/frequent motion. A summary of the measurement set-up, signal domain analysis, and results of the previously cited works has been compiled and presented in Table 2.3.

Table 2.3 Summary of the measurement setup, signal domain analysis, and results of reference works.

Ref.	Activity condition/method	Subj. (N)	Recording time	Resolution (fps)	SA	Overall results			
[19]	1 = sitting still 2 = natural motion (no large or rapid movements)	12	1 min	640x480 (15 fps)	FD			**r**	**RMSE (bpm)**
						HR	(1)	0.98*	2.29
							(2)	0.95*	4.63
[22]	1 = sitting still 2 = natural human-computer interaction	10	(1) 1 min (2) 20 min	640x480 (30 fps)	FD			**r**	**RMSE (bpm)**
						HR	(1)	0.99*	1.69
							(2)	0.93*	3.64
[23]	Sitting still in front of a laptop	12	1 min	640x480 (15 fps)	TD, FD†			**r**	**RMSE**
						HR (bpm)		1.00*	1.24
						LF (n.u.)		0.92*	12.3
						HF (n.u.)		0.92*	12.3
						LF/HF		0.88*	1.1
[24]	Resting conditions (palm of the subject's hand)	10	4 min	384x256 (200 fps)	TD, FD†			**r**	
						NNmean		0.998*	
						SDNN		0.874*	
						LF (n.u.)		0.971*	
						HF (n.u.)		0.978*	
						LF/HF		0.875*	
[25]	1 = sitting at rest 2 = sitting under stress	(1) 9 (2) 10	2 min	960x720 (30 fps)	TD, FD†			**r**	
								(1)	(2)
						HR (bpm)		1.00ˣ	1.00ˣ
						LF (n.u.)		0.87ˣ	0.97ˣ
						HF (n.u.)		0.87ˣ	0.97ˣ
						LF/HF		0.86ˣ	0.95ˣ
[26]	1 = sitting at rest 2 = sitting under stress	11	2 min	960x720 (30 fps)	TD			**MAPE (%)**	**MAE (s)**
						IBIs	(1)	3.11	0.027
							(2)	3.10	0.025
[27]	1 = supine position 2 = sitting position (controlled illumination, synchronized breathing and closed eyes)	(1) 12 (2) 8	5 min	640x480 (30 fps)	TD, FD†			**ρc**	
								(1)	(2)
						NNmean		0.9999	0.9999
						SDNN		0.9544	0.9108
						RMSSD		0.8398	0.5180
						pNN50		0.8635	0.5385
						LF (n.u.)		0.9752	0.7934
						HF (n.u.)		0.9498	0.7838
						LF/HF		0.8662	0.3186

Table 2.3   Continued

Ref.	Activity condition/method	Subj. (N)	Recording time	Resolution (fps)	SA	Overall results			
[28]	Sitting still in front of a camera (1 = ICA-based method 2 = EVM-based method)	12	2 min	720x480 (30 fps)	TD, FD[†]			r	
								(1)	(2)
						NNmean		0.999*	0.999[x]
						LF (n.u.)		0.8*	0.831[x]
						HF (n.u.)		0.84*	0.789[x]
						LF/HF		0.74*	0.256[x]
[29]	1 = sitting still and calm 2 = sitting with pre-defined head movements	12	35 s	320x240 (30 fps)	TD			r	RMSE (bpm)
						1	HRmean	1.00*	-
							IBIs	0.889*	1.97
						2	HRmean	1.00*	-
							IBIs	0.853*	2.33
[30]	1 = stationary 2 = reading 3 = watching video 4 = talking	5	80 s	1280x1024 (30 fps)	TD	RMSE (ms)			
							PRV		
						(1)	15.74		
						(2)	55.34		
						(3)	67.08		
						(4)	97.51		
[31]	1 = sitting still 2 = reading without motion 3 = reading without further instructions	4	2 min	800x600 (30 fps)	TD			MAE (ms)	
						IBIs	(1)	23.00	
							(2)	27.38	
							(3)	32.83	
[32]	1 = static 2 = static with makeup 3 = occasional motion 4 = frequent motion	(1) 4 (2) 3 (3) 3 (4) 2	1 min	640x480 (30 fps)	TD		MAE (ms)		
							SDNN	RMSSD	
						(1)	2.01	4.33	
						(2)	3.61	3.54	
						(3)	11.94	24.00	
						(4)	6.05	12.37	

N: number of subjects; fps: frames per second; SA: signal analysis; FD: frequency-domain; TD: time-domain (NN or PP time series); r: Pearson correlation coefficient; ρ_c: concordance correlation coefficient; *: p<0.001; n.u.: normalized units; [†]: only for the acquisition of frequency-domain parameters of HRV; PRV: pulse rate variability (time series between consecutive pulse beats). The MAE results from [31, 32] are calculated in this thesis to obtain a mean value of the individual results (note: the recordings of the occasional and frequent motion categories [32] were performed by the same subject). For comparison purposes, only were included the parameters that were calculated in this thesis.

2.4   Respiratory signal analysis by alternative methods

The interest of measuring physiological parameters using alternative methods also involves the respiratory signal analysis. In this respect, some authors have proposed several non-contact methods through the use of recording devices such as video [23-25, 29, 33-67], thermal/infrared [68-78] and depth imaging cameras [79-84], as well as other sensors [85-87] with the aim of measuring respiratory parameters. In these studies, several applications have been presented such as the monitoring of respiration in neonates in order to prevent the sudden infant death syndrome (SIDS) [44, 47, 51], as part of a non-contact vital sign monitoring system in a clinical environment [42, 54, 64], for sleep monitoring [77, 78], as a spirometry test [60], in telemedicine [40], for detection of vital signs in hazardous environments [66], etc.

Alternatively, other authors have acquired the respiratory signal from the ECG or PPG signal [88-99] or through the signals recorded with the built-in cameras of smartphones or tablets [100-105]. In these cases, the respiratory signal can be obtained because some components of the respiration such as the amplitude and frequency can be extracted from these recordings by performing an appropriate signal processing.

In the interest of our study, some works have proposed non-contact video camera-based methods to estimate the instantaneous and average respiratory rate and have reported statistical results of the measurements. Bousefsaf et al. [29] proposed a motion-tolerant method to measure the instantaneous breathing rate from the face area at rest and with predefined head movements. In this study, the respiratory signal was recovered from the instantaneous heart rate series. To achieve this, a linear interpolation was carried out on the beat-to-beat series and then a CWT was performed between 0.15 and 0.4 Hz, which is a frequency band commonly assessed in a HRV analysis. The respiratory signal was obtained

through a weighted reconstruction from the CWT and, subsequently, the instantaneous breathing rate was measured by computing the intervals between the respiratory cycles.

Alternatively, van Gastel et al. [53] presented a non-contact camera-based method to detect the instantaneous respiratory rate both in visible and dark lighting conditions. In both conditions, the subjects were in a sitting position and were asked to follow a breathing pattern that was displayed on a screen. The proposed algorithm by van Gastel et al. consisted of three processing steps: the tracking of a ROI (face) that is divided into subregions, a processing step in which weights are calculated for each subregion and the best are selected according to the SNR values of the pulse signals, and a scaling step for removing the influence of the pulse signal and thus only extract the respiratory signal.

Cobos and Abderrahim [58] proposed a method for the contactless measuring of heart and respiratory rates by employing the iPPG technique through wavelets. This method allows the analysis of variations within a ROI corresponding to the face and abdomen of the subject. In this study, the experiments were carried out in a sitting position in front of a computer with a built-in camera. The subjects were asked to follow an ascending controlled respiration from 14 to 16 breaths/min and, then, a descending pattern from 16 to 13 breaths/min. A manual selection of the ROI was performed in which the abdomen area was chosen to be analyzed by pixel averaging. The respiratory rate was measured by calculating the time interval between two peaks in the filtered signal.

In another study, Al-Naji and Chahl [57] presented a non-contact monitoring system using a digital camera to simultaneously obtain the cardiorespiratory signal from a group of people. The algorithm proposed by Al-Naji and Chahl is based on the skin color changes and head motion which are consequences of the

activities of the cardiovascular and respiratory system. The measurements were carried out in a stationary condition with as little movement as possible, and in a non-stationary scenario in which the subjects were allowed to move naturally, talking, blinking, and performing facial expressions. In general, the proposed algorithm includes the following processing methods: video magnification (color and head motion magnification based-methods), face detection, methods for removal of noise artifacts, frequency analysis and temporal filtering, and peak detection.

Moreover, some works that have proposed non-contact methods using a video camera have estimated other different respiratory parameters such as dynamic volume measurements [38], forced expiratory volume in the first second (FEV1), forced vital capacity (FVC), peak expiratory flow rate (PEF) [60] and tidal volume (VT) [62]. Although most of the previously cited works have estimated respiratory parameters in humans, a few of them have obtained their corresponding measurements from animals [55, 56].

All these studies represent a promising alternative technique in which the respiratory parameters can be measured by employing non-contact methods and using low-cost devices. The wide variety of potential applications of the technique encourages its implementation for measuring several respiratory parameters in the near future. To this end, novel methods are needed in order to obtain a higher accuracy in the measurement of these parameters such as that achieved by current reference systems.

2.5   Facial detection approaches

Nowadays, object detection algorithms are employed in many computer vision applications such as facial and activity recognition, information security,

automotive safety, surveillance, among others. The interest of applying these algorithms in our daily life activities has increased in the last few years in order to provide novel solutions to diverse problems. In this respect, several approaches have been proposed with the aim of detecting a human face on an image [106-109].

Since these algorithms are based on different methods, some authors have reviewed and classified them into categories according to their technical approach. Hjelmås and Low [106] classified the face detection methods as either feature-based or image-based algorithms. In summary, they stated that the feature-based approaches are appropriate for real-time systems in which the color and motion is present and, in the case of the image-based methods, these were considered more robust for detecting faces on gray-scale static images. In the study, a comparative assessment of the algorithms was not provided due to the lack of standardized tests.

Yang et al. [107] reviewed several techniques used to detect faces from a single intensity or color image, which were classified into four categories: knowledge-based methods (which encode information of a typical face), features invariant approaches (which look for structural features that remain in variations of pose, viewpoint or lighting conditions), template matching methods (which correlate an input image and stored patterns of faces), and appearance-based methods (which use models learned from a set of training images which constitute the variability of facial appearance). Yang et al. also declared the lack of consistency in the evaluation of the algorithms, which makes it difficult to conclude which face detection methods present the lowest error rates.

More recently, Zhang and Zhang [108] reviewed and evaluated the face detection algorithms proposed in the first decade of the 21st century. This evaluation showed that the state of the art achieved about 50-70% detection rate

with 0.5-3% of false positives. Among these methods, the Viola and Jones approach [20] was presented as the algorithm that had the most impact in the 2000's. This method is able to run in real-time [110] and suitable to be used in applications such as digital cameras and photo organization software. To achieve this, the work proposed by Viola and Jones introduced three contributions: the integral image, the AdaBoost-based algorithm, and the cascade classifier. These contributions and other information related to the algorithm are detailed in section 2.5.1.

Zafeiriou et al. [109] assessed several face detection methods in unconstrained conditions in a linked review to [108]. In this study, due to more recently developed methodologies, it was proposed to classify the algorithms into two major categories: rigid templates, which learned mainly by boosting-based methods or by using deep neural networks, and deformable models, which represent the face by its parts. The Viola and Jones approach and its variations [20, 110] is the main representative work of the boosting-based algorithms family, which, in turn, belongs to the category of rigid templates. A comparative analysis of performance is presented in this study, which includes a comparison of recent and older popular representative techniques such as the OpenCV version of the Viola and Jones approach [110, 111]. In this analysis, it is noted an increase in performance in the last few years attributed to contributions from the Viola and Jones algorithm and robust features [112].

The Viola and Jones algorithm is a well-known face detection approach that is able to run in real-time, characterized by processing images much faster and achieving high detection rates and, until today, it is widely used in several applications and research. Therefore, the implementation of this algorithm is proposed in this investigation to detect the ROI of the subjects in order to obtain the iPPG signals. But, as a drawback, the method is less effective in detecting faces in non-frontal poses because of the Haar features employed by the

algorithm [108]. For instance, the application of this approach may not be suitable for tracking the face of a subject in a real-life scenario. Hence, it is also proposed the application of the KLT algorithm [113-115] in conjunction with the Viola and Jones approach as a more robust face detection-tracking method.

2.5.1   Viola and Jones algorithm

Within the face detection approaches, the algorithm proposed by Viola and Jones [20] is a relevant and very commonly used method in computer vision software. It enables regions of the face to be detected on a video frame by using rectangular Haar-like features (Fig. 2.8). Different types of rectangle features are employed by the algorithm, which are classified according to the number of rectangles that compose them. These features can consist of two, three or four adjacent rectangles located on a horizontal or vertical position.
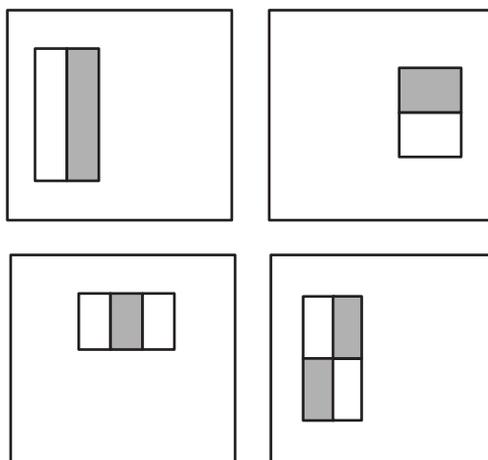
Fig. 2.8   Some examples of rectangle features: (top) two-rectangle features; (bottom-left) three-rectangle feature; (bottom-right) four-rectangle feature. (Source: P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features" [20]).

The features values are obtained from the difference between the sums of intensities of the pixels within the white and grey rectangles. For example, the value of a three-rectangle feature is the difference between the sum within the outside rectangles and the corresponding sum of the central rectangle. The employed detector of the algorithm has a base resolution of 24x24, which cause a very large number of rectangle features. Therefore, in order to perform a fast and efficient object detection, the Viola and Jones algorithm provides three main contributions: the integral image, the AdaBoost-based algorithm, and the cascade classifier.

- **Integral image:** because of the number of rectangle features is very large, which means a high computational cost, it was introduced the integral image (*ii*). By using this approach, it is possible to calculate the values of the rectangle features much faster. The integral image contains at the location (*x, y*) the sum of pixels both above and to the left of this location of the original image (*i*)

$$ii(x, y) = \sum_{x' \leq x, \ y' \leq y} i(x', y') \qquad \text{(Eq. 2.1)}$$

Thus, by using the values of the locations of the integral image, it is possible to obtain the sum of the pixel values within a specific region. For example, the sum of pixels of a rectangle *D* is calculated as *D=4+1-(2+3)*, in which the locations *1, 2, 3* and *4* are the sum of pixels of their corresponding rectangles (Fig. 2.9).
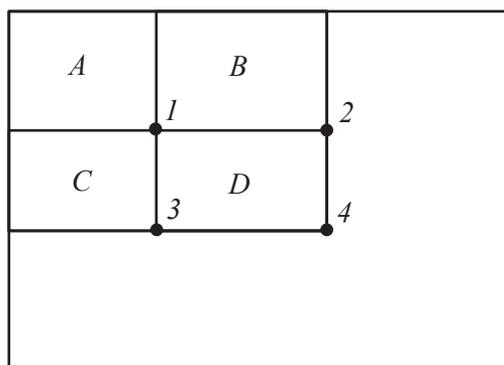
Fig. 2.9   The sum of pixels in a specific region using the integral image values. The sum of pixels in the rectangle D is calculated as *D=4+1-(2+3)*, in which the location *1* is the sum of pixels in *A*, *2* is the corresponding sum in *A+B*, *3* is the sum in *A+C*, and *4* is the sum in *A+B+C+D*. (Source: P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features" [20]).

▪ **AdaBoost-based algorithm:** because not all features are relevant, a learning algorithm based on AdaBoost is used to find relevant ones and to construct a strong classifier with the selected features. A relevant feature is that which causes a good distinction between an object and one that is not. The feature is part of a weak classifier only if it is able to detect more than half of the cases. It is defined as follows

$$h_j(x) = \begin{cases} 1 & if \ p_j f_j(x) < \ p_j \theta_j \\ 0 & otherwise \end{cases} \qquad \text{(Eq. 2.2)}$$

in which $h_j(x)$ is the weak classifier, x is the 24x24 pixel sub-window of the original image, p_j is a parity, f_j is the feature, and θ_j is a threshold. Fig. 2.10 shows examples of features selected by the AdaBoost-based algorithm and the comparison of intensities with regions of a face.
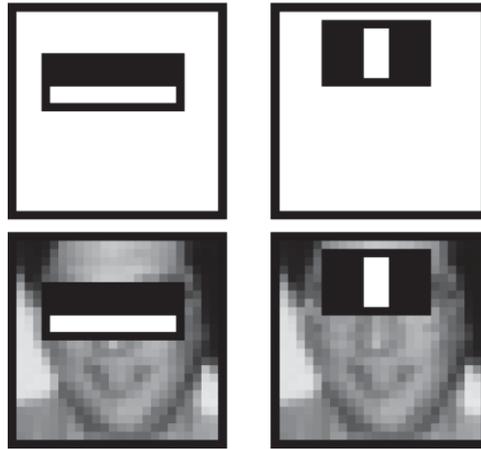
Fig. 2.10   Examples of features selected by the algorithm based on AdaBoost and the comparison of intensities with regions of a face: (top-left, bottom-left) feature with resemblance to the regions of eyes and cheeks (eyes are usually darker than the cheeks); (top-right, bottom-right) feature with resemblance to the regions of eyes and the bridge of the nose (eyes are usually darker than the nose). (Source: P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features" [20]).

- **Cascade classifier:** due to most of the sub-windows of the original image are negatives, a cascade classifier is used to find out quickly if the sub-window has the object or not. The cascade classifier is divided into several stages (classifiers), in which each stage contains a certain number of relevant features. The features are distributed into the stages instead of evaluating all of them in each sub-window. Each stage evaluates if the sub-window contains the object or not. If so, the sub-window passes to the next stage, otherwise, the sub-window is rejected (Fig. 2.11). The first stage of the cascade may include some of the most basic features with a resemblance to the regions of a face, as shown in Fig. 2.10.
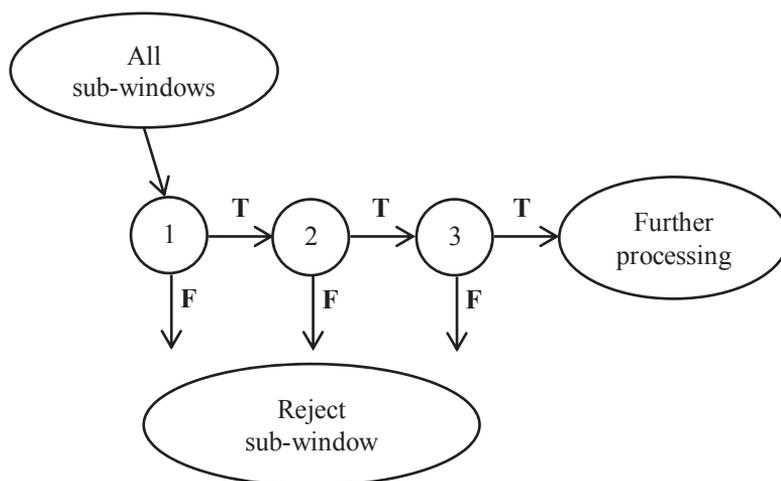
Fig. 2.11 Schematic representation of the cascade classifier. Each stage evaluates if the sub-window contains the object or not. If so, the sub-window passes to the next stage, otherwise, the sub-window is rejected. Further processing may vary depending on the detection system. (Source: P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features" [20]).

This cascade structure allows the fast rejection of a sub-window that does not contain the object, which reduces considerably the computational cost. The AdaBoost-based algorithm trains the stages of the cascade by adjusting the threshold in order to reduce the false negative cases.

2.5.2   KLT algorithm

Some studies of the state of the art have proposed the use of the Viola and Jones algorithm to carry out the ROI detection in each frame of the video sequence. However, the application of this algorithm to detect the ROI along a video sequence may be computationally expensive and not always robust, especially in the detection of faces in non-frontal poses. One approach proposed to deal with the high computational cost is the KLT algorithm [113-115]. This approach allows the tracking of a set of feature points along a video sequence. It is a

robust algorithm that is able to track objects during changes of factors such as the scale and orientation. Thus, the algorithm can be employed to track a moving object such as a person, and particularly, the face of the subject along a video sequence.

With the aim of performing a reliable tracking of an object, it is important to select good features to track. For this purpose, Shi and Tomasi [115] developed an algorithm in order to identify features points that could be reliably tracked. To achieve this, the criterion for selecting a good feature is established according to the following equation, which should be solved during the tracking

$$Zd = e \qquad \text{(Eq. 2.3)}$$

in which Z is a symmetric 2x2 matrix, d is the translation of the feature window's center, and e collects the last two entries of an error vector that depends on the difference between the two images involved. According to Shi and Tomasi, it is possible to track a window from frame to frame if the equation 2.3 represents good measurements, as well as if it can be solved reliably. In that case, the symmetric matrix Z has to be superior to the noise level and well-conditioned. In the noise level condition, the two eigenvalues λ_1 and λ_2 of Z must be large, which can represent features that can be tracked in a reliable manner. On the side of the well-conditioned requisite, both eigenvalues cannot be very different in terms of magnitude. Thus, if the criterion shown as follows

$$\min(\lambda_1, \lambda_2) > \lambda \qquad \text{(Eq. 2.4)}$$

is met, in which λ is a predefined threshold, then, the window is accepted. The value of λ is chosen halfway between lower and upper bounds, which are previously obtained by measuring the eigenvalues from a region of approximately uniform brightness and from a set of various types of features, respectively. Thus, this method for feature selection is focused on distinguishing between good and bad features in order to maximize the robustness of the tracking.

Furthermore, in order to make more robust the tracking, it is possible to enable a maximum bidirectional error due to the presence of noise along the video sequence. This bidirectional error is the distance calculated in pixels from the original location of the feature points to their final position after the backward tracking [116, 117]. If the error calculated is greater than the maximum bidirectional error set in the algorithm, then, the corresponding features points are dropped.

The estimation of a geometric transformation between the feature points of adjacent frames is established in the implementation of the algorithm in order to exclude outliers [118]. The M-estimator sample consensus (MSAC) algorithm [119] is the method employed to run this function. In case of a considerable loss of feature points during the tracking, the reacquisition of points could be performed as often as necessary.

Chapter 3

Materials and methods

This chapter describes the materials and methods used throughout the research, which are referenced in the thesis to provide a quick access to their description. Other specific information is described in the corresponding chapters.

3.1   Instrumentation

All measurements of this thesis were performed on a simultaneous recording of a video camera and a reference system in a sitting position. A representation of

this configuration is illustrated in Fig. 3.1. The measurements were conducted according to the principles defined in the Declaration of Helsinki.
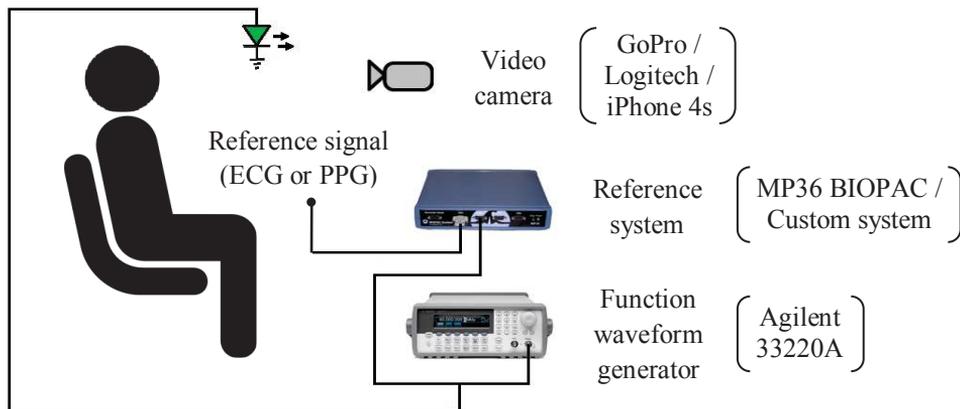
Fig. 3.1 Data acquisition set-up of the measurements.

3.1.1   Reference systems

A reference signal was recorded to evaluate the agreement of the measurements obtained by the video imaging method. The reference systems used to record the data were:

- **MP36 BIOPAC Systems, Inc.:** two input channels were used on this reference system (Fig. 3.2). The first channel was used to record the reference signal, and the second one to record a sawtooth signal in order to facilitate the synchronization of the heart rate signals (see section 4.2.4).

Fig. 3.2   MP36 BIOPAC Systems, Inc. reference system.

The ECG or PPG signal was recorded as the reference data of the measurements. The SS2LB BIOPAC Systems, Inc. lead set (Fig. 3.3 (left)) is a fully-shielded cable that was used to record the ECG signal and that minimizes interferences during the recording of the ECG signal. The lead II configuration, which consists of three electrodes around the chest area (Fig. 3.3 (right)), was used to record the ECG signals.
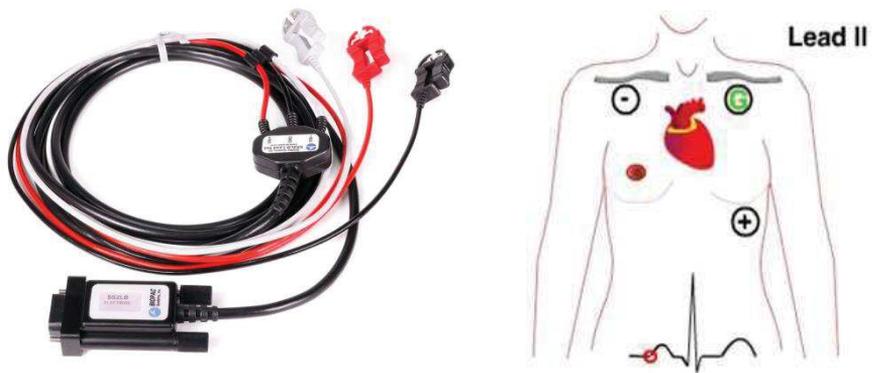
Fig. 3.3 (Left) SS2LB BIOPAC Systems, Inc. lead set.; (right) lead II ECG configuration (source: © RnCeus.com).

It is one of the most common monitoring configurations because it has a positive and well pronounced R wave to detect the RR intervals. The terminals of this configuration are connected to the corresponding lead set cables as follows:

- Positive terminal (+) → red cable
- Negative terminal (-) → white cable
- Ground terminal (G) → black cable

The SS4LA BIOPAC Systems, Inc. finger pulse sensor (Fig. 3.4 (left)) was used to record the PPG signal. This reflectance-mode PPG sensor is a transducer that consists of an infrared emitter and a photodiode detector (Fig. 3.4 (right)). It converts the capillary blood volume variations to an electrical pulse signal via optical methods. The reflected infrared light, which depends on the blood volume changes, is detected by the photodiode in order to measure these variations.

Fig. 3.4   (left) SS4LA BIOPAC Systems, Inc. finger pulse sensor; (right) diagram of the reflectance-mode PPG sensor.

In both reference signals, the data was recorded at a sampling frequency of 1 kHz in order to satisfy the established recommendations by the Task Force of the European Society of Cardiology and The North American Society of Pacing and Electrophysiology [2]. For a more detailed description of this requirement refer to section 2.2.3.

- **ECG acquisition system**: it was used as an alternative option to record the ECG signal. It consists of an ECG acquisition board, an isolated amplifier, and a NI USB-6212 DAQ device (Fig. 3.5), which are described below.
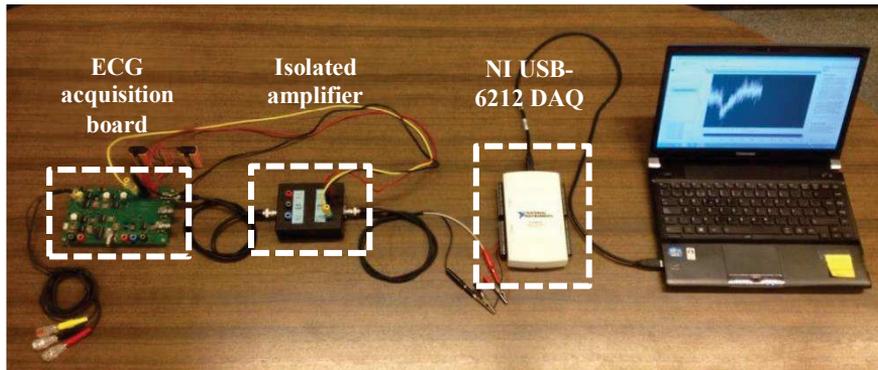
Fig. 3.5   ECG acquisition system that consists of an ECG acquisition board, an isolated amplifier, and a NI USB-6212 DAQ device.

> **ECG acquisition board:** in general, it consists of four circuit stages: a preconditioning and first amplification stage, a high-pass filter and second amplification stage, a notch filter, and a low-pass filter. For a more detailed description of each stage and its corresponding electrical diagram, refer to the appendix (section A.1). The input signal of the ECG acquisition board is obtained from an ECG lead set and the corresponding output is connected to the isolated amplifier (Fig. 3.6).
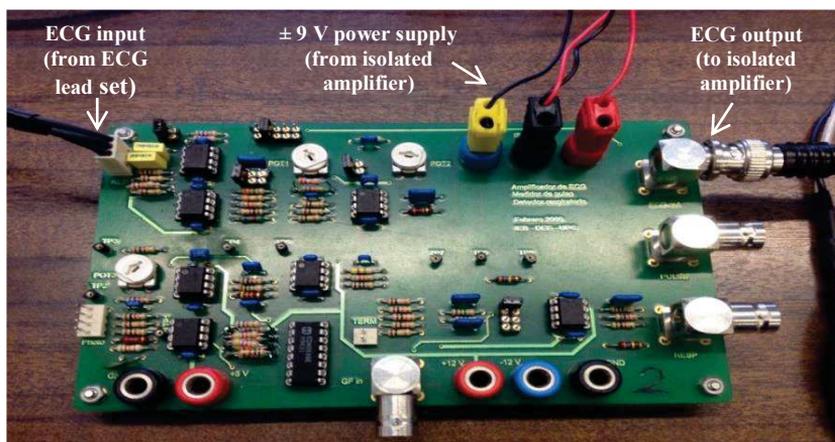
Fig. 3.6   ECG acquisition board.

- ➢ **Isolated amplifier:** its main function is to protect the subject from any accidental contact with the electrical power line (Fig. 3.7). Its input is supplied from the output of the ECG acquisition board and its output is connected to the NI USB-6212 DAQ. It is powered by a ± 12 V power supply and the ± 9 V isolated output powers the ECG acquisition board. The gain of the amplifier is 1.
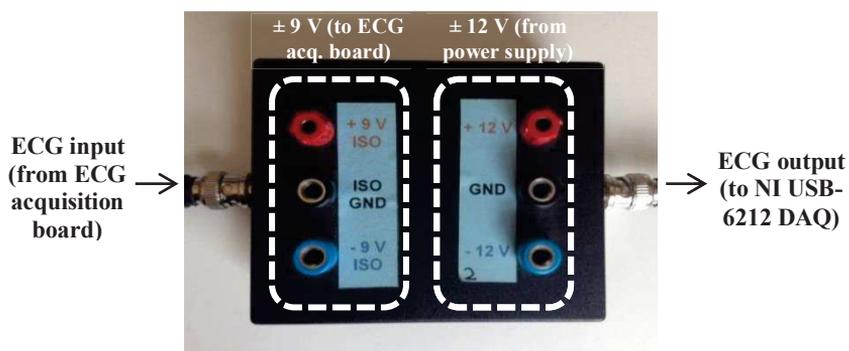
Fig. 3.7   Isolated amplifier powered by a ± 12 V power supply. The ± 9 V isolated output powers the ECG acquisition board. The gain of the amplifier is 1.

- ➢ **NI USB-6212 DAQ:** it is a multifunction data acquisition device that has 16 analog inputs (16-Bit, 400 kS/s), 2 analog outputs (16-Bit, 250 kS/s), and 32 digital I/O (Fig. 3.8). It is bus-powered which provides an easy portability and data transfer to computers.

Fig. 3.8   NI USB-6212 DAQ device.

40

Likewise, two inputs of this device were used for data acquisition. One input was utilized to record the ECG signal obtained from the isolated amplifier and the second one to record a sawtooth signal. The data were recorded at 1 kHz using the LabVIEW software.

▪ **RespiBand by plux of FICOSA International S.A.:** it is an inductive chest band capable of measuring the respiratory signal (Fig. 3.9). The band measures the displacements of the chest or abdomen caused by the respiration. A module is connected to the band, which records and sends the data by Bluetooth to the computer. The sampling frequency of the inductive band is 40 Hz.

Fig. 3.9   (Left) RespiBand by plux of FICOSA International S.A.; (right) subject using the band during the recordings.

3.1.2   Video cameras

One of the objectives of this research is the performance analysis of commercial cameras in obtaining the iPPG signal. Three camera models with different features were evaluated in the measurements (Fig. 3.10):

- **GoPro HERO3 silver edition:** it is a versatile model that has been widely used lately in different scenarios and has the capability of diverse video resolutions. Depending on the resolution, it is possible to record videos from 24 to 120 frames per second (fps). It has the capability of Wi-Fi connection which enables its operation by using its own mobile phone application. Its small size facilitates its portability.

- **Logitech HD Pro Webcam C910:** similarly to the GoPro model, this camera also has the capability of different video resolutions but with a lower number of fps. The maximum frame rate achievable is 30 fps, which is reached at the video resolution of 640x480 pixels. The camera enables the manual adjustment of diverse parameters such as focus, gain, exposure time, and others. It can only be operated by using a computer with a USB connection.

- **iPhone 4s:** it is a phone model that has back and front built-in cameras with the capability of recording at two video resolutions. The front camera records at 640x480 pixels and the back camera at 1920x1080 pixels, both up to 30 fps. It has the capability of autofocus, which may be disabled.

Fig. 3.10   Camera models: (Left) GoPro HERO3 silver edition; (middle) Logitech HD Pro Webcam C910; (right) iPhone 4s.

3.1.3   Function waveform generator

An Agilent 33220A function waveform generator (Fig. 3.11) was used to facilitate the synchronization of the heart rate signals. A sawtooth signal of 1 Hz was provided by the function waveform generator in order to achieve the alignment of the signals, as described in section 4.2.4.

Fig. 3.11   Agilent 33220A function waveform generator.

3.1.4   Light meter

An Amprobe LM-100 light meter (Fig. 3.12) was utilized to measure the light intensity in lux units that illuminated the face of the subjects (refer to section 6.1.2). Thus, it is possible to ensure the same illuminance, for example, between subjects or when different sources of illumination are compared.

Fig. 3.12   Amprobe LM-100 light meter.

43

3.2 Statistical parameters and plots

The following statistical parameters and plots were calculated to measure the agreement and the error between the reference system and the proposed method by video imaging.

3.2.1 Correlation coefficients

- **Pearson's correlation coefficient (r):** it is a statistical parameter used to measure the linear association between two continuous variables x and y. It is defined by

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \text{(Eq. 3.1)}$$

in which r ranges from the -1 to +1 interval. A value of +1 indicates a perfect positive association, 0 means no association, and -1 indicates a perfect negative association.

- **Intraclass correlation coefficient (ICC):** it is a statistical parameter that measures absolute agreement between two continuous variables

$$ICC = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} \qquad \text{(Eq. 3.2)}$$

in which T represents the mean of the replicate measurements of the variable of interest and, e is the difference between a single measurement X

and its underlying mean value T. In general, ICC values below 0.4 represent poor agreement, values between 0.4 and 0.75 indicate a good agreement, and values above 0.75 represent an excellent agreement in the measurements [120].

3.2.2 Measurement errors

▪ **Mean bias error (MBE):** it is a measure used to indicate the presence of bias in the measurements. It is defined as

$$MBE = \frac{1}{n} \sum_{i=1}^{n} p_i - o_i \qquad \text{(Eq. 3.3)}$$

in which p_i is the predicted value and o_i is the observed value.

▪ **Mean absolute error (MAE):** it is a measure that represents the average of the absolute errors between two continuous variables. It is defined as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |p_i - o_i| \qquad \text{(Eq. 3.4)}$$

in which p_i is the predicted value and o_i is the observed value.

- **Mean absolute percentage error (MAPE):** it is a measure that expresses accuracy as a percentage of the error between two continuous variables. It is defined as

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{p_i - o_i}{p_i} \right| \qquad \text{(Eq. 3.5)}$$

in which p_i is the predicted value and o_i is the observed value.

- **Root mean square error (RMSE):** it is a common measure of the differences between two continuous variables, but in comparison with MAE, RMSE punishes large errors. It is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2} \qquad \text{(Eq. 3.6)}$$

in which p_i is the predicted value and o_i is the observed value.

3.2.3 Bland-Altman plot

Bland-Altman plots with limits of agreement (LoA) of 95% were obtained to compare the measurements from the reference system and video, as well as to detect the presence of systematic errors. It is a plot used to assess the agreement between two methods of clinical measurement [121]. Each point on the plot is represented by the mean of the measurements x and y obtained by the two

methods on the X-axis, and by the difference between these measurements on the Y-axis

$$S(X,Y) = \left(\frac{x+y}{2}, x - y\right) \qquad \text{(Eq. 3.7)}$$

The LoA of 95% were calculated as the mean difference ± 1.96 the standard deviation of the differences

$$LoA\ (95\%) = \bar{d} \pm 1.96s \qquad \text{(Eq. 3.8)}$$

If the differences obtained between the methods are not considered clinically important, both may be used interchangeably.

Chapter 4

Video HRV analysis: initial measurements

In the last few years, some studies have measured HR or HRV parameters using a video camera [11-14]. This technique focuses on the measurement of the small changes in skin color caused by blood perfusion. In this chapter, the HR and HRV were measured using a video of the face of the subject. A cross-correlation analysis was performed with the aim of finding the face regions that provide more information on HR. The face areas that obtained high correlation values were the chosen positions to obtain the iPPG signal. The video recordings were obtained with two cameras at different video resolution. The statistical analysis showed a good agreement between the reference system and the video, as well as notable differences in performance between both cameras.

4.1   Data acquisition set-up

A summary of the measurement set-up is shown below:

- 12 subjects (3 women, 9 men) between 23 and 35 years old were recorded with an iPhone 4s camera at 640x480 pixels (low-resolution video) and 30 fps (see section 3.1.2).
- 15 subjects (3 women, 12 men) between 23 and 35 years old were recorded with a Logitech webcam at 1920x1080 pixels (high-resolution video) and 15 fps (see section 3.1.2).
- The recordings were obtained in stationary conditions (see Fig. 3.1).
- The recording length was 50 s.
- The approximate distance between the camera and the face was 40 cm in the iPhone 4s videos and 30 cm in the case of the Logitech recordings.
- The videos of both cameras were not performed on simultaneous recording.
- The recordings were performed indoors with sunlight as the lighting source.

4.2   Video and signal processing

4.2.1   Frame averaging

Although the measurements were performed in stationary conditions, a procedure of pixel averaging was carried out in order to attenuate small motion artifacts [24, 27]. The average of a number of rows and columns of pixels was performed to obtain a representative mean value by zone. The green channel of the video was used to perform the pixel averaging, as well as in the acquisition of the iPPG signals of this thesis [17, 18]. The videos recorded with the Logitech model were averaged on blocks of 40 pixels and those of the iPhone camera on blocks of 20 pixels. The frame average procedure was carried out on each frame

of the video sequence. For example, the video frames of 1920x1080 pixels were averaged to obtain reduced images of 48x27 blocks (Fig. 4.1).
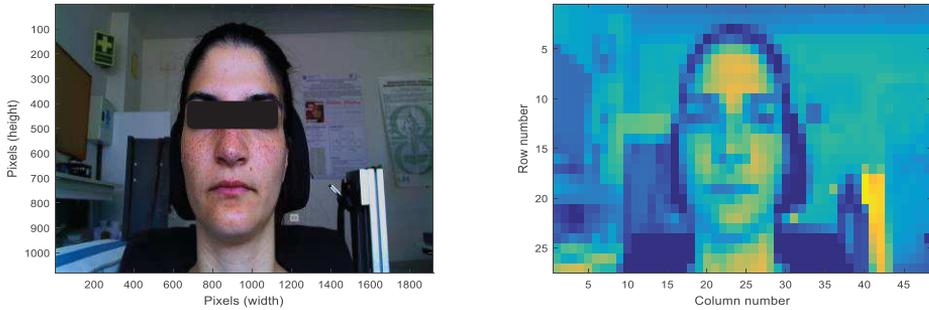
Fig. 4.1   (Left) frame of 1920x1080 pixels; (right) averaged frame of 48x27 blocks.

Once the video frames were averaged, a manual selection of the ROI was carried out by including most of the face area (Fig. 4.2). The selected ROI allows the acquisition of an iPPG reference signal in order to be employed in the cross-correlation analysis.
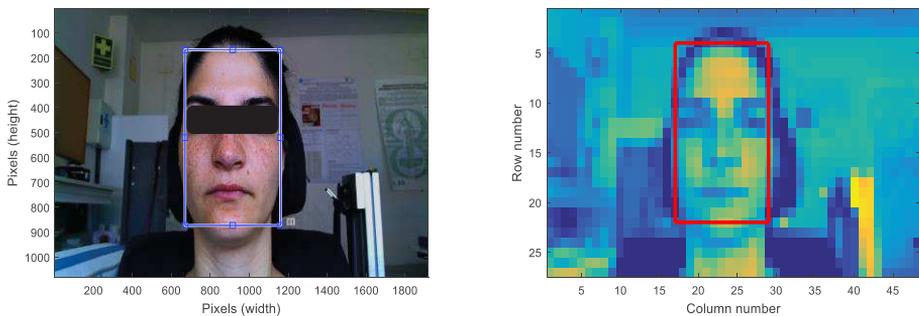
Fig. 4.2   (Left) selected ROI on video frame; (right) equivalent ROI on the averaged frame.

The values of the blocks within the ROI were averaged along the sequence of reduced images to obtain an iPPG reference signal. A first-order bandpass Butterworth filter between 0.6 and 2 Hz was applied to remove low-frequency (respiration and motion) and high-frequency components (electronic noise in the optical sensor of the camera) (Fig. 4.3). These cut-off frequencies correspond from 36 to 120 bpm, which cover normal values of heart rate at rest.
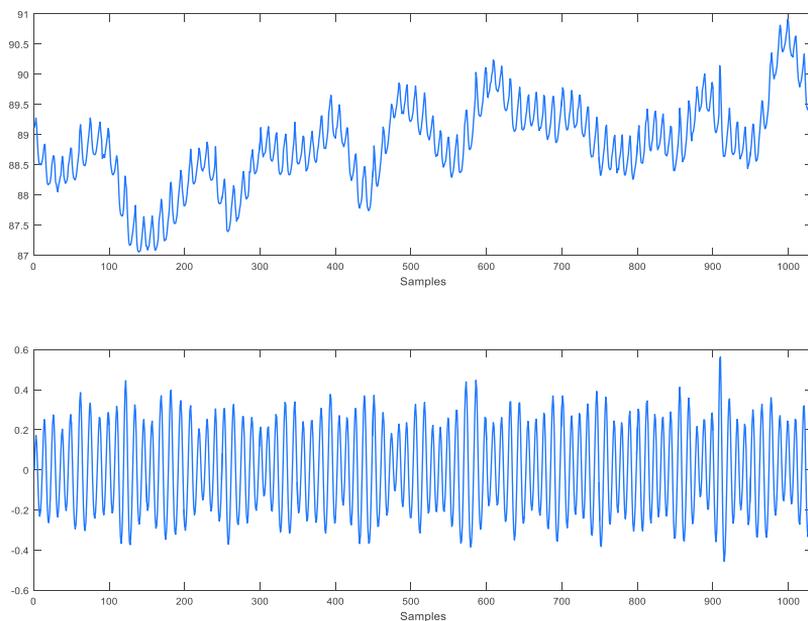
Fig. 4.3 (Top) raw iPPG reference signal; (bottom) iPPG reference signal after applying the filter.

4.2.2 Cross-correlation analysis

A cross-correlation analysis (Eq. 4.1) was performed with the aim of finding the blocks positions that could provide more information on HR

$$r_{xy}(l) = \sum_{n=-\infty}^{\infty} x(n)y(n-l) \qquad \text{(Eq. 4.1)}$$

in which *x(n)* is the iPPG reference signal and *y(n-l)* correspond to the signals formed by the values of each block position along the sequence of averaged frames. The similarity between the iPPG reference signal and all signals corresponding to each block position were measured by this method. The noise components of all these signals were removed by using the bandpass filter applied previously to the iPPG reference signal.

Clearly, the signals obtained from a distant block position from the face area achieved lower correlation results. As an example, the signal obtained from the first block position (row=1, column=1) achieved a maximum cross-correlation value of 0.2271 (Fig. 4.4).
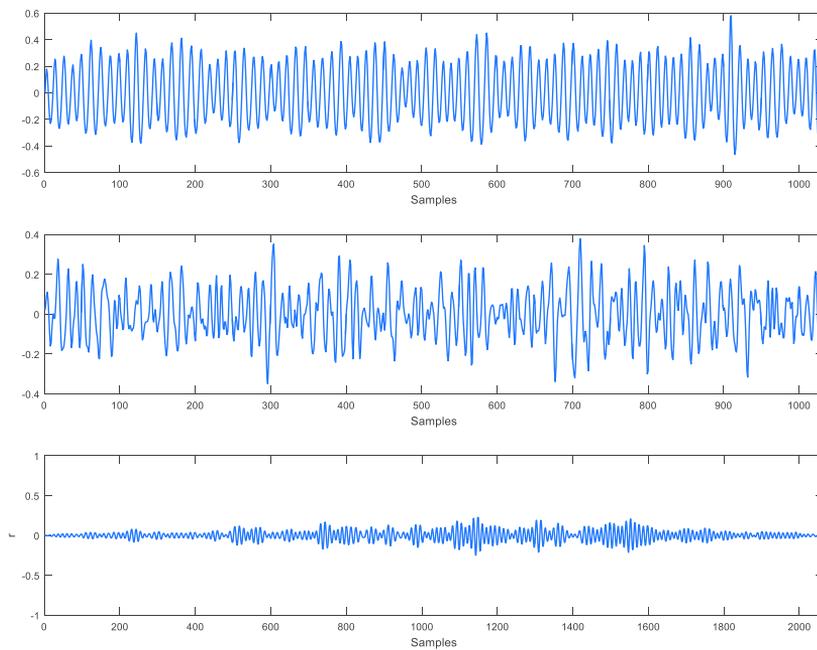
Fig. 4.4   (Top) iPPG reference signal; (middle) signal obtained from the first block position (row=1, column=1) along the sequence of averaged frames; (bottom) result of the cross-correlation analysis with a maximum value of 0.2271.

By contrast, the signals obtained from a block position of the face area achieved higher correlation results, especially from the forehead and cheek regions. These regions have a major blood perfusion in the face which provides more information on the heart rate [27]. Fig. 4.5 shows the signal obtained from a block position of the forehead region (row=9, column=24) and the result of the cross-correlation analysis, which achieved a maximum value of 0.9409.
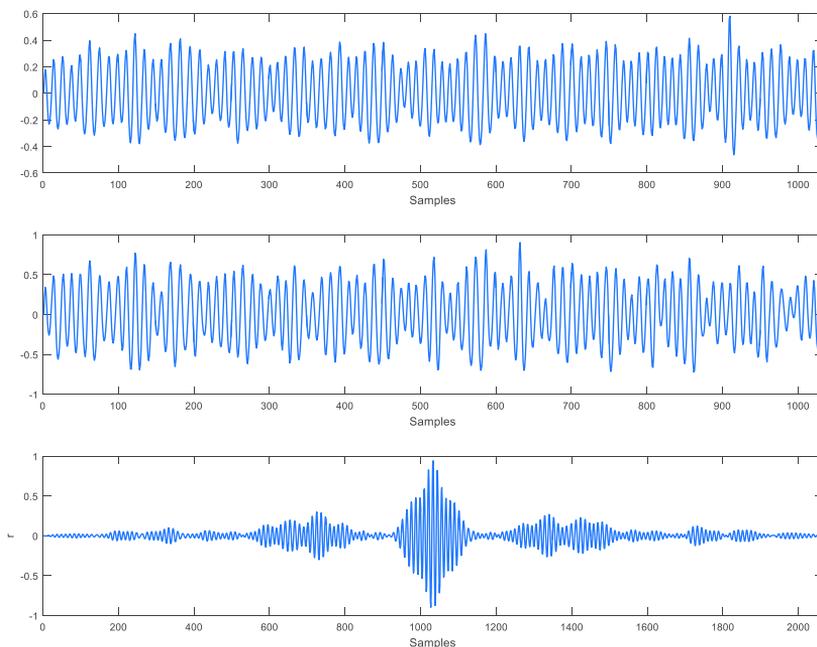
Fig. 4.5   (Top) iPPG reference signal; (middle) signal obtained from a block position of the forehead region (row=9, column=24) along the sequence of averaged frames; (bottom) result of the cross-correlation analysis with a maximum value of 0.9409.

The maximum value of the cross-correlation of each block position was stored in a matrix. These values were compared with a threshold of 0.7, within a range from 0 to 1, to distinguish between positions with high and low correlation results. On average, this threshold achieved better correlation and error results for all the video signals in comparison with lower and higher threshold values. It

presents a good compromise that allows a distinction between regions with a greater cardiac component (forehead and cheeks) in comparison with other parts of the face (eyes, mouth, and nose) and zones with the absence of a cardiac component. The values of block positions that were above the threshold and within the ROI were classified as high cross-correlation positions (Fig. 4.6).
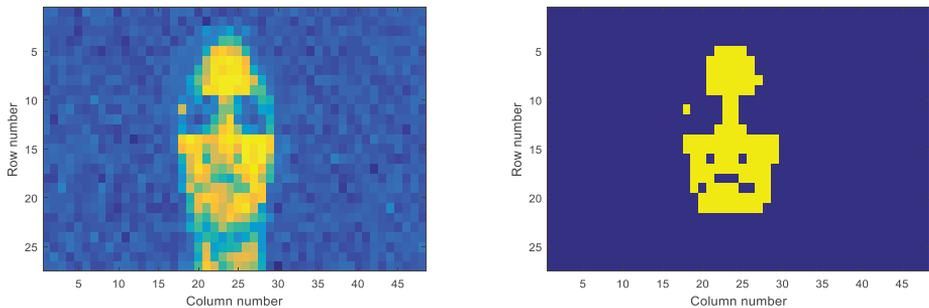
Fig. 4.6   (Left) matrix of maximum cross-correlation values; (right) matrix of high cross-correlation positions (yellow blocks).

4.2.3   iPPG signal acquisition

The values of the high cross-correlation positions were averaged on each reduced frame to obtain a raw iPPG signal. A first-order bandpass Butterworth filter between 0.6 and 2 Hz was applied to remove low and high-frequency noise components. Then, a cubic spline interpolation was performed to improve the temporal resolution from the sampling frequency of video to 1 kHz. Fig. 4.7 shows the iPPG signal results obtained from a subject.

According to the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, a minimum sampling rate is required to perform an appropriate HRV analysis [2]. An optimal range is established between 250-500 Hz or even higher. Therefore, the frequency of 1 kHz was chosen to record the reference signal, as well as to improve the

temporal resolution of the iPPG signal. The signal processing described hereinafter was carried out in all measurements of this study.
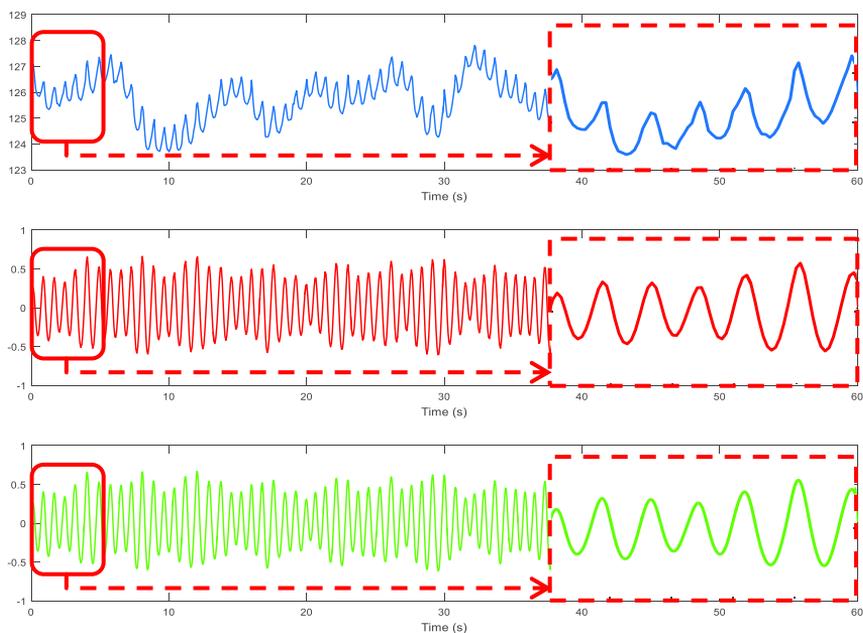
Fig. 4.7   iPPG signals obtained from a subject: (top) raw signal; (middle) filtered signal; (bottom) interpolated signal.

4.2.4   Synchronization of the signals

The alignment of the heart rate signals was necessary to perform a comparative HRV analysis since the camera and the reference system were not synchronized in time. A sawtooth signal of 1 Hz was obtained from a function waveform generator and recorded by the reference system. The sawtooth signal was also obtained by video from the green light emitting diode (LED) connected to the function waveform generator and positioned close to the participants (Fig. 3.1). The LED was turned on when both systems were recording and then turned off before the recordings ceased.

The sawtooth signal by video was obtained by measuring the changes in the light intensity of the LED along the video sequence (Fig. 4.8). Thus, by calculating the time offset between the sawtooth signals, the heart rate signals were synchronized adjusting the same delay between them (Fig. 4.9).

Fig. 4.8   Sawtooth signal acquisition: (top) frame with LED off; (bottom) frame with LED on.
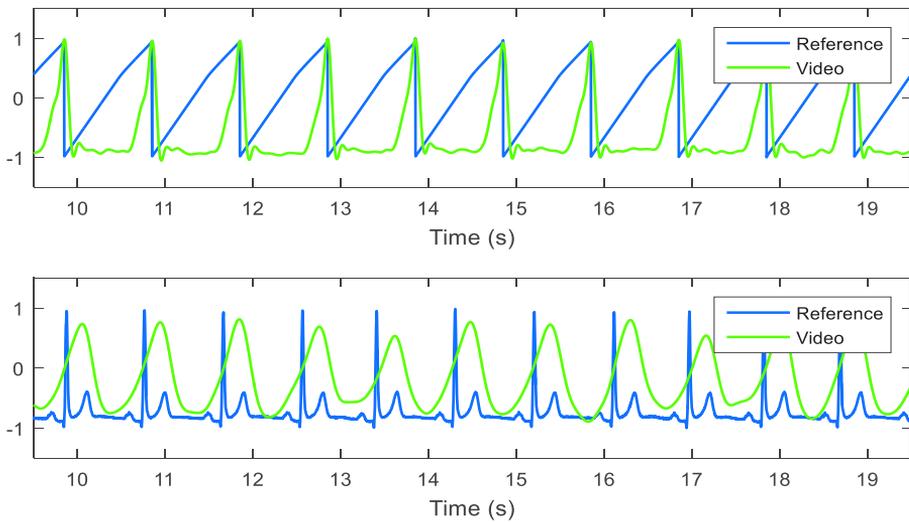
Fig. 4.9   Synchronized signals: (top) sawtooth signals; (bottom) heart rate signals.

4.2.5 Artifacts correction

Once the signals were synchronized, the local maxima were detected in order to calculate the NN intervals. The Pan-Tompkins algorithm was employed to detect the peak of the QRS complexes of the ECG signals [122]. In the case of the PPG and iPPG signals, the local maxima were detected. A representation of the local maxima detection is shown in Fig. 4.10. Subsequently, an artifact correction was carried out when false positive or false negative measurements affected the NN time series. The NN intervals falling outside established thresholds were identified as artifacts and replaced with new values. The thresholds were defined as the median ± 4 standard deviations of the NN time series.
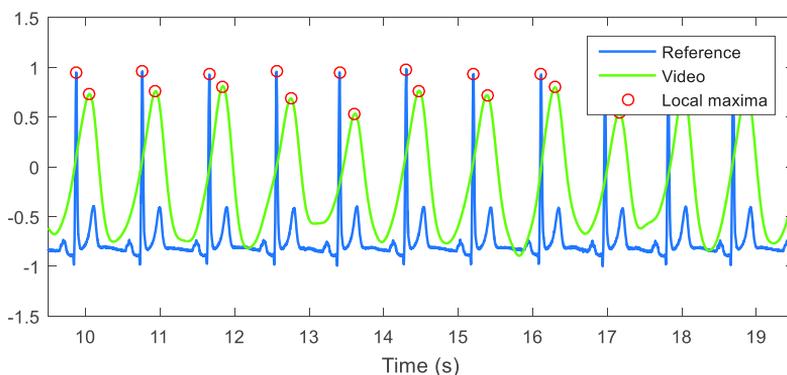
Fig. 4.10   Local maxima detection on the heart rate signals.

When a false positive (false beat) was detected, two incorrect intervals were originated in the NN time series. These erroneous measurements are usually shorter than the expected values because they correspond to one NN interval. Thus, the incorrect measurements were replaced with the sum of their corresponding values. In the case of NN measurements above the upper threshold caused by a false negative (missing beat), they were replaced with the average of the five previous NN intervals.

Furthermore, the absolute values of the differences between adjacent NN measurements (NN-NN) that were above the median + 4 standard deviations of the NN-NN series were identified as outliers. These outliers were also replaced by following the same procedure as that in a false negative case.

4.3   HRV analysis

A HRV analysis was performed according to the established recommendations (refer to section 2.2.3). Some of the most common time and frequency domain parameters of HRV were calculated in the analysis (refer to sections 2.2.1 and 2.2.2). The time-domain parameters were calculated in MATLAB® and the frequency-domain parameters were obtained with the Kubios HRV software (version 2.2) by using the calculated NN time series. A comparative example of the NN intervals obtained by the calculation of time between consecutive local maxima is illustrated in Fig. 4.11. The LF and HF components were reported in n.u. These parameters were obtained by the Welch's method, which employs the FFT for the calculation of PSD.
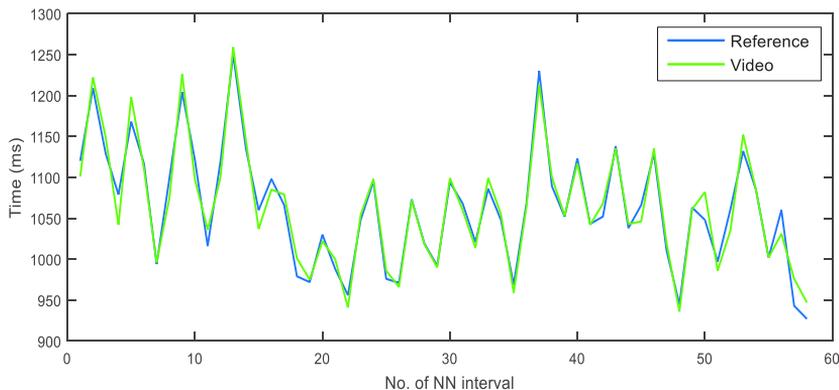
Fig. 4.11   NN intervals of reference and video obtained from a subject.

4.4   Results and discussion

4.4.1   Low-resolution video

Table 4.1 presents the statistical results of the NN time series achieved by each subject using the iPhone 4s front camera (640x480 pixels and 30 fps). It is noted a wide variety of results. The correlation coefficients reached maximum values of 0.984 and 0.973 and, in contrast, lower correlation values of 0.736 and 0.753 were obtained as well. The mean ± standard deviation ($\bar{x} \pm s$) of the individual results were calculated to obtain an overall perspective of the data. The Pearson and ICC measures were 0.847 ± 0.088 and 0.837 ± 0.096, respectively. The mean of the errors achieved a MAPE = 2.49 %, which corresponds to a MAE = 22.27 ms or an RMSE = 29.44 ms.

Table 4.1   Statistical results of the NN time series by subject (iPhone 4s)

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,917	0,912	-0,80	0,01	19,41	1,04	1,83	24,31	1,31
2	0,861	0,852	-0,32	-0,01	20,07	0,88	1,71	25,90	1,16
3	0,779	0,766	-0,25	-0,06	24,02	1,65	2,57	33,48	2,30
4	0,984	0,984	0,15	-0,03	11,07	0,61	1,06	13,62	0,75
5	0,973	0,972	0,12	-0,05	10,88	0,94	1,30	13,70	1,23
6	0,736	0,704	0,52	-0,34	28,88	3,86	4,28	36,06	4,94
7	0,784	0,752	-0,14	-0,20	29,09	3,17	3,89	35,58	3,93
8	0,925	0,924	1,24	-0,17	24,98	1,97	2,85	31,54	2,56
9	0,785	0,772	0,55	-0,13	35,64	1,77	3,23	51,80	2,57
10	0,753	0,750	0,29	-0,06	21,79	1,47	2,31	34,26	2,20
11	0,778	0,776	0,79	-0,11	25,96	1,89	2,84	33,51	2,51
12	0,883	0,874	-0,16	-0,05	15,40	1,58	2,01	19,49	2,05
Mean	**0,847**	**0,837**	**0,17**	**-0,10**	**22,27**	**1,74**	**2,49**	**29,44**	**2,29**
S.D.	**0,088**	**0,096**	**0,55**	**0,10**	**7,44**	**0,95**	**0,99**	**10,78**	**1,20**

*: All p < 0.001. Refer to section 3.2 for statistical parameters definitions.

Fig. 4.12 shows the correlation scatter plots of the HR and HRV parameters obtained by the iPhone 4s camera, and Table 4.2 presents the corresponding statistical results of the data. The HR, NNmean, SDNN, and frequency-domain parameters achieved the best correspondence results and, conversely, the time-domain parameters RMSSD and pNN50 obtained a noticeable lower agreement.
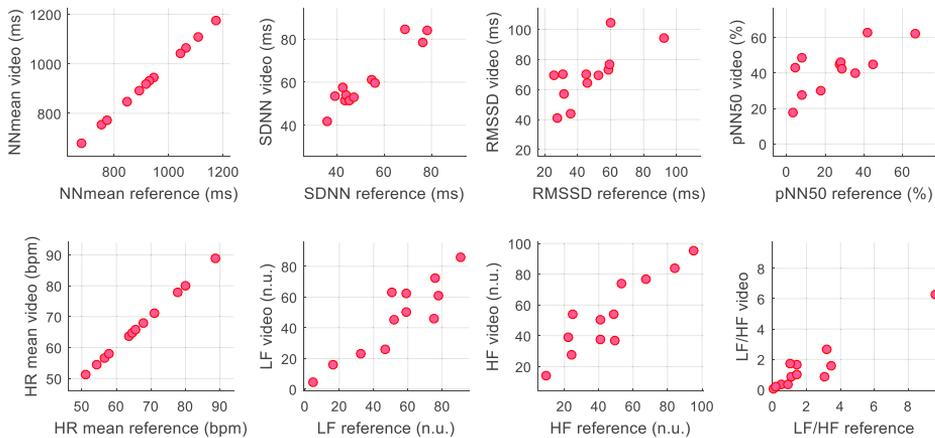
Fig. 4.12   Correlation scatter plots of the HR and HRV parameters (iPhone 4s). Refer to Table 4.2 for r and ICC results.

Table 4.2   Statistical results of the HR and HRV parameters (iPhone 4s)

Parameter	r	ICC	MAE	RMSE
HRmean (bpm)	1,000*	1,000*	0,10	0,14
NNmean (ms)	1,000*	1,000*	0,44	0,56
SDNN (ms)	0,948*	0,811*	8,29	9,39
RMSSD (ms)	0,729**	0,425**	22,41	25,96
pNN50 (%)	0,716**	0,453**	17,11	20,86
LF (n.u.)	0,900*	0,872*	9,69	12,96
HF (n.u.)	0,900*	0,873*	9,62	12,88
LF/HF	0,940*	0,826*	0,86	1,33

*: $p < 0.001$, **: $p < 0.05$. Refer to section 3.2 for statistical parameters definitions.

61

Moreno et al. [27] also performed a video HRV analysis carrying out a procedure of pixel averaging prior to obtaining the iPPG signals. In this study, the video resolution and frame rate were the same as those used in our work. In the case of the sitting posture results (refer to Table 2.3), the time-domain parameters reported by Moreno et al. achieved a higher correlation agreement in comparison with the ICC results presented in Table 4.2. By contrast, the frequency-domain parameters obtained lower correlation values than those reported in our study, particularly the LF/HF component.

With the aim of conducting the analysis in as fair a way as possible, the results obtained throughout this work are contrasted with data reported in reference works carried out in similar conditions. Therefore, in this case, only the results obtained in the sitting position by Moreno et al. are discussed in this section. Also, it is important to note that in terms of agreement measurement, the ICC measure is almost identical to the concordance correlation coefficient (ρc) [123] measured by Moreno et al. Nevertheless, in both works, the RMSSD and pNN50 parameters obtained a significant low agreement among all the HRV parameters.

On the other hand, Sun et al. [24] conducted a PRV analysis focused on the palm of the subject's hand. In the study, a procedure of reduced frames by pixel averaging was carried out to attenuate small motion artifacts. The PRV parameters achieved good Pearson's correlation values (refer to Table 2.3), which present a better correspondence in comparison with the results reported in Table 4.2. The correlation data of the RMSSD and pNN50 components were not reported in the study, which are two parameters usually more affected in the statistical analysis. Based on the results reported by Sun et al., the method proposed by them seems a good option to perform a video PRV analysis.

4.4.2   High-resolution video

Table 4.3 presents the statistical results of the NN time series achieved by each subject using the Logitech camera (1920x1080 pixels and 15 fps). In comparison with the results obtained with the iPhone 4s, most of the subjects obtained correlation coefficients above 0.9. The $\bar{x} \pm s$ of the Pearson and ICC measures increased to 0.926 ± 0.030 and 0.924 ± 0.032, respectively. The mean of the errors was improved to a MAPE = 1.99 %, which corresponds to a MAE = 16.85 ms or an RMSE = 22.02 ms.

Table 4.3   Statistical results of the NN time series by subject (Logitech)

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,943	0,942	0,20	-0,04	18,64	0,96	1,72	24,95	1,28
2	0,926	0,926	0,39	-0,05	11,57	0,93	1,33	16,02	1,28
3	0,892	0,889	0,29	-0,07	10,88	1,27	1,52	15,26	1,81
4	0,877	0,875	0,06	-0,17	27,88	3,12	3,78	37,14	4,24
5	0,930	0,926	0,25	-0,13	16,32	1,67	2,11	21,54	2,37
6	0,975	0,976	-0,30	0,00	12,98	1,04	1,49	16,31	1,35
7	0,915	0,908	0,16	-0,17	20,77	2,07	2,67	27,37	2,80
8	0,960	0,960	-0,21	0,00	13,24	1,29	1,68	16,04	1,60
9	0,920	0,917	-0,43	-0,06	22,06	1,71	2,52	29,96	2,31
10	0,923	0,921	0,34	-0,10	15,91	1,66	2,09	21,23	2,25
11	0,950	0,951	1,11	-0,05	15,50	0,86	1,49	19,83	1,15
12	0,954	0,955	-0,40	0,01	17,84	0,84	1,58	21,70	1,03
13	0,879	0,870	1,12	-0,13	22,39	1,42	2,29	28,44	1,86
14	0,902	0,900	0,11	-0,07	17,38	1,89	2,33	22,09	2,41
15	0,949	0,947	0,06	-0,03	9,37	0,91	1,19	12,49	1,19
Mean	0,926	0,924	0,18	-0,07	16,85	1,44	1,99	22,02	1,93
S.D.	0,030	0,032	0,46	0,06	5,00	0,61	0,67	6,64	0,85

*: All p < 0.001. Refer to section 3.2 for statistical parameters definitions.

Fig. 4.13 shows the correlation scatter plots of the HR and HRV parameters obtained by the Logitech camera, and Table 4.4 presents the corresponding statistical results of the data. High correlation results were obtained in all

parameters with values above 0.9 in most of them. The correlation scatter plots show an excellent correspondence in all parameters.
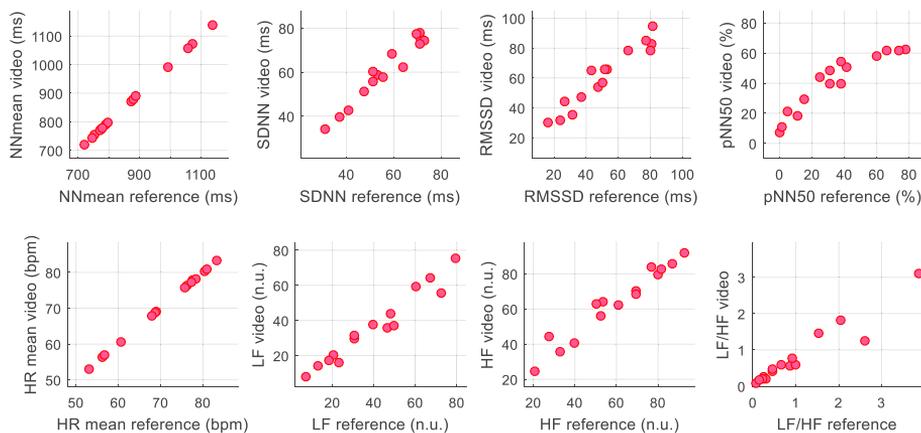

Fig. 4.13   Correlation scatter plots of the HR and HRV parameters (Logitech). Refer to Table 4.4 for r and ICC results.

Table 4.4   Statistical results of the HR and HRV parameters (Logitech)

Parameter	r*	ICC*	MAE	RMSE
HRmean (bpm)	1,000	1,000	0,07	0,09
NNmean (ms)	1,000	1,000	0,36	0,48
SDNN (ms)	0,977	0,935	4,41	5,11
RMSSD (ms)	0,962	0,867	10,21	11,54
pNN50 (%)	0,934	0,865	10,66	11,91
LF (n.u.)	0,974	0,954	4,30	6,49
HF (n.u.)	0,975	0,954	4,27	6,47
LF/HF	0,957	0,900	0,23	0,43

*: All $p < 0.001$. Refer to section 3.2 for statistical parameters definitions.

By comparing the data showed in Table 4.4 with the corresponding results obtained with the iPhone 4s camera, it is noted an improvement of the statistical

data, especially in the RMSSD and pNN50 parameters. Likewise, the correlation results were improved in comparison with those obtained in the studies of Moreno et al. [27] and Sun et al. [24], which are discussed in section 4.4.1. Despite the higher frame rate of 200 fps utilized in the work of Sun et al., the results did not prove to be better. The lower video resolution and the analyzed ROI may have affected the measurements. Also, it is important to note that the sample size was larger in our work, but the recording length was longer in the study performed by Sun et al.

In the case of the study of Moreno et al., one of the factors that may have an influence on the measurements is the video resolution, as the results obtained with the iPhone 4s camera. Moreno et al. recorded the videos with a resolution of 640 × 480 pixels, which is lower in comparison with the video resolution of the Logitech camera recordings. Moreover, it seems that the higher frame rate of 30 fps has not been significant in improving the measurements.

4.4.3   Bland-Altman plots

Fig. 4.14 shows the Bland-Altman plots with LoA of 95% representing the agreement between the NN measurements of reference and video obtained by the two cameras. As expected, it is noted a difference in results between both datasets. The measurements obtained with the low-resolution videos presented larger differences between the reference and video than those achieved with the high-resolution recordings. The low-resolution recordings obtained LoA from -60.82 to 61.19 ms (-5.45 to 5.23 bpm), and the high-resolution recordings achieved LoA from -44.72 to 45.07 ms (-4.30 to 4.16 bpm). Practically, no systematic errors were identified in the measurements obtained by both cameras.
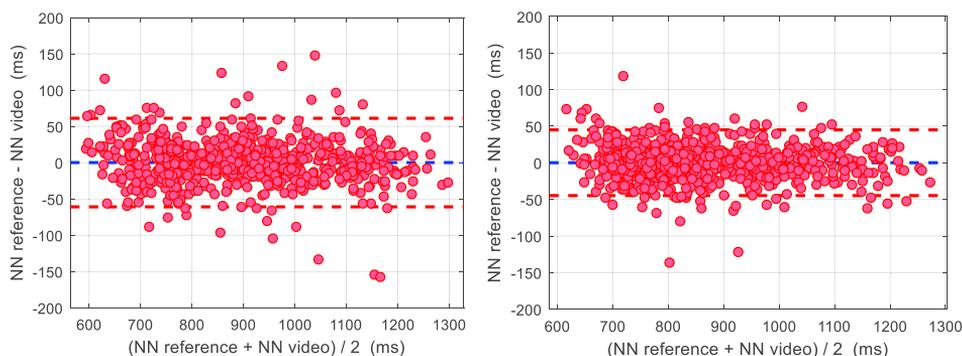
Fig. 4.14 Bland-Altman plots with LoA of 95% representing the agreement between the reference and video (data of all subjects combined in one dataset): (left) low-resolution video (iPhone 4s camera) with a mean difference of 0.18 ms (-0.11 bpm) and LoA from -60.82 to 61.19 ms (-5.45 to 5.23 bpm); (right) high-resolution video (Logitech camera) with a mean difference of 0.17 ms (-0.07 bpm) and LoA from -44.72 to 45.07 ms (-4.30 to 4.16 bpm).

4.5 Conclusions

In this chapter is presented a video HRV analysis performed in stationary conditions. The method employs a cross-correlation analysis with the aim of finding the face regions that could provide more information on heart rate. The high cross-correlation positions of the averaged frames were the regions of the face that presented a greater cardiac component. Mainly, the forehead and cheeks were the areas that achieved high cross-correlation values, in comparison with other regions of the face such as the nose and eyes.

The videos were recorded with two video cameras at different video resolutions, in which the results achieved by them presented noticeable differences. The low-resolution recordings obtained a wider variety of results between subjects on the measurement of the NN time series. The statistical results of the HR and HRV parameters also presented differences among them. The HR, NNmean, SDNN, and frequency-domain parameters achieved higher

correlation results and, in contrast, the RMSSD and pNN50 components obtained a lower correspondence.

In the case of the high-resolution recordings, the results were considerably better. Most of the subjects obtained correlation coefficients above 0.9 on the measurement of the NN time series. Also, high correlation values were obtained on the HR and HRV parameters with values above 0.9 in most of them. In this case, although the RMSSD and pNN50 components achieved good statistical results, they presented the lower correspondence among all parameters. In general, the correlation results of the HR and HRV parameters measured with the Logitech camera were improved in comparison with the corresponding data obtained in the studies of Moreno et al. [27] and Sun et al. [24].

Despite the higher frame rate of the low-resolution recordings, this feature did not yield better results. Taking into account that the same method was carried out to obtain the measurements of the low-resolution and high-resolution recordings, the video resolution seems to be significant in the difference of results. A higher resolution presents a major number of pixels to be analyzed and, hence, more data to obtain reliable measurements. Moreover, the frame reduction by pixel averaging could affect the results of the low-resolution recordings more, since the amount of data available to obtain the iPPG signal was reduced.

Although the high-resolution recordings obtained good results, the employed method has the limitation that it can only be used in stationary conditions due to it identifies and evaluates fixed positions on the averaged frames to obtain the iPPG signal. Therefore, alternative methods are necessary to perform a video HRV analysis in motion conditions.

Chapter 5

Video HRV analysis in stationary and motion conditions

The state of the art shows that most of the works that have employed the video imaging technique have obtained HRV parameters in stationary conditions (refer to section 2.3.2), and there are practically no studies that obtain these parameters in motion scenarios and by conducting an in-depth statistical analysis. Therefore, this chapter proposes a selective tracking method using the Viola-Jones and KLT algorithms, with the aim of carrying out a robust video HRV analysis in stationary and motion conditions. Furthermore, given the importance of the sampling rate in a HRV analysis and the low temporal resolution of commercial cameras, an analysis of two models was carried out to evaluate their performance in the measurements.

5.1 Data acquisition set-up

A summary of the measurement set-up is shown below:

- 15 subjects (3 women, 12 men) between 23 and 35 years old were recorded with a Logitech webcam at 1920x1080 pixels and 15 fps (see section 3.1.2).
- 5 subjects (1 woman, 4 men) between 23 and 35 years old were recorded with a GoPro camera at 1280x720 pixels and 60 fps (see section 3.1.2).
- The analyzed recordings were obtained in stationary and motion conditions (refer to section 5.1.2).
- The recording length was 50 s.
- The distance between the camera and the face was 0.3 m approximately.
- The videos of both cameras were performed on simultaneous recording.
- The recordings were performed indoors with sunlight as the lighting source.

5.1.1 Performance analysis of video cameras

The first study consisted of a comparative analysis of two commercial cameras with different features in order to evaluate their performance in the measurements. In this analysis, 5 subjects (1 woman and 4 men) were assessed in stationary and motion conditions (refer to section 5.1.2).

One of the cameras (GoPro HERO3 silver edition) is a versatile model that has been widely used lately in different scenarios and has the capability of diverse video resolutions. Depending on the resolution, it is possible to record videos at different fps. A resolution of 1280x720 pixels was chosen in order to record at 60 fps, a higher number of fps than most of the commercial video devices are able to record. The aim of recording at a higher frame rate was to assess its influence on the measurements.

The other camera (Logitech HD Pro Webcam C910) also has the capability of different video resolutions, but with a lower number of fps. In this model, the videos were recorded at 1920x1080 pixels and 15 fps, which is the maximum frame rate achievable at this video resolution. This camera was chosen because it enables the manual adjustment of diverse parameters. The focus, gain, and exposure time were fixed in order to obtain an adequate focus and illumination of the face. Fixing these parameters also avoids automatic adjustments caused by the movements of the subject or ambient light changes. This manual configuration was not possible in the GoPro model.

5.1.2 Stationary and motion analysis

The recordings analyzed in this chapter were carried out in stationary and motion conditions. Firstly, a video was recorded asking the participants to remain still throughout the acquisition. Since the application of a tracking algorithm and the influence of motion were under evaluation, it was necessary to conduct an analysis in stationary conditions for comparison. Secondly, the participants were asked to perform lateral and forward/backward movements, always looking toward the camera and avoiding fast gestures. A total of 15 subjects (3 women and 12 men) were analyzed in this chapter. The recordings of the Logitech webcam that were analyzed in chapter 4 are part of this study.

5.2 Video and signal processing

5.2.1 ROIs detection

In order to perform an automatic ROI detection, instead of the manual selection done in chapter 4, the Viola and Jones algorithm was employed to detect the ROIs on the video recordings (refer to section 2.5.1). Once the algorithm was implemented, some problems arose such as multiple, incorrect or misdetections

of face or eyes. As shown in Fig. 5.1, the algorithm can incorrectly detect more than one ROI due to the eyebrows (Fig. 5.1 (left)) and the nose region (Fig. 5.1 (right)) present a resemblance with the eyes area from the algorithm perspective. This occurs because the Viola and Jones approach employs Haar-like features that are compared in similarity to the image in terms of intensity (see Fig. 2.10), which may lead to an incorrect ROI detection.
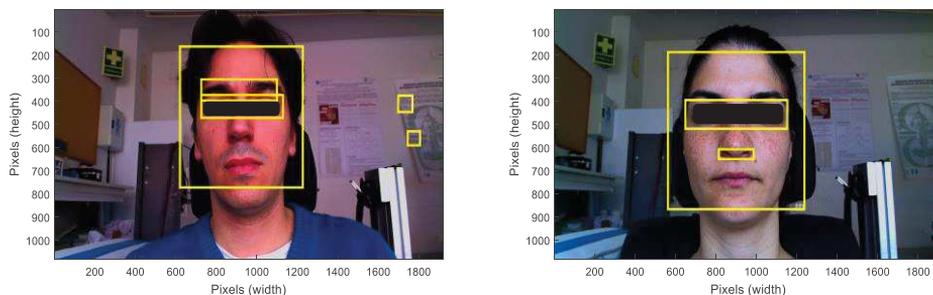
Fig. 5.1   Examples of multiple and incorrect ROIs detection using the Viola and Jones algorithm. The shown boxes aim to enclose the ROIs corresponding to the face and eyes regions.

Some conditions were then established on the basis of the four-element vector specified in pixels, which defines the top-left coordinate (x, y), width (w), and height (h) of the ROI. The w and h of the ROI of the face ($ROIf$) and eyes ($ROIe$) were verified within the following pixel values:

$$450 > ROIf_w > 800, \quad 450 > ROIf_h > 800$$

$$250 > ROIe_w > 550, \quad 50 > ROIe_h > 150$$

If the w or h value failed to meet the criterion, the ROI was rejected. In this case, a new detection attempt was carried out in the following frame until both ROIs were found. Fig. 5.2 shows examples of the ROIs selected in two subjects based on the established pixel values of w and h.
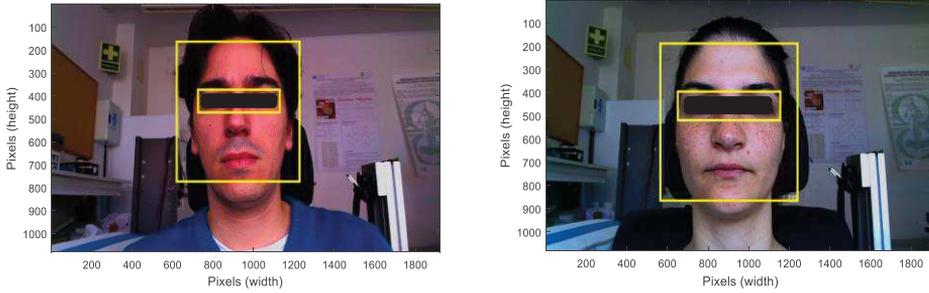
Fig. 5.2 Selected ROIs in two subjects based on the established values of width (w) and height (h).

If both ROIs were within the established values, the four-element vectors [x, y, w, h] were modified according to the following percentages:

$$ROIf = [ROIf_x + 0.2 \; ROIf_w, \quad 1.1 \; ROIf_y, \quad 0.6 \; ROIf_w, \quad 0.6 \; ROIf_h]$$

$$ROIe = [ROIe_x - 0.1 \; ROIe_w, \quad 0.85 \; ROIe_y, \quad 1.2 \; ROIe_w, \quad 1.8 \; ROIe_h]$$

The coordinates were modified with the aim of analyzing only the forehead and cheek regions (Fig. 5.3). Moreno et al. [27] showed that the forehead and cheeks have a greater cardiac component in comparison with other regions of the face. By using the coordinates of the *ROIe*, the eye area was excluded in order to reduce the artifacts produced by blinking.
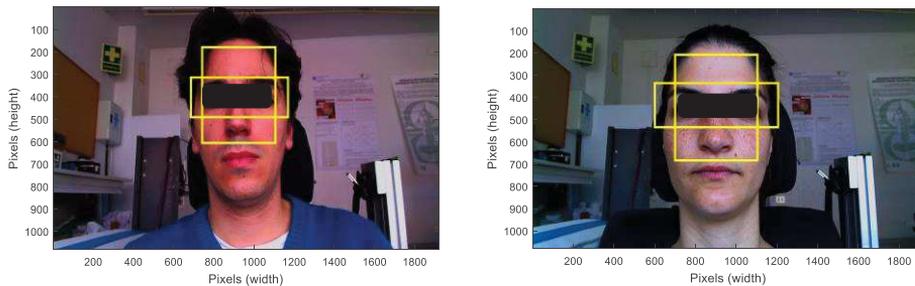
Fig. 5.3   Modified ROIs in two subjects according to the established percentages of the initial four-element vectors [x, y, w, h].

5.2.2 Tracking of feature points

The KLT algorithm was used to perform the tracking of the face and eyes along the video sequence (refer to section 2.5.2). In a study of the state of the art, Huang and Dung [32] proposed the Viola and Jones algorithm to detect the ROI in each frame of the video. This ROI detection method can be computationally expensive and not always robust, especially if the subject makes some type of movements. Therefore, the KLT algorithm was applied in this work as a more efficient tracking method. This approach enables the tracking of feature points in a video sequence, which gives the possibility to detect the ROI in each frame, even if the subject performs a tilting of the head (Fig. 5.4).
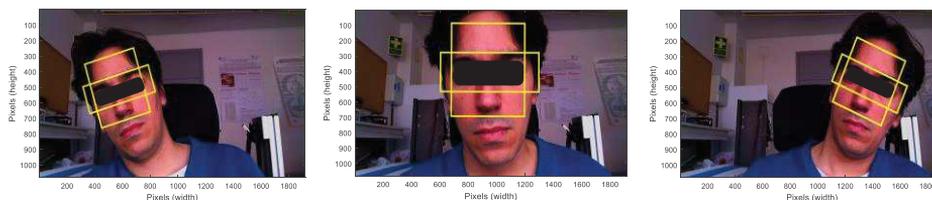
Fig. 5.4   Tracking of the face and eyes in a subject: (left) lateral movement to the right; (middle) forward movement; (right) lateral movement to the left.

The ROI size is adaptable in accordance with movements of the head or some facial expressions. For this reason, the feature points detected within the *ROIe* were eliminated to avoid changes in the ROI size caused by blinking, which may affect the measurements (Fig. 5.5). A great loss of feature points did not occur during the tracking, but the performing of faster and sudden movements would likely result in a major loss of points. In such a case, a reacquisition of feature points would be necessary in order to ensure a reliable tracking of the ROI.
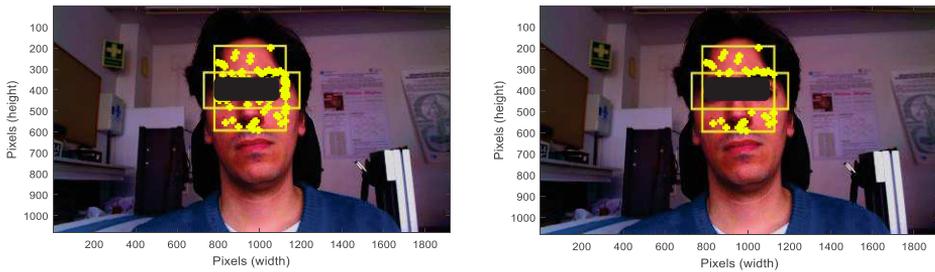
Fig. 5.5   (Left) frame with feature points (in yellow) detected within the *ROIf*; (right) frame without feature points within the *ROIe*.

5.2.3   Video image processing and signal acquisition

The video image processing was aimed at analyzing the pixels corresponding to the skin and excluding regions that may contribute with artifacts. The frames were therefore converted into threshold-based binary images to highlight the skin from darker areas. Thus, the pixels within the *ROIf* and corresponding to the skin were analyzed and areas such as hair, eyebrows, and beard were excluded from the analysis. The pixels outside the *ROIf* and within the *ROIe* were also excluded. The corresponding image results were multiplied to obtain the final image to be processed, as illustrated in Fig. 5.6.
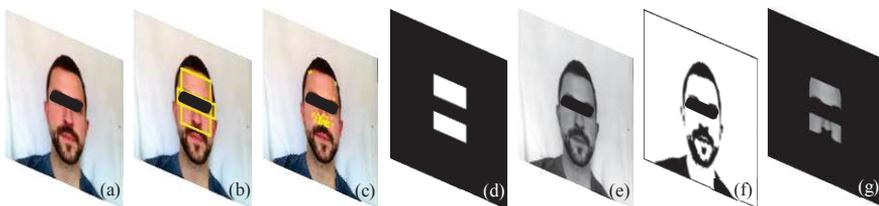
Fig. 5.6   Video image processing: (a) original frame; (b) *ROIf* and *ROIe*; (c) feature points detected on frame; (d) ROI filter; (e) green channel of frame; (f) binary image; (g) resulting image of multiplication of (d), (e), and (f) images.

In the resulting image, all non-rejected pixels (pixels with non-zero values) were averaged across all the video sequence to obtain a raw iPPG signal. The

filtering, interpolation, and further signal processing steps were carried out according to the section 4.2.3.

5.3 Results and discussion

5.3.1 Performance analysis of video cameras

Fig. 5.7 shows the Pearson's correlation coefficient of the NN measurements obtained by the two cameras in stationary and motion conditions. Table 5.1 shows the statistical results of the NN measurements achieved by both cameras. Based on data results, it should be pointed out that the Logitech model achieved better results than the GoPro camera in both conditions. Moreover, the results of the webcam were similar in both conditions of the study.
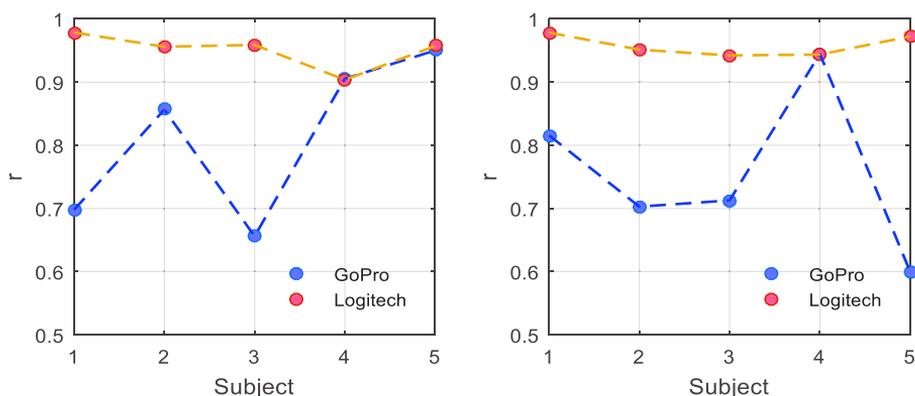
Fig. 5.7 Pearson's correlation coefficient (r) of the NN measurements obtained by the cameras: (left) stationary conditions; (right) motion conditions. All p-values < 0.001.

A determining factor that greatly affected the recordings obtained by the GoPro camera was the lighting condition on a partly cloudy day. Although recordings were carried out indoors with sunlight as the only source of illumination, some unpredictable lighting changes influenced some video

recordings. After video processing, it was observed that some of the signals obtained by the GoPro model were more affected by these illumination disturbances. Because of the small changes in skin color are being measured, these disturbances cause alterations in the signal that make the acquisition of reliable measurements difficult. A comparative illustration of the iPPG signals obtained by the two cameras under lighting disturbances is shown in Fig. 5.8.

Table 5.1   Statistical results of the NN measurements obtained by the cameras ($\bar{x} \pm s,\ N = 5$)

STATIONARY CONDITIONS					
Camera	**r**	**ICC**	**MAE**	**MAPE**	**RMSE**
GoPro	0.813 ± 0.129	0.794 ± 0.154	23.65 ± 8.58 ms (2.10 ± 0.61 bpm)	2.83 ± 0.75 %	36.16 ± 21.70 ms (3.14 ± 1.18 bpm)
Logitech	0.951 ± 0.028	0.950 ± 0.028	12.55 ± 5.66 ms (1.16 ± 0.72 bpm)	1.54 ± 0.82 %	17.03 ± 8.97 ms (1.57 ± 1.08 bpm)
MOTION CONDITIONS					
Camera	**r**	**ICC**	**MAE**	**MAPE**	**RMSE**
GoPro	0.755 ± 0.131	0.729 ± 0.151	34.76 ± 12.11 ms (2.80 ± 1.32 bpm)	4.03 ± 1.69 %	45.54 ± 18.53 ms (3.66 ± 1.92 bpm)
Logitech	0.958 ± 0.017	0.957 ± 0.017	13.23 ± 4.52 ms (1.09 ± 0.48 bpm)	1.54 ± 0.59 %	16.76 ± 5.52 ms (1.39 ± 0.59 bpm)

Refer to section 3.2 for statistical parameters definitions.

Although the lighting condition is an external factor, it is very important to take into account because it may affect the camera sensors differently. Lighting conditions should also be taken into consideration if the application in real environments is contemplated. Additionally, a slight blur in the videos obtained by the GoPro model was observed that could affect the measurements. This issue arose because the face was positioned at a short distance from the camera and the impossibility to adjust the focus. The Logitech model did not present this problem because the focus was manually adjusted before the recordings.

Despite the difference in the frame rate of the cameras, it appears that this parameter has little effect on the measurements. As shown, the signals acquired with the higher frame rate did not yield better results, at least in stationary conditions and without the presence of lighting disturbances. Thus, the better overall results obtained by the webcam and the possibility of making manual settings were decisive for choosing this camera in the later recordings.
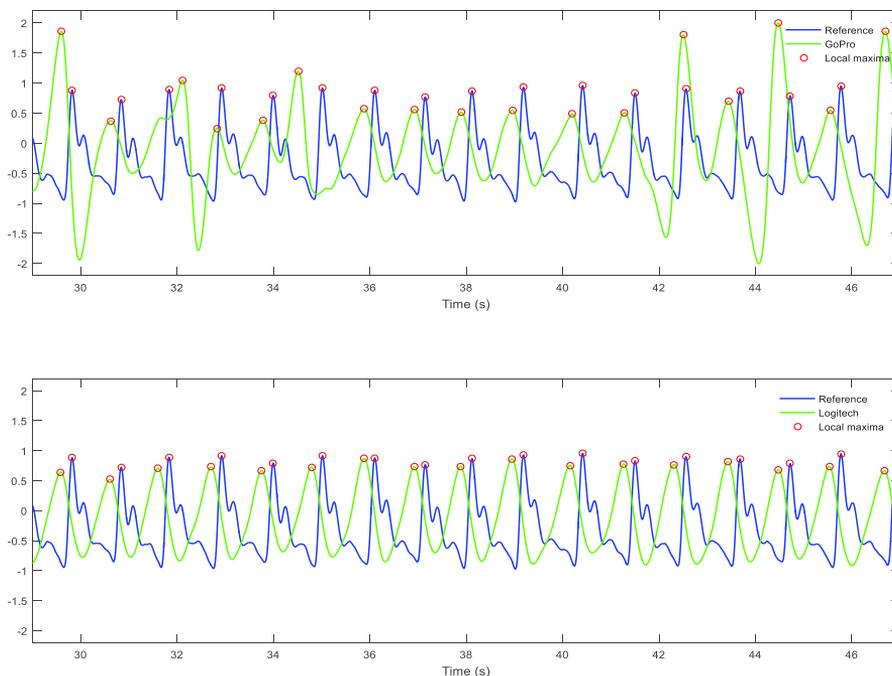
Fig. 5.8   iPPG signals obtained by the cameras in simultaneous recording under lighting disturbances: (top) GoPro signal; (bottom) Logitech signal.

5.3.2   HRV analysis in stationary conditions

Once the first part of the study was completed, a HRV analysis was conducted under the same conditions using the chosen camera (Logitech model). Table 5.2 shows the statistical results of the NN measurements obtained by each subject in stationary conditions. As shown, most of the subjects achieved correlation

coefficients above 0.9. The $\bar{x} \pm s$ of the individual data presented Pearson and ICC measures of 0.937 ± 0.035 and 0.935 ± 0.036, respectively. The mean of the errors achieved a MAPE = 1.83 %, a MAE = 15.45 ms, and a RMSE = 19.99 ms. In general, these mean results are slightly better that those reported in chapter 4 using the same video records.

Table 5.2   Statistical results of the NN measurements in stationary conditions

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,978	0,979	0,42	-0,02	10,91	0,56	1,00	14,57	0,73
2	0,956	0,956	0,41	-0,04	9,30	0,74	1,07	12,32	0,99
3	0,959	0,956	0,04	-0,03	7,69	0,90	1,07	9,51	1,11
4	0,904	0,903	-0,03	-0,10	22,09	2,36	2,95	32,41	3,42
5	0,957	0,955	0,16	-0,07	12,73	1,26	1,63	16,33	1,61
6	0,970	0,970	-0,43	0,02	14,43	1,15	1,66	17,86	1,46
7	0,906	0,896	-0,58	-0,13	23,71	2,40	3,06	29,31	3,11
8	0,958	0,956	-0,45	-0,01	13,42	1,30	1,70	17,06	1,68
9	0,952	0,951	-0,56	-0,01	17,33	1,33	1,96	22,51	1,75
10	0,939	0,934	0,25	-0,11	15,50	1,62	2,04	19,61	2,08
11	0,932	0,933	1,52	-0,07	18,48	1,00	1,75	23,09	1,27
12	0,949	0,950	0,44	-0,02	17,84	0,85	1,59	22,43	1,09
13	0,951	0,950	1,71	-0,12	12,33	0,76	1,24	16,68	1,03
14	0,846	0,845	-1,38	0,08	22,06	2,41	2,97	27,95	3,12
15	0,898	0,893	0,29	-0,08	14,00	1,40	1,81	18,27	1,85
Mean	0,937	0,935	0,12	-0,05	15,45	1,34	1,83	19,99	1,75
S.D.	0,035	0,036	0,79	0,06	4,76	0,62	0,68	6,35	0,84

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

Some participants were observed to have achieved better results than others. Since the analysis was carried out in stationary conditions, some factors such as the different anatomical and physiological characteristics of the participants may have influenced the measurements. Kumar et al. [30] performed PRV estimations for different skin tones (fair, olive, and brown) in which 4 subjects were analyzed according to the skin category. The fair and olive skin tones presented quite similar results with RMSE values of 13.61 ms and 13.36 ms,

respectively, while for the brown category a higher RMSE of 48.91 ms was obtained. Based on the results obtained from the fair and olive skin categories and the similarity of fair skin tone of the subjects in our study, the differences in the results between subjects may not be due to this physical characteristic.

There are other factors besides skin color which may also influence the measurements such as blood pressure, heart rate, respiration, etc. Moreover, although the participants were asked to keep still during the recording, some had difficulties in remaining motionless or avoiding facial expressions. Sunlight affected some subjects causing a major blinking in them.

Some authors have reported statistical results of the NN measurements (or IBIs) obtained in stationary conditions [26, 29–31], whose corresponding results are summarized in Table 2.3. By examining these data and the corresponding mean results presented in Table 5.2, it is noted that our method achieved a good performance. The NN time series measured by our method achieved a higher correlation in comparison with the result presented by Bousefsaf et al. [29]. With regard to the error results reported in these works, only the study of Kumar et al. [30] achieved a lower error compared with the 19.99 ms obtained by our method. It is important to note that our study presented a larger sample size than the cited works. Thus, the sample size should also be taken into account since the results may vary considerably between subjects, even if there is a minimal presence or an absence of motion artifacts (Table 5.2).

Fig. 5.9 shows the correlation scatter plots of the HR and HRV parameters obtained in stationary conditions, and Table 5.3 presents their corresponding statistical results. High correlation values were obtained by the proposed method, most of which achieved results above 0.9. In general, these parameters present similar statistical results than those obtained with the Logitech camera in chapter 4. Taking into account this, the conclusions obtained in contrast with the

studies discussed in section 4.4.1 are very similar (Moreno et al. [27] and Sun et al. [24]). In comparison with the results presented by Poh et al. [23], the RMSE of HR was reduced from 1.24 bpm to 0.07 bpm and the frequency-domain parameters also presented lower errors.
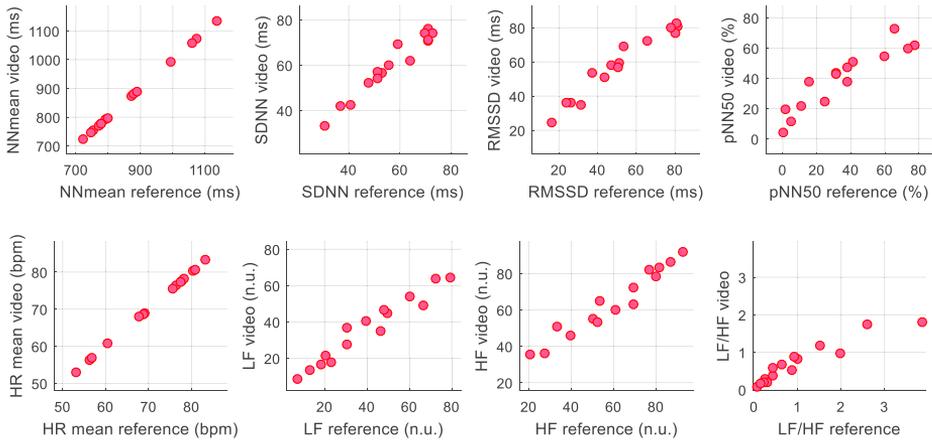
Fig. 5.9   Correlation scatter plots of the HR and HRV parameters obtained in stationary conditions. Refer to Table 5.3 for r and ICC results.

Table 5.3   Statistical results of HR and HRV parameters in stationary conditions

Parameter	r*	ICC*	MAE	RMSE
HRmean (bpm)	1,000	1,000	0,06	0,07
NNmean (ms)	1,000	1,000	0,58	0,77
SDNN (ms)	0,975	0,948	3,64	4,40
RMSSD (ms)	0,974	0,911	7,61	8,92
pNN50 (%)	0,925	0,877	9,79	11,46
LF (n.u.)	0,970	0,930	5,65	7,67
HF (n.u.)	0,970	0,930	5,63	7,66
LF/HF	0,944	0,724	0,35	0,65

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

McDuff et al. [25], besides the HR measurement, they presented the results of the frequency-domain parameters of PRV at rest and under stress. They used a five band digital camera in which the combination of cyan, green and orange bands yielded a higher correlation with the reference sensor. Curiously, the results under stress achieved higher correlation values than those obtained at rest. The best results obtained of this work were similar in comparison with the data achieved in our study, in which only the green channel of the video was analyzed.

The study of Alghoul et al. [28] presented a comparison between two approaches to measure HRV parameters from the face in stationary conditions. On one side, the results reported in the study showed that ICA-based method yielded lower errors in the HF and LF/HF components and, on the other hand, the LF parameter achieved better results with the EVM-based approach. Although the correlation values presented by Alghoul et al. were lower than those achieved by our method, some corresponding errors were better than the data in our study, which appear to make no sense, likely, due to a transcription error.

Huang and Dung [32] only presented single results of HRV parameters using absolute error measures. Therefore, the MAE results shown in our study were calculated to obtain a mean value of the individual results. These data are slightly better than the corresponding results achieved by our method, but using a smaller sample size in comparison with our analysis. Thus, the application of the proposed method by Huang and Dung appears promising in the acquisition of measurements in stationary conditions. It would be interesting the application of this approach to estimate time and frequency domain parameters of HRV with a larger sample size.

5.3.3   HRV analysis in motion conditions

Although HRV is normally measured at rest, recently, the interest of measuring physiological variables in everyday activities by using alternative methods has increased. Most of the daily life scenarios present the motion inherently, which must be considered because it may affect the measurements, especially when is used the video imaging technique. For example, the drowsiness in drivers is one the causes of traffic accidents all over the world. Therefore, some works [124, 125] have proposed the measurement of HRV parameters in drivers during alert and drowsy or fatigued periods, in which some parameters showed significant differences between both states.

Other studies have proposed the analysis by video to measure physiological parameters while driving [126–128] and other ones to detect cardiac arrhythmias [129, 130]. Thus, the video imaging technique may eventually become a method to detect these events in drivers with the aim of preventing traffic road accidents. Also, since the HRV analysis is an excellent indicator of physical and psychological stress, this technique may be used to evaluate the stress level of people during working hours or as part in a polysomnography study.

In this part of the study, the same analysis was conducted as that for stationary conditions. Table 5.4 shows the statistical results of the NN measurements obtained by each subject in motion conditions. In general, the correlation results were lower in comparison with the stationary conditions, but they also varied according to the participant; even so, results of above 0.9 were obtained with several subjects. The $\bar{x} \pm s$ of the individual data presented Pearson and ICC measures of 0.912 ± 0.048 and 0.910 ± 0.050, respectively. The mean of the errors achieved a MAPE = 2.00 %, a MAE = 17.26 ms, and a RMSE = 21.64 ms. In comparison with the stationary conditions, these mean results were not greatly affected by the presence of motion.

Table 5.4   Statistical results of the NN measurements in motion conditions

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,978	0,979	0,14	-0,01	11,90	0,55	1,04	15,27	0,69
2	0,951	0,952	0,07	-0,01	9,17	0,81	1,11	11,98	1,08
3	0,942	0,940	-0,02	-0,03	10,67	0,99	1,33	13,72	1,29
4	0,944	0,944	0,52	-0,04	20,76	1,78	2,46	26,11	2,23
5	0,972	0,973	0,36	-0,05	13,64	1,33	1,74	16,71	1,68
6	0,972	0,970	0,64	-0,16	14,34	1,28	1,74	19,47	1,72
7	0,901	0,899	-0,74	0,00	18,54	1,79	2,34	25,07	2,59
8	0,926	0,927	-0,30	0,00	18,12	1,49	2,13	22,54	1,86
9	0,886	0,888	-0,07	0,02	19,84	1,83	2,46	23,46	2,17
10	0,878	0,879	0,68	-0,07	15,75	1,67	2,09	19,11	2,02
11	0,906	0,904	1,85	-0,12	15,98	0,87	1,52	19,15	1,03
12	0,829	0,825	-1,96	0,06	30,84	1,60	2,87	36,45	1,88
13	0,884	0,885	0,78	-0,07	17,37	1,19	1,85	23,12	1,62
14	0,865	0,866	0,61	-0,10	26,18	2,49	3,27	32,09	3,14
15	0,844	0,823	0,65	-0,13	15,73	1,53	2,00	20,31	1,98
Mean	**0,912**	**0,910**	**0,22**	**-0,05**	**17,26**	**1,41**	**2,00**	**21,64**	**1,80**
S.D.	**0,048**	**0,050**	**0,84**	**0,06**	**5,67**	**0,49**	**0,63**	**6,58**	**0,63**

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

In comparison with the stationary posture, it is noted how some subjects obtained similar results; on the other hand, some subjects did not present results as good as those under the stationary condition. The motion conditions of our study consisted of lateral and forward/backward movements that added the difficulty of obtaining the iPPG signal at different positions and distances from the camera. These are common movements that are performed while working in front of a PC or in other activities with similar moves. Moreover, these movements presented the particularity that they were performed at a steady and non-stop pace.

The employed tracking algorithm performed very well in accordance with the different movements carried out by the subjects. However, despite the participants were asked to execute the same type of movements, in practice, their gestures tended to vary slightly; the type of head inclination, the pace of the

movements or the distance between the camera and the face during the forward/backward motions are some examples of these variations. Thus, it should be pointed out that some particular movements may significantly affect the iPPG signals, thereby making the acquisition of NN measurements more difficult.

To date, most of the works of the state of the art have obtained HRV measurements in stationary conditions, therefore, the number of studies that have performed an in-depth statistical analysis in motion conditions is very limited. In the literature, there are works that have measured physiological parameters by video in motion scenarios [36, 131–134], but with assessments that make difficult a comparative analysis with the studies reviewed in this thesis. In other works, some authors have reported statistical results of the IBIs obtained in motion conditions [29–31], whose corresponding results are summarized in Table 2.3.

Bousefsaf et al. [29] obtained IBIs measurements with predefined head movement conditions. Although the results obtained in a sitting still and calm condition were better, it seems that the motion scenario did not affect to a great extent the measurements. These results obtained a lower correlation than that presented in Table 5.4. The motion-tolerant method proposed by Bousefsaf et al. appears to perform well in both conditions, although the low video resolution of the recordings could have influenced the measurements.

The study of Kumar et al. [30] performed the acquisition of signals in three natural motion scenarios. The reading scenario achieved an RMSE of 55.34 ms and the watching video and talking activities obtained the higher errors (see Table 2.3). Although none of the three motion scenarios is equivalent to the motion conditions of our study, the reading scenario may have a closer resemblance due to a greater presence of moderated movements. This may

explain the better result achieved by this scenario. Nevertheless, a reasonably higher error was obtained in comparison with the result achieved in our study (21.64 ms). Besides the difference of the motion scenarios, it is also important to take into consideration the difference in the sample size of both studies, as mentioned in other cases.

Antink et al. [31] performed a beat-to-beat estimation by means of different signals and their fusion. Within these signals, we have focused on the video signal obtained from the changes of the skin color. The conditions of the third trial of the study consisted in the acquisition of measurements during reading without further instructions, unlike the second trial (reading-task without motion). Thus, the difference in results between the trials 2 and 3 is, likely, because of the presence of motion artifacts during the reading task. The MAE results shown in Table 2.3 are calculated in this work to obtain a mean value of the individual results reported by Antink et al. in each trial. The MAE achieved by our method in motion conditions was lower than the mean error calculated in the third trial, as well as in comparison with the recordings performed without motion.

Likewise, the acquisition of HR and HRV parameters was carried out for purposes of comparison with data reported in other works and the results obtained in stationary conditions. Fig. 5.10 shows the correlation scatter plots of the HR and HRV parameters obtained in motion conditions, and Table 5.5 presents their corresponding statistical results. Most of the parameters obtained correlation results above 0.9, but with lower values in comparison with the stationary condition. The RMSSD, pNN50, and frequency-domain parameters yielded the lower correlation results, in particular, the LF/HF ratio that was the more affected component in both conditions.
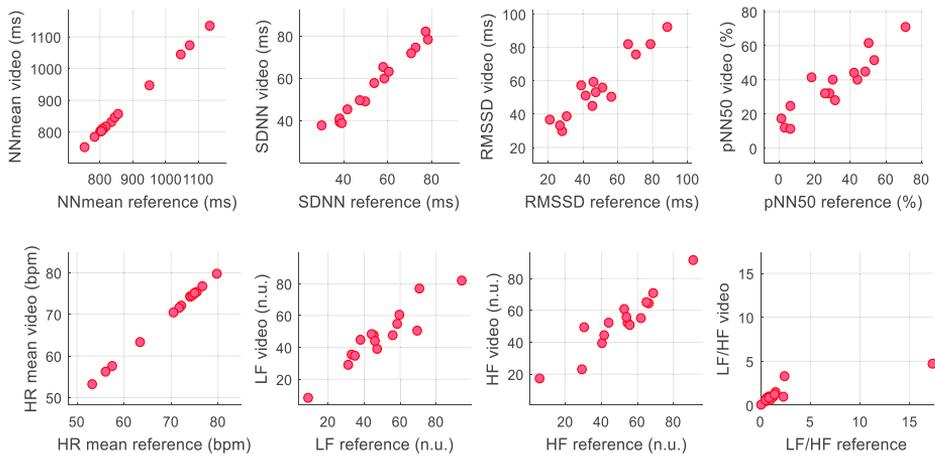
Fig. 5.10   Correlation scatter plots of the HR and HRV parameters obtained in motion conditions. Refer to Table 5.5 for r and ICC results.

Table 5.5   Statistical results of HR and HRV parameters in motion conditions

Parameter	r	ICC	MAE	RMSE
HRmean (bpm)	1,000*	1,000*	0,06	0,07
NNmean (ms)	1,000*	1,000*	0,62	0,84
SDNN (ms)	0,986*	0,970*	2,91	3,77
RMSSD (ms)	0,943*	0,886*	7,97	9,59
pNN50 (%)	0,921*	0,861*	7,76	10,20
LF (n.u.)	0,937*	0,929*	5,18	7,23
HF (n.u.)	0,938*	0,931*	5,13	7,14
LF/HF	0,858*	0,449**	1,08	3,28

*: $p < 0.001$; **: $p < 0.05$. Refer to section 3.2 for statistical parameters definitions.

Huang and Dung [32] recently presented a study with measures of HRV parameters obtained during occasional and frequent motion. Only single results of HRV parameters using absolute error measures were reported in their study. The sample size of the motion categories were $N = 3$ and $N = 2$, respectively, which were performed by the same subject. In comparison with both categories,

the MAE results of our study were lower than the data calculated from the results reported by them.

The occasional motion category consisted of three different recordings. One of them presented motions that were very similar to some of the movements performed in our study, in which the subject moved away from the camera and then moved back. In another recording, the subject shook the head three times but it was not specified how these movements were performed and, in the last one the subject talked and turned the head. In the case of the frequent motion category, the two recordings presented movements that were repeated along the video sequence.

In the case of the occasional motion category, this presented a higher MAE in comparison with the corresponding data of the frequent motion condition. This occurred because of the third recording of the occasional motion category obtained a considerable absolute error of 28.87 ms, likely, due to the presence of motion artifacts caused by the talking scenario. The face detection method could also affect the measurements because it is not always robust if the subject makes some type of gestures. Moreover, the small sample size with one subject and the motion conditions of the recordings, which were completely different from each other, make the data not suitable for obtaining reliable statistical results.

5.3.4   Bland-Altman plots

Fig. 5.11 shows the Bland-Altman plots with LoA of 95% representing the agreement between the NN measurements obtained by the reference system and the video in stationary and motion conditions. The stationary posture achieved LoA from -41.25 to 41.37 ms (-3.97 to 3.87 bpm), and the motion scenario obtained LoA from -43.69 to 44.15 ms (-3.84 to 3.74 bpm). No large differences

were found between both conditions of the study, and also, no systematic errors were identified in the measurements.
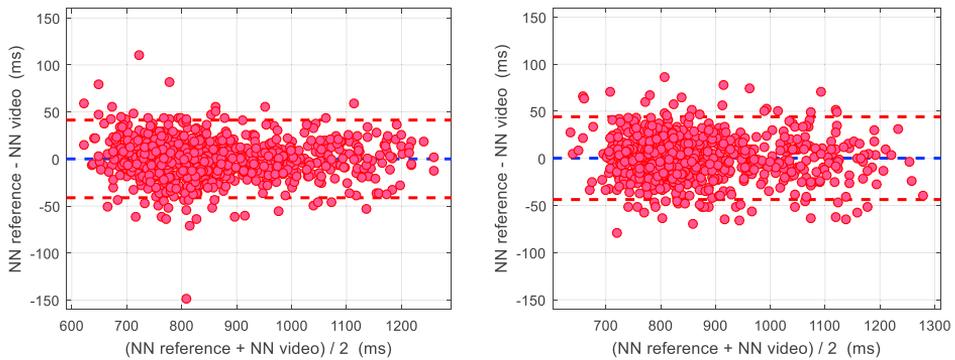
Fig. 5.11   Bland-Altman plots with LoA of 95% representing the agreement between the reference and video (data of all subjects combined in one dataset): (left) stationary conditions with a mean difference of 0.06 ms (-0.05 bpm) and LoA from -41.25 to 41.37 ms (-3.97 to 3.87 bpm); (right) motion conditions with a mean difference of 0.23 ms (-0.05 bpm) and LoA from -43.69 to 44.15 ms (-3.84 to 3.74 bpm).

The LoA obtained in stationary conditions were slightly improved in comparison with the results achieved in chapter 4 using the same camera (from -44.72 to 45.07 ms (-4.30 to 4.16 bpm)). Thus, the method proposed in this chapter seems a good alternative to obtain HRV measurements in stationary conditions, as well as in motion since no large differences were obtained between both scenarios. According to the Bland-Altman analysis, if the differences obtained by the measurement systems are not regarded as clinically important, both may be used interchangeably.

5.4 Conclusions

In this chapter, a video HRV analysis was carried out in stationary and motion conditions. An initial performance analysis of cameras validates the use of the webcam as a better option in this study. In the GoPro camera, it is noted a limitation to set parameters such as the focus or the automatic gain control that influenced its performance in comparison with the Logitech model. Some feature differences such as these could be related to the aim for which the cameras were designed. It is shown that it is possible to obtain good results by using a video recorded at a lower frame rate. Moreover, the higher the frame rate, the higher the computational cost of the video processing. In addition to analyzed factors such as frame rate and motion, it is also shown that the measurements could be affected by some physiological characteristics of the participants, lighting conditions, focus, resolution or the measuring distance.

The statistical analysis shows a good agreement between the reference system and the proposed method. In stationary conditions, the results of HRV parameters are improved by our method in comparison with data reported in related works, in which most of the parameters obtained correlation results above 0.9. Most of the HRV parameters also achieved good correlation results in the motion analysis, but with lower values in relation to the stationary condition. Although the participants were asked to perform the same type of movements, some differences such as the head inclination, the speed of the movements or the distance between the camera and the face during the forward/backward motions could also influence the results between the subjects. An overall comparative analysis of HRV parameters in motion conditions was more limited due to the lack of studies or studies containing insufficient data analysis. The larger sample size and the in-depth statistical analysis of our study provide greater reliability of the data.

The RMSSD, pNN50, and frequency-domain parameters yielded the lower correlation results, in particular, the LF/HF ratio that was the more affected component in both conditions of the study. Based on the similar results obtained using this approach in comparison with the method employed in chapter 4, the method proposed herein seems a good alternative for measuring HR or HRV parameters in stationary and motion conditions. Thus, the method appears to be a better option than the proposed approach in chapter 4, which requires additional video and signal processing and can only be applied in stationary conditions.

This study is a first assessment of the proposed method in motion conditions, with movements that can be performed while working in front of a PC or in other activities with similar moves. In general, the results between stationary and motion conditions do not differ significantly, although it is important to note that the measurements were carried out following specific movements. It is therefore necessary to conduct an evaluation of the proposed method considering a wider variety of motion, as well as the development of more robust algorithms in order to be applied in more realistic scenarios.

Chapter 6

Video HRV: real-time acquisition and lighting conditions analysis

In this chapter, a video HRV analysis is performed through the iPPG signal acquisition in real-time. A real-time acquisition has some advantages such as a quick video HRV analysis and the assessment of video parameters, motion or lighting influences to the iPPG signal in real-time. The recordings were carried with three sources of illumination separately: sunlight, LED light, and fluorescent light. The lighting conditions analysis allows the evaluation of different sources of illumination on the measurements and, based on the results, the choice of the best lighting source in order to perform a video HRV analysis.

6.1   Data acquisition set-up

A summary of the measurement set-up is shown below:

- 25 subjects (12 women, 13 men) between 22 and 54 years old were recorded with a Logitech webcam at 1280x720 pixels and 30 fps (see section 3.1.2).
- The videos frames were obtained via the MATLAB® software.
- The recordings were obtained in stationary conditions (see Fig. 3.1).
- The recording length was 5 min.
- The distance between the camera and the face was 0.3 m approximately.
- The recordings were performed indoors using three lighting sources:
  - Sunlight
  - LED light (*bulb*: Verbatim, ref. 52130; *color*: warm white; *wattage*: 6.5 W; *luminous flux*: 480 lm; *beam angle*: 130º, *CCT*: 3000 K).
  - Fluorescent light (*bulb:* Pro-Lite, ref. HELIXT2/20W/ES/64; *wattage:* 20 W; *luminous flux:* 875 lm, *CCT*: 6400 K).
- The technical characteristics of the PC used to record the measurements are: *processor*: Intel® Core™ i7 @ 3.40 GHz; *RAM*: 8 GB; *operating system*: Windows 7 Enterprise 64 bits; *graphics card*: Intel® HD Graphics 4000.

6.1.1   Video camera settings

First, the webcam settings were established in order to ensure the video acquisition in real-time. A video input object was created using the *videoinput* function of MATLAB®. This object is the connection between MATLAB® and the video camera. It is necessary to specify the adaptor name available on the system, in our case *winvideo* (Windows video), to carry out the communication

with the camera. By default, the acquisition parameters of the video input object were:

```
Summary of Video Input Object Using 'Logitech HD Pro Webcam C910'.

 Acquisition Source(s): input1 is available.

Acquisition Parameters: 'input1' is the current selected source.
                        10 frames per trigger using the selected source.
                        'RGB24_1280x720' video data to be logged upon START.
                        Grabbing first of every 1 frame(s).
                        Log data to 'memory' on trigger.

   Trigger Parameters: 1 'immediate' trigger(s) on START.

               Status: Waiting for START.
                       0 frames acquired since starting.
                       0 frames available for GETDATA.
```

Some of these default parameters were changed before the video acquisition. The number of frames per trigger was set to 1 in order to acquire only one frame when it is executed. The trigger was executed until the *stop* function was called to stop the video input object. The trigger type was set to *immediate,* so the trigger was executed automatically after calling the *start* function in order to use the video camera associated with the video input object.

Using the *getselectedsource* function, some specific camera properties were set as summarized below:

```
Display Summary for Video Source Object:

    General Settings:
      Parent = [1x1 videoinput]
      Selected = on
      SourceName = input1
      Tag = [0x0 string]
      Type = videosource

    Device Specific Properties:
      BacklightCompensation = off
      Brightness = 128
      Contrast = 128
      Exposure = -7
      ExposureMode = auto
```

```
Focus = 68
FocusMode = manual
FrameRate = 30.0000
Gain = 64
HorizontalFlip = on
Saturation = 128
Sharpness = 128
Tilt = 0
VerticalFlip = off
WhiteBalance = 4000
WhiteBalanceMode = manual
Zoom = 100
```

6.1.2 Lighting conditions

Due to the lighting conditions may affect the signal acquisition by video, it is proposed the assessment of different light sources. The face of the subjects was illuminated using three light sources separately: sunlight, LED light, and fluorescent light (Fig. 6.1). The room blinds were closed with the LED and fluorescent light sources to avoid the influence of sunlight. In these two light sources, a lamp with a shadow was used to power and hold the bulbs.

Fig. 6.1   A subject's face illuminated using different sources of illumination: (left) sunlight; (middle) LED light; (right) fluorescent light. The videos were recorded at 200 lux of illuminance with each lighting source.

The videos were recorded at 200 lux in order to have an adequate and the same illuminance of the face with the three lighting sources. To ensure this, before each recording, an Amprobe LM-100 light meter (refer to section 3.1.4) was utilized to measure the light intensity that reached the face. The subjects were allowed to close the eyes during the recordings obtained using artificial

light in order to avoid discomfort. Fig. 6.2 shows the spectrums of the sunlight, LED light and fluorescent light sources.
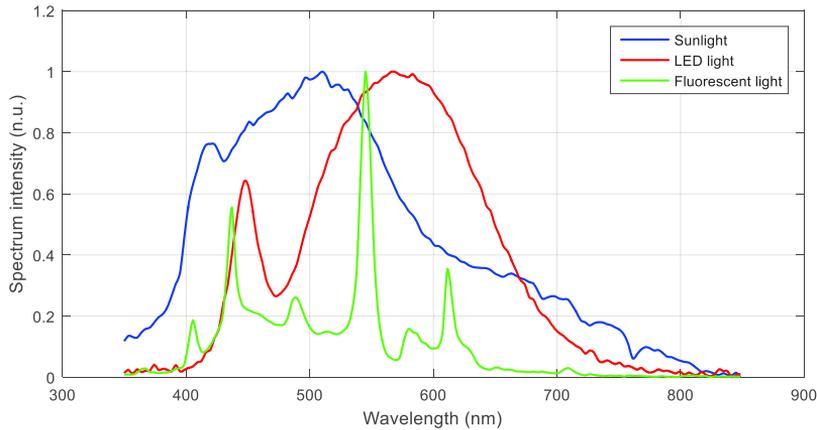
Fig. 6.2   Spectrums of the sunlight, LED light and fluorescent light sources.

6.2   Video and signal processing

6.2.1   ROIs detection

Once the camera settings were defined, the video acquisition began by using the *start* function. A first video frame was acquired by using the *getdata* function in order to detect the face and eyes (refer to section 5.2.1). As in previous chapters, the forehead and cheeks were analyzed because they have a greater cardiac component in comparison with other regions of the face [27].

6.2.2   Real-time acquisition

The video and signal acquisition started when both ROIs were found. The time stamp of each frame was stored to be used in the interpolation of the signal. The video frames captured were composed of the RGB components, but only the

green channel was analyzed because it has the best iPPG signal in comparison with the other channels [17, 18]. The video image and signal processing were performed as detailed in section 5.2.3. An example of the video image and the iPPG signal obtained in real-time are shown in Fig. 6.3.
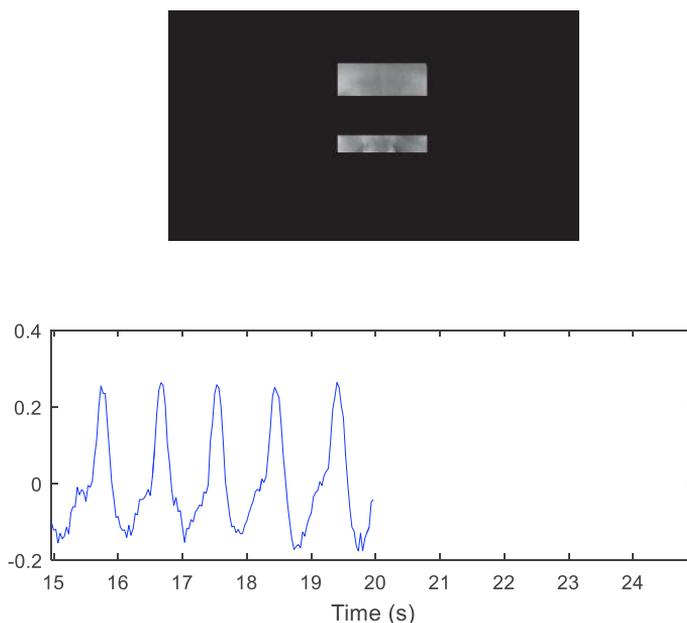
Fig. 6.3   Video and signal acquisition in real-time: (top) video image showing the face regions under analysis; (bottom) iPPG signal obtained from the video image.

Taking into account the aim of the real-time acquisition, it was important that the video and signal processing were performed within specified time conditions. Then, some test calculations were carried out to evaluate the video and signal acquisitions. The technical characteristics of the PC used to record the measurements are listed in section 6.1. The average calculation time to obtain the iPPG signal values from the video frames was around 3 ms, a value considerably lower than the theoretical sampling period of 33.3 ms, according to the frame rate of 30 fps.

The test calculations that were performed using the stored time stamp proved that the records were obtained at the specified recording length and frame rate. Theoretically, the time elapsed between frames should be 33.3 ms (0.0333 s) but, in practice, the sampling period was not constant (Fig. 6.4). It is noted that most of the frames were captured a slight instant before (around 0.032 s) the ideal time, as well as the sampling period suffered a periodic delay every 12 frames. Even so, the mean of the sampling period was 0.0333 s.
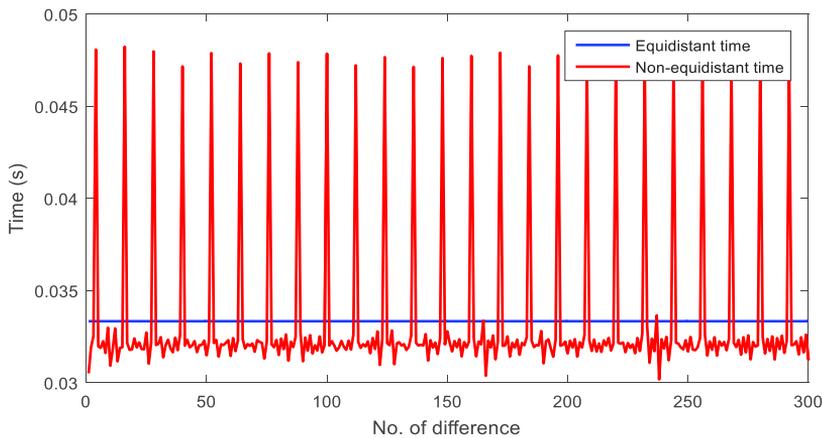
Fig. 6.4   Time between video frames obtained during the real-time acquisition in a subject. The sampling period was not constant (red plot) as the expected value of 0.0333 s (blue plot).

In order to evaluate the acquisition method, a test was carried out in which a sawtooth signal of 1 Hz was obtained from a function waveform generator and recorded by video at 30 fps from a green LED, similarly to the method explained in section 4.2.4. The sawtooth signal presents 30 samples between peaks of the signal, which ensures no loss of samples during the signal acquisition (Fig. 6.5). Representatively, Fig. 6.6 shows the signals obtained from a subject, in which the filtering and interpolation were carried out according to the section 4.2.3.
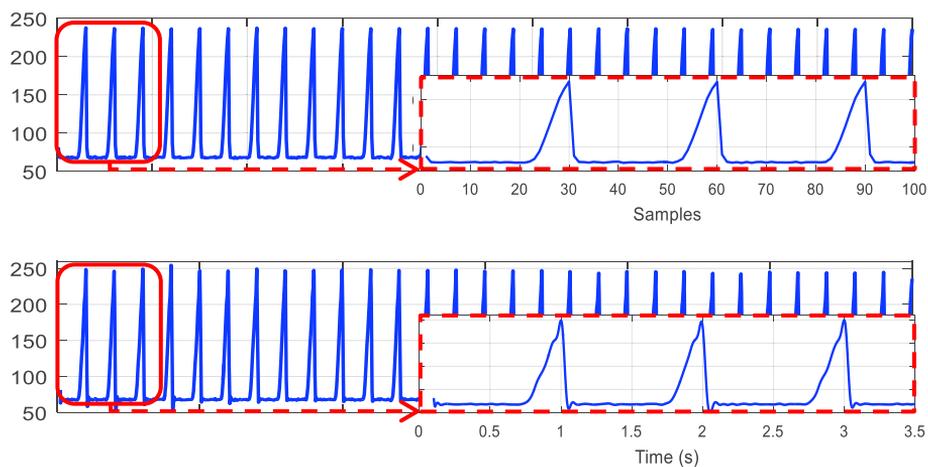
Fig. 6.5   (Top) sawtooth signal of 1 Hz obtained from a function waveform generator and recorded by video at 30 fps from a green LED; (bottom) sawtooth signal shown on top after interpolation to 1 kHz (the signals were shifted to visually match the peaks with the values in samples and seconds, respectively).
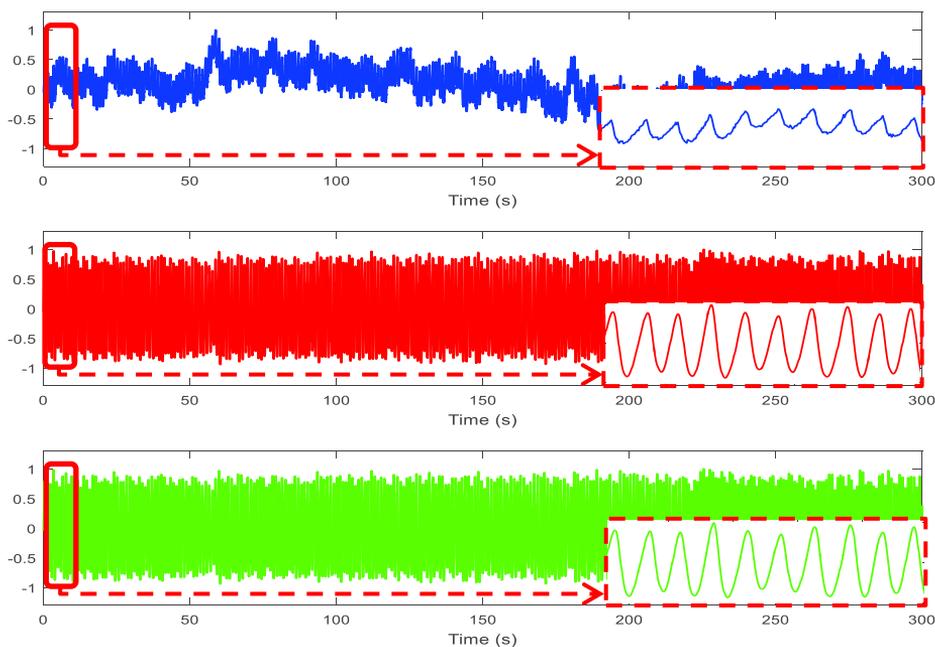
Fig. 6.6   iPPG signals of a subject obtained with the real-time acquisition: (top) raw signal; (middle) filtered signal; (bottom) interpolated signal.

6.2.3   Synchronization of the signals

The start time of the reference and video recordings were stored in order to perform the synchronization of the signals. These starting times corresponded to the computer's operating system in which were obtained both recordings. The corresponding data were obtained with the BSL PRO 3.7 (MP36 BIOPAC Systems, Inc. software) and MATLAB® for each recording. Thus, the difference between the respective start times was the required synchronization of the heart rate signals.

6.2.4   Artifacts correction

Once the signals were synchronized, the local maxima detection and artifacts correction were carried out according to the section 4.2.5. The NN intervals were obtained by calculating the time between consecutive local maxima. Fig. 6.7 shows a comparative of the NN time series obtained from a subject. The posterior HRV analysis was performed according to the section 4.3.
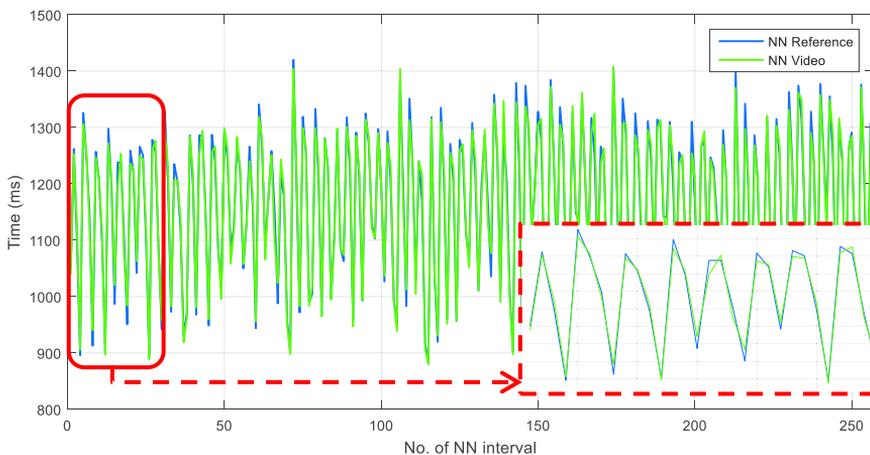
Fig. 6.7   NN intervals of reference and video obtained from a subject.

6.3   Results and discussion

6.3.1   Sunlight condition

Table 6.1 presents the statistical results of the NN time series obtained by each subject in sunlight condition. All subjects achieved correlation coefficients above 0.9. The $\bar{x} \pm s$ of the Pearson and ICC measures were 0.962 ± 0.023 and 0.961 ± 0.023, respectively. The MAPE achieved a mean value of 1.14 %.

Table 6.1   Statistical results of the NN time series by subject (sunlight condition)

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,970	0,970	0,29	-0,01	12,24	0,39	0,89	17,46	0,55
2	0,975	0,974	0,64	-0,09	11,70	0,92	1,33	16,68	1,33
3	0,972	0,970	0,01	-0,03	8,06	0,81	1,04	10,52	1,07
4	0,989	0,989	0,15	-0,04	8,68	0,74	1,03	11,14	0,95
5	0,934	0,929	0,02	-0,04	11,06	1,19	1,48	13,94	1,52
6	0,961	0,961	0,50	-0,04	9,45	0,66	1,01	13,01	0,90
7	0,947	0,945	0,10	-0,03	11,06	0,88	1,27	14,90	1,20
8	0,990	0,990	0,13	0,03	15,44	0,69	1,32	19,70	0,87
9	0,994	0,994	0,08	-0,01	7,70	0,43	0,74	9,74	0,56
10	0,945	0,943	0,07	-0,05	12,83	1,30	1,66	16,32	1,66
11	0,969	0,969	-1,80	0,13	11,03	1,00	1,36	18,19	1,67
12	0,957	0,957	-0,45	0,04	6,73	0,57	0,81	11,03	1,06
13	0,963	0,962	0,15	-0,02	9,03	0,78	1,08	11,07	0,97
14	0,965	0,965	0,08	-0,01	9,89	0,65	1,03	12,95	0,86
15	0,933	0,933	0,14	-0,02	9,13	0,66	1,00	11,48	0,83
16	0,913	0,909	0,08	-0,03	9,87	0,97	1,26	12,65	1,24
17	0,970	0,970	0,20	-0,01	10,45	0,54	0,97	13,56	0,71
18	0,939	0,939	-1,42	0,07	18,27	0,85	1,63	29,43	1,57
19	0,933	0,929	0,07	-0,02	8,26	0,70	0,98	10,73	0,91
20	0,927	0,927	1,16	-0,13	10,46	1,00	1,30	20,63	1,93
21	0,975	0,974	0,16	-0,03	9,22	0,71	1,04	11,99	0,94
22	0,976	0,976	0,13	-0,03	9,12	0,81	1,11	11,66	1,06
23	0,985	0,985	0,10	-0,01	8,12	0,60	0,90	10,22	0,76
24	0,977	0,977	0,16	0,02	14,57	0,74	1,33	18,03	0,91
25	0,992	0,992	0,13	0,01	8,96	0,64	0,98	11,18	0,82
Mean	0,962	0,961	0,04	-0,01	10,45	0,77	1,14	14,33	1,07
S.D.	0,023	0,023	0,57	0,05	2,63	0,21	0,24	4,47	0,36

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

Fig. 6.8 and Table 6.2 show the correlation scatter plots and the corresponding results of the HR and HRV parameters, respectively. It is noted an excellent correspondence in which very high correlation values were obtained in all HRV parameters. Most of the parameters achieved correlation coefficients above 0.99. The RMSSD and pNN50 components obtained the higher errors among the time-domain parameters. In general, the results were improved in comparison with the data obtained in stationary conditions in chapters 4 and 5.
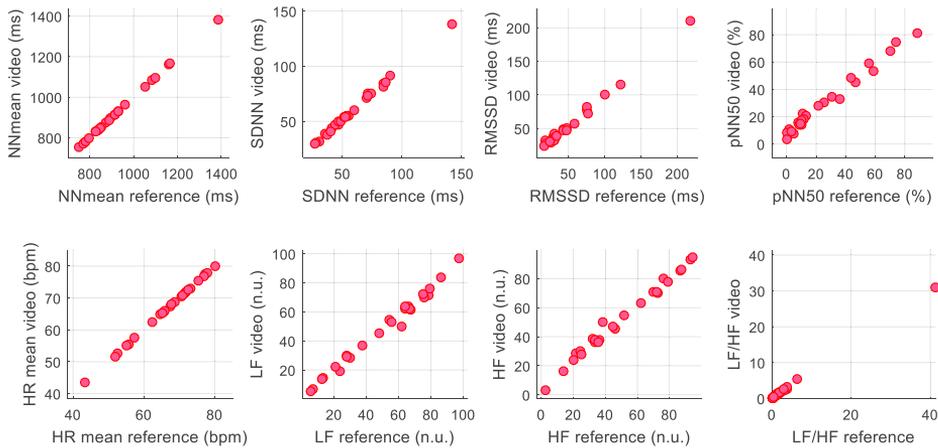
Fig. 6.8   Correlation scatter plots of the HR and HRV parameters (sunlight condition). Refer to Table 6.2 for r and ICC results.

Table 6.2   Statistical results of the HR and HRV parameters (sunlight condition)

Parameter	r*	ICC*	MAE	RMSE
HRmean (bpm)	1,000	1,000	0,04	0,05
NNmean (ms)	1,000	1,000	0,33	0,56
SDNN (ms)	0,998	0,996	1,80	2,17
RMSSD (ms)	0,997	0,989	5,41	6,22
pNN50 (%)	0,992	0,972	5,31	5,88
LF (n.u.)	0,994	0,990	2,77	3,82
HF (n.u.)	0,994	0,990	2,73	3,78
LF/HF	1,000	0,956	0,67	2,12

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

6.3.2 LED light condition

Table 6.3 presents the statistical results of the NN time series achieved by each subject in LED light condition. In comparison with the sunlight source, in this case, not all subjects obtained correlation coefficients above 0.9, although most of them did it. The $\bar{x} \pm s$ of the Pearson and ICC measures decreased to 0.922 ± 0.058 and 0.918 ± 0.063, respectively. The mean of the errors achieved a MAPE = 1.45 %, a MAE = 13.34 ms, and an RMSE = 19.11 ms.

Table 6.3   Statistical results of the NN time series by subject (LED light condition)

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,967	0,967	-0,44	0,00	15,31	0,45	1,07	20,93	0,63
2	0,970	0,968	0,14	-0,05	11,16	0,85	1,25	15,30	1,21
3	0,943	0,939	0,19	-0,04	9,24	0,83	1,13	12,14	1,09
4	0,973	0,972	0,10	-0,04	12,01	1,00	1,41	17,64	1,55
5	0,841	0,825	0,11	-0,10	15,72	1,71	2,11	23,19	2,54
6	0,896	0,896	-0,32	0,01	14,47	1,06	1,60	22,14	1,67
7	0,805	0,794	-0,20	-0,02	14,11	1,06	1,58	20,95	1,60
8	0,985	0,983	0,09	0,09	17,24	0,90	1,60	22,93	1,21
9	0,979	0,979	0,11	0,00	9,52	0,50	0,89	13,51	0,73
10	0,900	0,893	-0,40	-0,02	13,72	1,53	1,87	19,23	2,16
11	0,883	0,873	0,69	-0,13	14,60	1,25	1,74	23,38	2,09
12	0,969	0,969	0,43	-0,05	7,33	0,60	0,85	9,60	0,78
13	0,904	0,901	0,07	-0,06	15,61	1,47	1,95	23,75	2,26
14	0,833	0,821	0,21	-0,07	18,61	1,34	2,04	27,60	2,01
15	0,777	0,758	0,15	-0,05	15,52	1,15	1,72	21,61	1,63
16	0,915	0,912	0,08	-0,03	11,56	1,00	1,39	14,87	1,29
17	0,926	0,924	0,55	-0,05	17,11	0,87	1,57	23,71	1,22
18	0,959	0,958	0,06	-0,06	18,04	0,93	1,67	28,72	1,55
19	0,912	0,907	0,17	-0,04	10,83	0,82	1,21	14,56	1,12
20	0,919	0,917	-0,01	-0,04	12,32	1,30	1,63	17,14	1,83
21	0,966	0,966	0,03	-0,02	10,71	0,72	1,13	15,23	1,02
22	0,984	0,984	0,09	-0,01	6,45	0,57	0,78	8,33	0,75
23	0,980	0,980	0,12	-0,01	6,48	0,48	0,72	8,42	0,62
24	0,929	0,929	0,13	0,00	21,73	1,16	2,05	30,34	1,63
25	0,928	0,928	0,10	-0,02	14,18	0,85	1,41	22,42	1,32
Mean	0,922	0,918	0,09	-0,03	13,34	0,98	1,45	19,11	1,42
S.D.	0,058	0,063	0,25	0,04	3,87	0,34	0,41	6,09	0,53

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

Fig. 6.9 and Table 6.4 show the correlation scatter plots and the corresponding results of the HR and HRV parameters, respectively. It is noted that the correspondence of the data decreased in comparison with the sunlight condition, particularly in the pNN50 and frequency-domain components. Even so, high correlation values were obtained in the HRV parameters. The RMSSD and pNN50 components also obtained the higher errors among the time-domain parameters and almost twice as the results achieved in the sunlight condition.
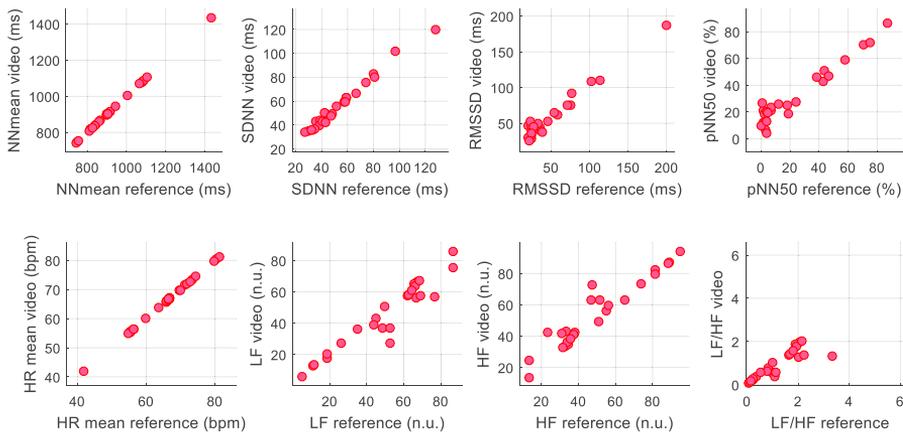
Fig. 6.9   Correlation scatter plots of the HR and HRV parameters (LED light condition). Refer to Table 6.4 for r and ICC results.

Table 6.4   Statistical results of the HR and HRV parameters (LED light condition)

Parameter	r*	ICC*	MAE	RMSE
HRmean (bpm)	1,000	1,000	0,04	0,05
NNmean (ms)	1,000	1,000	0,20	0,26
SDNN (ms)	0,993	0,981	3,68	4,41
RMSSD (ms)	0,981	0,944	10,78	13,07
pNN50 (%)	0,969	0,903	8,60	11,14
LF (n.u.)	0,952	0,930	5,52	8,62
HF (n.u.)	0,953	0,932	5,39	8,48
LF/HF	0,903	0,846	0,41	0,84

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

6.3.3 Fluorescent light condition

Table 6.5 presents the statistical results of the NN time series achieved by each subject in fluorescent light condition. Similarly to the LED light condition, most of the subjects obtained correlation coefficients above 0.9. The $\bar{x} \pm s$ of the Pearson and ICC measures were 0.951 ± 0.032 and 0.950 ± 0.033, respectively. These results are better in comparison with the LED light condition, but lower than the sunlight condition. The mean of the errors achieved a MAPE = 1.23 %, a MAE = 11.55 ms, and an RMSE = 16.43 ms.

Table 6.5   Statistical results of the NN time series by subject (fluorescent light condition)

Subj.	r*	ICC*	MBE (ms)	MBE (bpm)	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)
1	0,967	0,964	0,23	-0,03	15,64	0,47	1,11	21,39	0,66
2	0,968	0,967	1,04	-0,11	11,96	0,85	1,29	18,63	1,30
3	0,967	0,965	0,13	-0,04	8,31	0,82	1,06	10,64	1,06
4	0,988	0,988	0,18	-0,03	9,88	0,74	1,10	14,27	1,08
5	0,959	0,958	0,11	-0,04	10,09	1,00	1,29	12,87	1,28
6	0,983	0,982	0,07	-0,02	8,47	0,63	0,94	12,22	0,91
7	0,913	0,911	0,27	-0,04	10,95	0,85	1,25	14,90	1,15
8	0,983	0,982	0,00	0,06	14,73	0,76	1,36	22,66	1,21
9	0,951	0,950	0,06	-0,01	12,39	0,73	1,23	16,46	0,97
10	0,952	0,951	-0,27	-0,01	12,37	1,30	1,63	16,12	1,72
11	0,976	0,976	0,16	-0,03	9,40	0,69	1,04	12,50	0,94
12	0,967	0,965	0,21	-0,03	6,48	0,52	0,75	8,71	0,72
13	0,943	0,943	0,67	-0,06	10,04	0,83	1,17	14,66	1,20
14	0,865	0,863	0,42	-0,05	15,02	1,06	1,63	20,91	1,47
15	0,936	0,935	0,06	-0,01	9,21	0,64	0,99	11,46	0,81
16	0,869	0,861	0,17	-0,05	15,42	1,28	1,81	19,33	1,59
17	0,937	0,936	0,17	-0,02	14,89	0,71	1,32	20,42	0,98
18	0,921	0,921	-0,02	-0,01	19,62	0,88	1,70	33,57	1,52
19	0,952	0,950	0,17	-0,03	9,57	0,72	1,07	11,94	0,90
20	0,931	0,931	0,15	-0,03	11,58	1,15	1,48	20,12	1,88
21	0,982	0,982	0,02	-0,02	9,46	0,64	1,00	12,25	0,83
22	0,966	0,966	0,16	-0,01	9,92	0,82	1,16	16,57	1,34
23	0,955	0,955	0,12	-0,01	8,24	0,56	0,87	10,28	0,70
24	0,978	0,978	0,17	0,01	12,65	0,77	1,27	16,07	1,00
25	0,969	0,969	0,02	0,00	12,56	0,83	1,31	21,81	1,46
Mean	0,951	0,950	0,18	-0,02	11,55	0,81	1,23	16,43	1,15
S.D.	0,032	0,033	0,24	0,03	3,02	0,21	0,26	5,39	0,33

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

Fig. 6.10 and Table 6.6 show the correlation scatter plots and the corresponding results of the HR and HRV parameters, respectively. The correspondence of the parameters increased compared to the LED light source, but not as the results of the sunlight condition. High correlation results were obtained in the HRV parameters and most of them achieved values above 0.98. The RMSSD and pNN50 remained as the time-domain parameters with higher errors, whose results were closer to the sunlight condition.
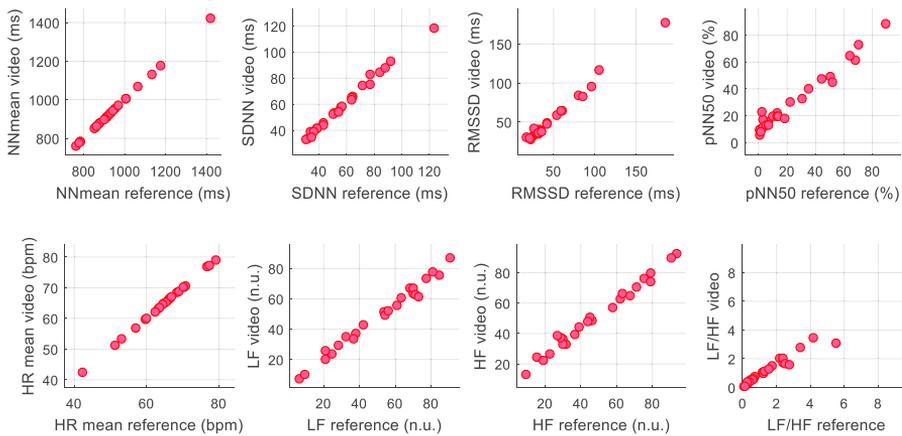
Fig. 6.10   Correlation scatter plots of the HR and HRV parameters (fluorescent light condition). Refer to Table 6.6 for r and ICC results.

Table 6.6   Statistical results of the HR and HRV parameters (fluorescent light condition)

Parameter	r*	ICC*	MAE	RMSE
HRmean (bpm)	1,000	1,000	0,03	0,04
NNmean (ms)	1,000	1,000	0,20	0,30
SDNN (ms)	0,997	0,994	2,10	2,55
RMSSD (ms)	0,994	0,981	6,25	7,13
pNN50 (%)	0,983	0,951	6,53	7,79
LF (n.u.)	0,993	0,979	3,65	4,67
HF (n.u.)	0,993	0,980	3,61	4,62
LF/HF	0,984	0,878	0,46	0,91

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.

Some authors also employed sources of artificial light to perform their experiments [27, 30]. In the case of Moreno et al. [27], the artificial light was used in order to control the effect of light on the video recordings. The lamp used in the experiments (Sylvania Satin H7/H 230V 40W) was chosen for its low amplitude variations in illumination, which can produce low-frequency components in the iPPG signal. The measurement set-ups of this work and our study present some similarities such as the same recording length and frame rate, the artificial illumination and recordings that were carried out in a sitting still position with closed eyes. By comparing the results achieved in both works, it is noted that our study achieved considerable better results with the two artificial light sources. Therefore, as the video resolution differed in both works, it could be a determinant factor in the obtained results, as discussed in previous chapters.

In the study of Kumar et al. [30], it was carried out an experiment in which the illumination level was varied from 50 lux up to 650 lux using fluorescent light (no other information of the light source was detailed), with increments of 50 lux and duration of 40 seconds in each light level. Two subjects with pale-white and brown skin tones were measured in this study. The experimental results show SNR (dB) variations at the different light levels, although they were mostly due to uncontrolled experimental conditions.

The method proposed by Kumar et al. (distancePPG) achieved on average an SNR improvement of 6.5 dB for the brown skin tone and 1.9 dB in the case of the pale-white tone, both in comparison with the results obtained by employing the face averaging method. No other results of this experiment were reported in the study. On the other hand, the results obtained in the ambient light condition show that our method achieved an improvement in the NN intervals measurement by obtaining an average RMSE value of 14.33 ms, compared with the 19.99 ms achieved in chapter 5 and the 15.74 ms obtained in the work of Kumar et al.

6.3.4   Bland-Altman plots

Fig. 6.11 shows the Bland-Altman plots representing the agreement between the NN intervals of the reference system and video in the three lighting conditions.
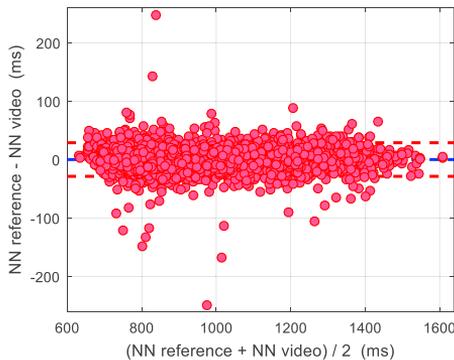
Sunlight condition results:
Mean difference = 0.04 ms (-0.02 bpm)
Lower LoA = -28.74 ms (-2.28 bpm)
Upper LoA = 28.82 ms (2.25 bpm)
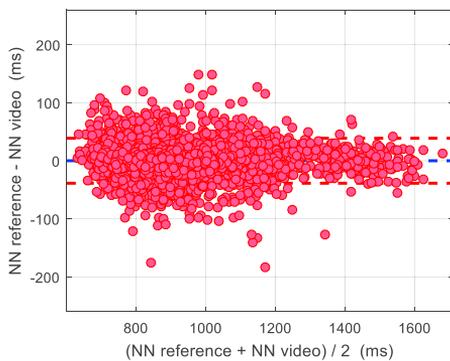
LED light condition results:
Mean difference = 0.09 ms (-0.03 bpm)
Lower LoA = -38.64 ms (-3.08 bpm)
Upper LoA = 38.82 ms (3.01 bpm)
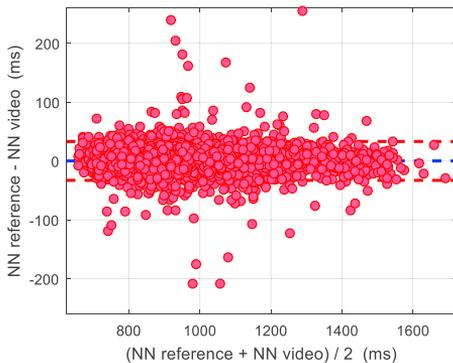
Fluorescent light condition results:
Mean difference = 0.18 ms (-0.03 bpm)
Lower LoA = -32.84 ms (-2.40 bpm)
Upper LoA = 33.19 ms (2.34 bpm)

Fig. 6.11   Bland-Altman plots with LoA of 95% representing the agreement between the reference and video in the three lighting conditions (data of all subjects combined in one dataset): (top) sunlight condition; (middle) LED light condition; (bottom) fluorescent light condition.

As the plots show, the sunlight condition achieved the minor differences between the measurements obtained by the reference system and video, while the LED light source obtained the larger differences. No systematic errors were identified in any of the three cases.

6.4   Conclusions

In this chapter was presented a video HRV analysis through the iPPG signal acquisition in real-time. The analysis performed in this chapter demonstrates a reliable signal acquisition in this condition. The real-time acquisition offers a quick video HRV, as well as the monitoring of the iPPG signal under external influences in real-time. The recordings of this study were carried out indoors with three different light sources in order to evaluate their influence on the measurements. In general, the three lighting conditions obtained good measurements but the sunlight achieved the best results.

The statistical results show that the NN time series and the HRV parameters obtained in the sunlight condition achieved an excellent agreement compared to the reference data. These results were improved in comparison with the data obtained in stationary conditions in chapter 4 and 5, in which the sunlight was also used to illuminate the face of the participants. A factor that could have influenced the results is a lesser amount of small movements of the head. Due to the recording length was longer, the subjects rested the head on the seat back in order to minimize the motion. Furthermore, the videos were recorded at a higher frame rate but with a lower video resolution in comparison with the recordings in chapter 5. By finding a balance between these two parameters could provide better results compared to other settings.

Most of the HRV parameters achieved correlation coefficients above 0.99 in the sunlight condition. These results are considerably better than those reported in most of the works of the state of the art. Conversely, the LED light source obtained the minor agreement in the measurements, but still with very good results. The sunlight source has an energy spectrum that is more evenly distributed in comparison with the LED and fluorescent lights, which present their energy more concentrated in some wavelengths. Thus, hypothetically, the energy level provided by the light source at the analyzed wavelength could be a factor that influences the quality of the signals.

Therefore, based on the results, the sunlight condition seems the best light option to perform a video HRV analysis. Moreover, this lighting condition is clearly more accessible anywhere than other sources, besides the cost involved. Additionally, the sunlight source avoids the contribution of high-frequency artifacts from some artificial sources to the iPPG signal. Even so, this study shows that some artificial lights can achieve very good and close results as those obtained with the sunlight source.

Chapter 7

Video HRV analysis in a real HCI scenario

Although HRV is normally measured at rest, recently, the interest of measuring physiological parameters in everyday activities by using alternative methods has increased. In this chapter is presented a video HRV analysis performed in a real HCI scenario. Most of these scenarios present a great amount and variety of motion and other environmental factors, which must be taken into account as they may significantly affect the measurements. In this respect, the progress of the video imaging technique may eventually provide an innovative, low-cost, contactless, and reliable alternative for measuring HR or HRV parameters in real-life scenarios.

7.1 Data acquisition set-up

A summary of the measurement set-up is shown below:

- 28 subjects (13 women, 15 men) between 22 and 56 years old were recorded with a Logitech webcam at 1920x1080 pixels and 15 fps (see section 3.1.2).
- The recordings were carried out in a sitting position with complete freedom of movement.
- The recordings included two tasks in front of a computer: a video viewing and the completion of a questionnaire.
- The recording length was variable according to the time required to complete the tasks (see Table 7.1).
- The distance between the camera and the face was 0.5 m approximately.
- The recordings were performed indoors with sunlight as the lighting source.

7.1.1 Human-computer interaction scenario

The aim of this study is to perform a video HRV analysis on a subject in a real HCI scenario. In order to assess the proposed method in this scenario, this chapter proposes the completion of two tasks in front of a computer, one after the other, while a video and the ECG signal of the subject are recorded. During the recordings, the subjects were allowed a complete freedom of movement.

The first task consisted of watching a video of the TED organization called "The happy secret to better work", which is one of the most popular TED talks. In this activity, the subjects were asked to watch the entire video and, once the video finished, the subjects were allowed to start the second task. The length of the video was 12 min 17 s.

In the second task, the subjects were asked to fill in a questionnaire that contains questions related to the psychological status, dietary habits, physical activity, and general data such as age, gender, weight, among others. In order to facilitate the filling out of the questionnaire, some inquiries presented a pull-down list with various answers to be selected. The complete questionnaire is included in the appendix (section A.2).

7.2   Video and signal processing

7.2.1   ROIs and feature points detection

The first step to perform the tracking of the subject is the ROIs detection. The automatic detection of the ROIs was carried out according to the section 5.2.1, in which the Viola and Jones algorithm was employed [20]. The conditions established in the aforesaid section were also applied to obtain the *ROIf* and *ROIe*. The ROIs detection is required in order to obtain the feature points, which are necessary to perform the tracking of the subject along the video sequence. The KLT algorithm [113-115] was the approach used to carry out this task, as detailed in section 5.2.2. Fig. 7.1 shows the ROIs and feature points detected in one participant of the study.
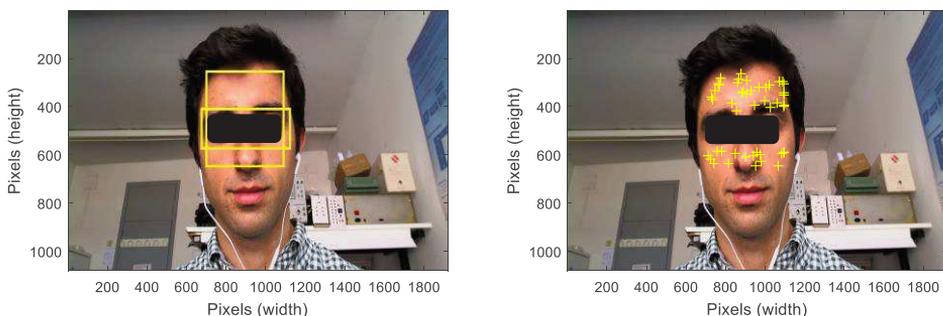

Fig. 7.1   (Left) ROIs detected in a subject during the HCI scenario; (right) feature points detected in the face to carry out the tracking of the subject along the video sequence.

7.2.2 Face tracking improvements

A real-life scenario is very likely to include a great amount and variety of motion. Clearly, the presence of this factor makes more difficult the tracking of the subject and, subsequently, the acquisition of a reliable iPPG signal. Therefore, some improvements are necessary to carry out a more robust face tracking and thus avoid issues such as a distortion or a loss of the ROIs.

In this chapter, some conditions were established with the aim of achieving a more robust face tracking. A great loss of feature points is likely to occur when a subject is tracked using the KLT algorithm in a real-life scenario. In such a case, a reacquisition of points is necessary in order to ensure a reliable tracking of the face. If the number of feature points falls below a certain amount, the risk of having an unreliable tracking increases. Then, in this study, if the number of feature points was below 100 at some time during the tracking, a reacquisition of points was carried out to improve the face tracking.

During the tracking, if the condition of the minimum number of feature points was met, then, the tracking of these points was carried out in the current frame. If not, a reacquisition of feature points was performed, in which the ROIs detection was first required in order to obtain the points. The ROIs detection and the reacquisition of the feature points were performed in the same way as described in section 7.2.1. If both ROIs were not detected, the current ROIs and remaining feature points were kept until the two ROIs were found in following frames. The *ROIf* coordinates obtained in each frame were stored to be used in a posterior analysis.

However, despite a great amount of motion during the recordings, a high number of points could remain along the video sequence. In such a case, a possible consequence is the distortion of the ROIs, which may influence the acquisition of the iPPG signal. Therefore, an automatic ROIs detection was only

carried out in case there were large movements. For this purpose, the face displacements were measured every second by calculating the pixel differences of the top-left coordinate of the *ROIf* between the current and the corresponding previous frame. The presence of a large movement was considered when there was a difference superior to 100 pixels in any of the two axes (x or y) of motion. Likewise, if both ROIs were not detected in this step, the current ones were kept until the two ROIs were found in following frames.

7.2.3   Video image processing and signal acquisition

The video image processing and signal acquisition were carried out as described in section 5.2.3, in which only the pixels that correspond to the skin were analyzed and those that may contribute with artifacts were excluded. The signal filtering and interpolation were carried out according to the section 4.2.3. In the case of the presence of a Not-a-Number (NaN) in the vector containing the video signal, these values were replaced with zeros. The NaN values could result from frames in which were not possible to calculate the signal values because of the ROIs were not detected.

7.2.4   Quality distinction of the video signal

In this chapter, the subjects were allowed to perform natural movements during the recordings. The amount and type of motion carried out by the participants varied from person to person, which in some cases considerably affected the video signals. For this reason, the presence of a great amount of motion artifacts incurred a more difficult acquisition of the measurements. Therefore, a quality distinction of the video signal is proposed in this chapter in order to only analyze the segments that contain a minor amount of motion artifacts.

The *ROIf* coordinates obtained along the face tracking provided relevant information of the motion performed by the subjects. In order to identify the segments with a minor amount of motion artifacts, every 5 frames it was checked if the values of the top-left coordinate of the *ROIf*, both on the x-axis or y-axis, presented a difference equal to or less than 3 pixels (Fig. 7.2). If the differences met this condition, the corresponding indexes were identified with a value of 1; otherwise, the assigned value was 0 (Fig. 7.3 (a, b)). These *x* and *y* values were multiplied to analyze both axes of motion in one variable (Fig. 7.3 (c)).
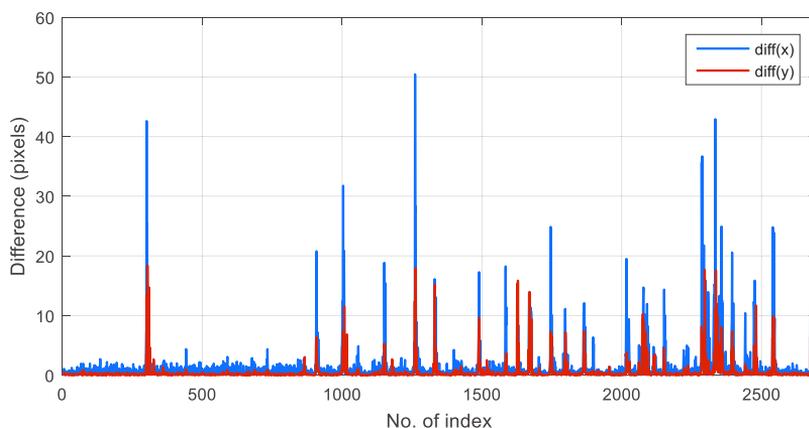
Fig. 7.2   Differences measured every 5 frames from the top-left coordinate values of the *ROIf* of a subject. If the differences presented a difference equal to or less than 3 pixels, the corresponding indexes were identified with a value of 1; otherwise, the assigned value was 0 (see Fig. 7.3 (a, b)).

This last variable of motion information represents a preliminary sequence of segments. The final selection was obtained according to the segments that met a corresponding minimum length of 5 s (Fig. 7.3 (d)). Half a second at each end of the segments was excluded to avoid the presence of remaining artifacts. Thus, the video signal is analyzed where the segments present values equal to 1 and, conversely, it is not assessed where the corresponding values are 0.
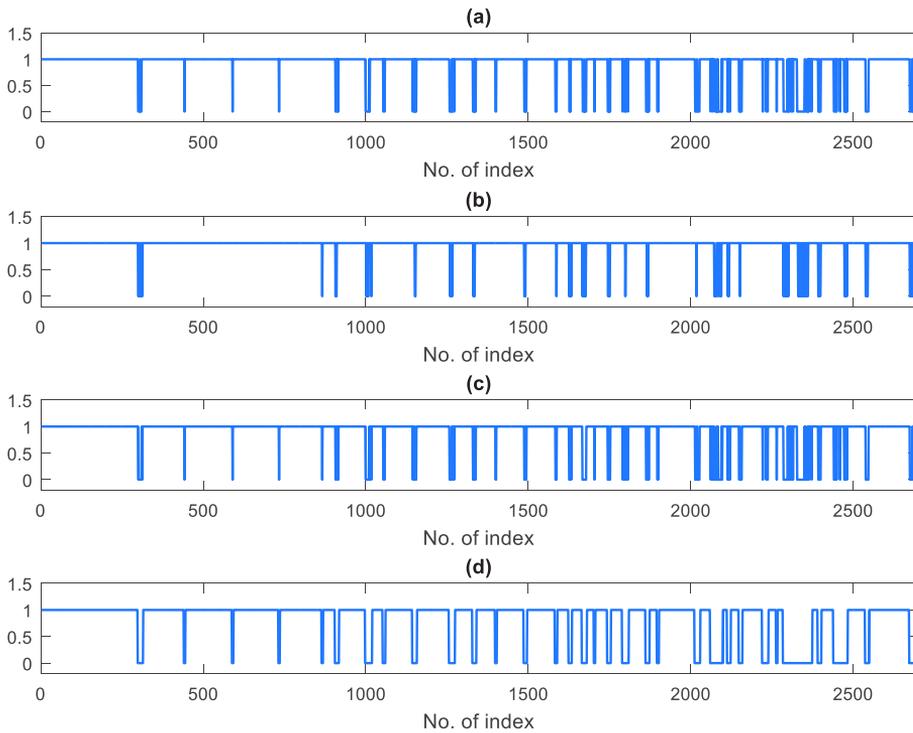
Fig. 7.3 (a, b) representation of the values obtained in a subject according to the motion condition on the x-axis and y-axis, respectively (see Fig. 7.2); (c) result of the multiplication of (a) and (b); (d) final sequence of segments (the signal is analyzed where the segments present values equal to 1 and, conversely, it is not assessed where the corresponding values are 0).

Fig. 7.4 compares the final selection of the segments with the corresponding amount of motion performed by a subject during the recording. It is shown that the segments with a large amount of motion are not considered in the analysis. Fig. 7.5 shows the illustration of a video signal and the corresponding sequence of segments to be analyzed, as well as a zoomed section of them. The sequence of segments was adjusted according to the length of the video signal. The zoomed section shows more clearly the motion artifacts affecting the iPPG signal. The local maxima were only detected in the segments that presented a minor amount of motion. The Pan-Tompkins algorithm was employed to detect the peak of the QRS complexes of the reference signal [122].
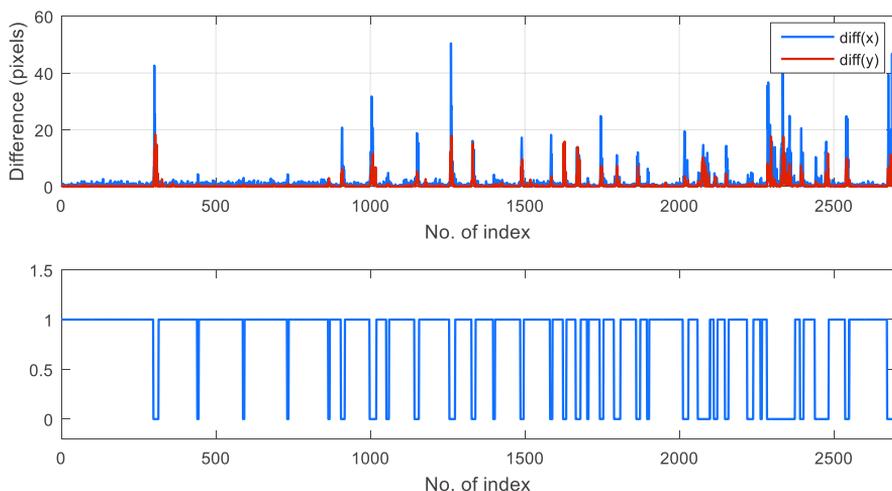
Fig. 7.4 (Top) differences measured every 5 frames from the top-left coordinate values of the *ROIf* of a subject; (bottom) final sequence of segments (the signal is analyzed where the segments present values equal to 1 and, conversely, it is not assessed where the corresponding values are 0).
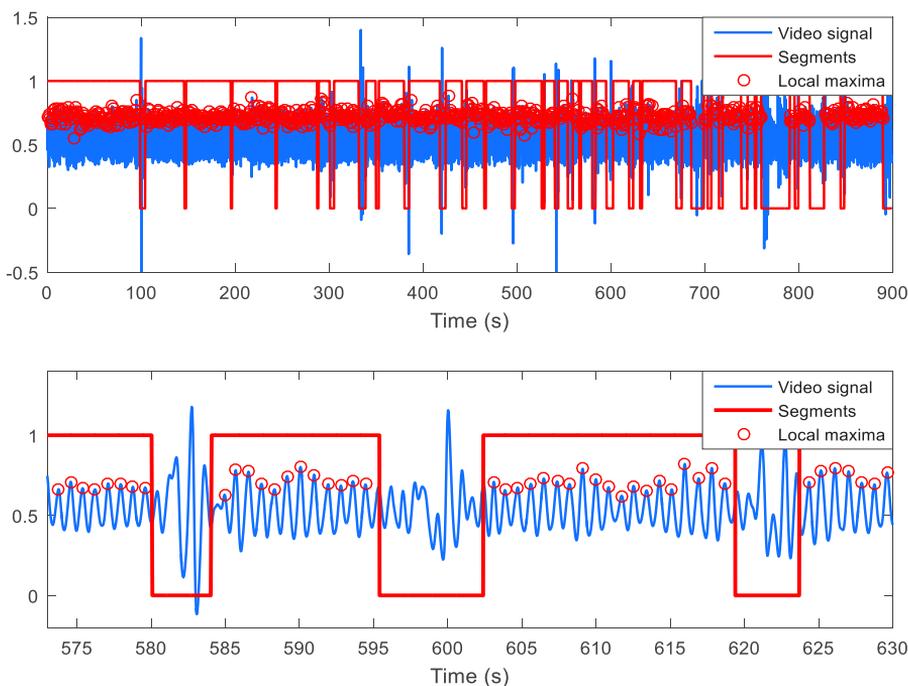
Fig. 7.5 (Top) video signal and the sequence of segments; (bottom) zoomed section of the video signal and the sequence of segments showing the motion artifacts affecting the signal (the signals were analyzed where the segments presented values equal to 1).

7.2.5 Artifacts correction

The NN intervals of the video and reference were obtained by calculating the time between consecutive local maxima and the subsequent artifacts correction was carried out according to the section 4.2.5. The NN-NN intervals that were above the median + 6 standard deviations of the NN-NN series were also identified as outliers, and the corresponding NN intervals were replaced with the average of the five previous NN intervals (Fig. 7.6). This correction step was repeated until no more outliers were found.
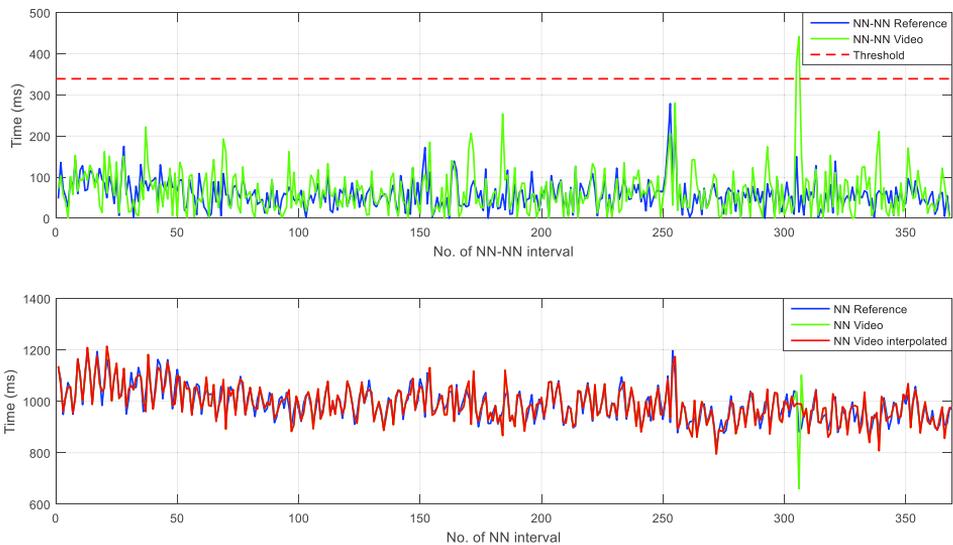
Fig. 7.6 (Top) absolute values of the differences between adjacent NN measurements (NN-NN) of a subject; (bottom) NN intervals of reference and video, which correspond to the NN-NN intervals shown on the top of the figure, before and after the correction of artifacts.

In this case, only the NNmean and SDNN parameters were calculated as part of the HRV analysis (refer to section 2.2.1). The other time-domain parameters and the frequency-domain components were not obtained due to the discontinuity of the NN intervals sequence.

7.3   Results and discussion

Table 7.1 presents the statistical results of the NN time series obtained by each subject. A high variety of results were obtained by the participants, in which most of them achieved correlation coefficients above 0.8. The $\bar{x} \pm s$ of the Pearson and ICC measures were 0.830 ± 0.088 and 0.819 ± 0.098, respectively.

Table 7.1   Statistical results of the NN time series obtained by subject in the HCI scenario

Subj.	r*	ICC*	MAE (ms)	MAE (bpm)	MAPE (%)	RMSE (ms)	RMSE (bpm)	R. length (min:s)	R. length analyzed (%)
1	0,831	0,816	17,22	1,32	1,94	23,13	1,83	15:47	57
2	0,928	0,922	15,36	1,12	1,69	19,72	1,44	14:35	22
3	0,938	0,935	17,31	1,34	1,96	22,62	1,79	16:16	20
4	0,815	0,806	18,58	1,47	2,13	25,19	2,02	15:05	24
5	0,707	0,700	29,65	3,22	3,99	44,55	4,64	14:54	23
6	0,912	0,911	21,30	1,32	2,16	28,35	1,77	15:01	41
7	0,865	0,855	20,05	1,62	2,31	26,41	2,19	15:32	29
8	0,947	0,946	14,26	0,76	1,34	19,12	1,02	15:43	14
9	0,740	0,716	20,56	2,68	3,01	27,43	3,62	15:09	34
10	0,740	0,718	27,89	1,63	2,75	35,05	2,05	15:24	69
11	0,901	0,899	19,36	1,82	2,42	25,32	2,46	16:05	56
12	0,774	0,763	29,96	1,58	2,82	39,22	2,10	15:10	53
13	0,800	0,781	15,62	1,48	1,96	21,02	2,03	16:59	16
14	0,649	0,620	41,04	2,75	4,32	54,46	3,70	17:12	37
15	0,725	0,712	18,74	2,16	2,58	25,91	3,03	15:22	23
16	0,859	0,855	21,48	2,19	2,81	28,07	2,86	15:22	19
17	0,852	0,847	11,92	1,03	1,43	15,55	1,35	15:44	57
18	0,763	0,745	27,41	2,31	3,25	38,83	3,31	15:32	35
19	0,892	0,887	14,47	1,10	1,63	19,05	1,45	15:48	62
20	0,913	0,911	14,67	1,37	1,83	19,06	1,78	17:02	38
21	0,856	0,847	15,72	1,11	1,70	19,29	1,37	17:04	15
22	0,656	0,597	31,69	4,01	4,59	41,94	5,39	15:59	40
23	0,807	0,789	24,16	1,92	2,78	31,95	2,59	14:59	77
24	0,961	0,962	20,68	1,32	2,12	28,33	1,97	14:56	32
25	0,811	0,810	27,88	2,03	3,08	38,88	2,95	15:30	21
26	0,960	0,959	20,23	0,94	1,77	25,74	1,20	15:06	53
27	0,836	0,825	16,78	1,45	2,01	22,53	1,98	15:47	49
28	0,802	0,793	21,99	1,80	2,56	28,81	2,37	15:53	64
Mean	0,830	0,819	21,28	1,74	2,46	28,41	2,37	15:40	39
S.D.	0,088	0,098	6,60	0,73	0,83	9,21	1,03	0:42	18

*: All p < 0.001. R. length: record length. Refer to section 3.2 for statistical parameters definitions.

The variety of results obtained by the participants in this study is wide, in which some subjects achieved high correlation results and, on the other hand, the results obtained by other participants were not as good as expected. Clearly, due to the conditions of the study, a relevant factor to take into account in these measurements is the motion. The subjects were allowed a total freedom of movement during the recordings, which influenced the analyzed record length, as shown in Table 7.1.

On one side, some subjects with a higher percentage of analyzed record length, for instance, the participants 11 and 26 achieved good correlation results and others obtained a lower agreement in the measurements, as the participants 10 and 12, for example. On the other side, some subjects with a lower percentage of analyzed record also obtained both good and lower agreement results, as noted in the participants 3, 8, and 5, 15, respectively. Therefore, in this study, the analyzed record length was not related to the agreement achieved in the measurements.

Unlike the proposed HCI scenario, the motion conditions of chapter 5 were certainly more specific in which the subjects performed lateral and forward/backward movements, always looking toward the camera, and avoiding fast gestures. In contrast to these conditions, some recordings of this chapter presented a wider variety of motion that included faster movements, different positions of the face, more changes in shape and size of the ROIs, as well as a less uniform illumination over the face, which could affect the results.

Moreover, based on the quality distinction of the signal, the segments with a minor amount of motion were identified with the aim of excluding the segments with more motion artifacts. However, some subjects performed particular facial expressions or laughed during the recordings which could also have influenced the measurements. Based on the above, the difference in results may be

influenced by the aspects discussed here, but also, due to other factors such as the particular anatomical and physiological characteristics of the subjects.

Fig. 7.7 and Table 7.2 show the correlation scatter plots and the corresponding results of the NNmean and SDNN parameters, respectively. Due to the discontinuity of the NN intervals sequence, other time-domain parameters and the frequency-domain components were not calculated in this chapter.
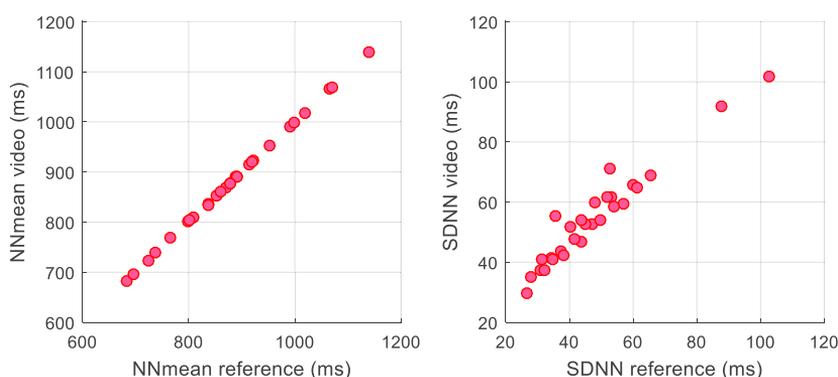
Fig. 7.7   Correlation scatter plots of the NNmean and SDNN parameters. Refer to Table 7.2 for r and ICC results.

Table 7.2   Statistical results of the NNmean and SDNN parameters

Parameter	r*	ICC*	MAE	RMSE
NNmean (ms)	1,000	1,000	0,59	0,89
SDNN (ms)	0,964	0,886	7,04	8,28

*: All p < 0.001. Refer to section 3.2 for statistical parameters definitions.

The SDNN parameter achieved a good Pearson's correlation coefficient but the ICC measure was more affected in comparison with the results reported in previous chapters, in which the component obtained ICC values above 0.9. Only the measurements obtained with the iPhone 4s camera in stationary conditions in

chapter 4 achieved a lower agreement. The agreement achieved in this parameter is also reflected in the increase of the reported error results.

Some authors also obtained measurements by video imaging that included a HCI scenario or the presence of natural movements [19, 22, 30, 31]. Firstly, Poh et al. [19] conducted a study in which the subjects were asked to move naturally by simulating the interaction with a laptop (no large or rapid movements were allowed). In the study of Monkaresi et al. [22], the measurements were obtained in a natural HCI scenario in which the participants were allowed to explore the internet and other resources in order to complete a requested task. In both works, it is important to note that the measurements were acquired by a frequency-domain signal analysis, which is not comparable with the signal analysis of our study.

The work of Kumar et al. [30] included recordings carried out in three different motion conditions: reading, watching video and talking. From these scenarios, the watching video situation presents a greater similarity with the HCI scenario of our study. According to the description in the study of Kumar et al., this scenario involved intermittent facial expressions such as smiling, getting amazed, sad, etc., and lateral movement of the head.

Unlike the works of Poh et al. and Monkaresi et al., Kumar et al. performed a time-domain signal analysis in which the PP intervals were measured. Due to the presence of large motion, some peaks were not detected in the iPPG signal (5.68 % in the watching video scenario). These PP measurements achieved an average RMSE of 67.08 ms, which is more than twice the error achieved in the measure of the NN time series in our study. Taking into account the more similar results achieved in stationary conditions in both works, some particular movements performed by the participants could greatly affect the statistical results in the study of Kumar et al.

On the other hand, Antink et al. [31] performed a beat-to-beat heart rate estimation in a reading scenario without motion and without further instructions. This work involved the acquisition of different signals and their fusion, which include the video signals obtained from the changes of the skin color. In the case of the reading scenario without further instructions, although the average MAE result (see Table 2.3) is higher than the corresponding error achieved in the HCI scenario of our study, it is closer in comparison with those obtained in other reference works. The results are improved when considering the data achieved from the fusion of all the different signals, although the error continues to be higher. It is important to note that the works of Antink et al. and Kumar et al. present a smaller sample size in comparison with other reference works, which makes it more difficult to obtain a reliable statistical analysis.

Fig. 7.8 shows the Bland-Altman plot with LoA of 95% representing the agreement between the NN intervals of reference and video. The results presented a mean difference of -0.11 ms (-0.06 bpm) and LoA from -58.82 to 58.61 ms (-5.16 to 5.04 bpm).
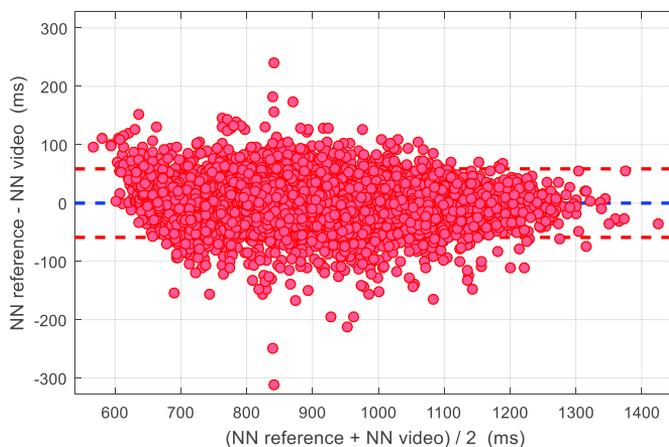

Fig. 7.8   Bland-Altman plot with LoA of 95% representing the agreement between the NN intervals (data of all subjects combined in one dataset). The results presented a mean difference of -0.11 ms (-0.06 bpm) and LoA from -58.82 to 58.61 ms (-5.16 to 5.04 bpm).

No systematic errors were identified in the measurements. The differences between the measurements obtained by both methods were increased compared with the results achieved in chapter 5 in which the subjects performed specific movements. Some sort of movements performed by the subjects, facial expressions, as well as the particular results of some participants not influenced by the motion could affect these overall data.

7.4   Conclusions

In this chapter is proposed a method to perform a video HRV analysis on a person in a real HCI scenario. During the proposed tasks, the amount and type of motion carried out by the participants varied from person to person, which in some cases greatly affected the video signals. Therefore, the method employed in this study is a selective algorithm which aims to analyze the segments of the video signal that contain a minor amount of motion artifacts.

Due to the wider variety of motion, some improvements are proposed in order to achieve a more robust face tracking of the participants. An automatic reacquisition of feature points was established in case of a great loss of points during the motion scenarios. Also, a distortion of the ROIs is very likely to occur in presence of a great amount of motion, which subsequently may affect the acquisition of the signal. Therefore, an automatic ROIs detection was carried out every second in the case that the subjects performed certain large movements.

The statistical analysis of the NN time series obtained by each participant shows a high variety of results, both good and bad, however, most of the subjects achieved correlation coefficients above 0.8. The difference in results between subjects may be influenced by several aspects related to the motion conditions of the HCI scenario but, also, due to other factors such as the

particular anatomical and physiological characteristics of the subjects. In this study, the analyzed record length was not related to the agreement achieved in the measurements. Based on the general results, the proposed method achieved lower errors in comparison with those obtained in related works, which obtained measurements in a HCI scenario or in the presence of natural movements.

The NNmean and SDNN parameters obtained Pearson's correlation coefficients above 0.9, of which, the SDNN parameter was particularly affected in the measure of ICC. The Bland-Altman plot shows larger differences between the NN intervals of reference and video, compared with the results achieved in chapter 5 in which the subjects performed specific movements.

The proposed quality distinction of the video signal performed very well but, in a few cases, the distinction was not completely accurate because it is based on the amount of motion. On one side, some segments of the signal with good quality were rejected due to the presence of a great amount of motion. Conversely, some segments were analyzed because of the subject did not perform large movements of the head, despite some facial expressions or the laugh affected the measurements. In this respect, an objective for future work could be the implementation of an automatic identification of facial expressions in order to exclude the artifacts produced by them.

Based on the above, more improvements are needed in order to carry out a better analysis in a real-life scenario. The development of more robust algorithms is necessary to suppress, as much as possible, the contribution of artifacts caused by the factors present in a real environment, as well as a better signal quality distinction with the aim of performing a more reliable video HRV analysis in these scenarios.

Chapter 8

Respiratory rate analysis by video imaging

The respiratory signal analysis by video, just as the heart rate and HRV analysis, also has been explored in the last few years (refer to section 2.4). In this chapter is proposed the application of a tracking algorithm in order to obtain the respiratory signal by video. The tracking of the body movements caused by the respiration, which are reflected mostly in the torso area, may have a correspondence to the respiratory signal of the subject. Thus, the respiratory rate analysis by video could provide a low-cost, contactless, and reliable alternative to traditional methods, just as the video HRV analysis.

8.1 Data acquisition set-up

A summary of the measurements set-up is shown below:

- 25 subjects (13 women, 12 men) between 22 and 54 years old were recorded with a GoPro camera at 1920x1080 pixels and 30 fps (see section 3.1.2).
- The reference signal was recorded using the RespiBand by plux of FICOSA International S.A. at a sampling frequency of 40 Hz (see section 3.1.1).
- The recordings were obtained at rest by using a seat belt of a car seat.
- The recording length was 5 min.
- The distance between the camera and the torso was 0.5 m approximately.
- The recordings were performed indoors with sunlight as the lighting source.

8.2 Video and signal processing

8.2.1 Selection and tracking of feature points

The respiratory signal by video was obtained using the KLT algorithm [113-115], which has also been employed previously to track the face of the subjects. As mentioned earlier, this algorithm allows the tracking of feature points along a video sequence. In this study, a seat belt of a car seat was used as a reference, which was marked with small white spots in order to obtain the feature points. The tracking of a different number (n) of reference marks was carried out in each subject because of the different physical constitution of the participants. Fig. 8.1 shows an example of two people with a different number of reference seat belt marks.

Fig. 8.1   Subjects with a different number (*n*) of reference seat belt marks (inside the green circles): (left) subject with six marks; (right) subject with ten marks. The reference marks were numbered from left to right and from top to bottom.

The acquisition of feature points was carried out through the manual selection of *n* seat belt marks on the first video frame. Then, eight extra feature points were obtained around the locations (x_i, y_i) of the selected points in order to perform a more robust tracking as follows

$$\begin{pmatrix} x_i - 1, y_i - 1 & x_i, y_i - 1 & x_i + 1, y_i - 1 \\ x_i - 1, y_i & x_i, y_i & x_i + 1, y_i \\ x_i - 1, y_i + 1 & x_i, y_i + 1 & x_i + 1, y_i + 1 \end{pmatrix} = \begin{pmatrix} x_{i_1}, y_{i_1} & x_{i_2}, y_{i_2} & x_{i_3}, y_{i_3} \\ x_{i_4}, y_{i_4} & x_{i_5}, y_{i_5} & x_{i_6}, y_{i_6} \\ x_{i_7}, y_{i_7} & x_{i_8}, y_{i_8} & x_{i_9}, y_{i_9} \end{pmatrix}$$

in which $i = 1,2,3,...n$. Fig. 8.2 shows the feature points obtained on the seat belt marks in a subject.
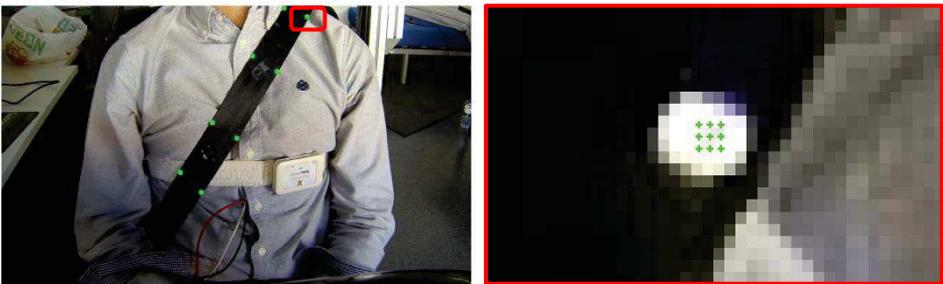

Fig. 8.2   (Left) features points (green marks) obtained on the seat belt marks (red rectangle encloses the image on the right); (right) set of features points on a seat belt mark.

The presence of additional features points allows a more reliable signal acquisition, given the possibility of loss of points along the video sequence. A point tracker was established which enables the bidirectional error of the feature points due to the presence of noise. The maximum bidirectional error was set at 3 pixels. Once the tracking of points was accomplished along the video sequence, the locations of the feature points of each reference mark were averaged to obtain a central position by frame.

$$x_{i_{avg}} = \frac{\sum_{k=1}^{9} x_{i_k}}{9} \quad , \quad y_{i_{avg}} = \frac{\sum_{k=1}^{9} y_{i_k}}{9} \qquad \text{(Eq. 8.1)}$$

8.2.2 Respiratory signal acquisition

With the aim of measuring the body displacements caused by the respiration, the distances between feature points were calculated as shown in Fig. 8.3. Thus, a longer distance between points corresponds to the inspiratory phase of the breath, and conversely, a shorter distance to the expiratory phase. The calculation was carried out in all frames to obtain the raw respiratory signals.

Fig. 8.3   Distances calculated between features points (dashed green arrows) to obtain the respiratory signals. The number of signals depended on the number of available feature points.

Given the averaged locations of the feature points on each frame, the distance between two points was calculated by using the following formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad \text{(Eq. 8.2)}$$

The number of respiratory signals depended on the number of available feature points of each subject. Fig. 8.4 shows the respiratory signals obtained from a subject by calculating the corresponding distances between the feature points.
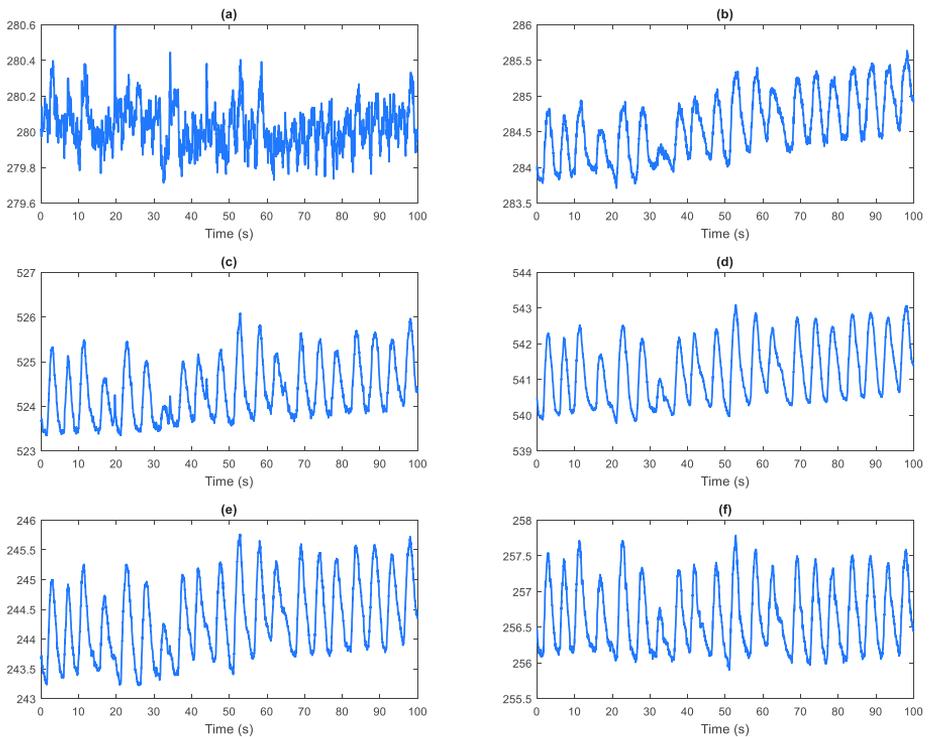
Fig. 8.4 Respiratory signals obtained from a subject by calculating the distances between the feature points (first 100 s). The signals (*s*) are named with subscripts according to the two reference marks numbers used for their calculation (see Fig. 8.1): (a) s_{13}; (b) s_{24}; (c) s_{15}; (d) s_{26}; (e) s_{35}; (f) s_{46}.

A second-order bandpass Butterworth filter between 0.05 and 0.5 Hz was applied to remove low and high frequency noise components. The cutoff frequencies correspond to the range from 3 to 30 breaths per minute (BPM). A cubic spline interpolation was performed to improve the temporal resolution of the video (30 fps) to the sampling frequency of the reference signal (40 Hz). The reference and video signals were normalized. Fig. 8.5 shows the respiratory signals of a subject after applying the filter and the interpolation.
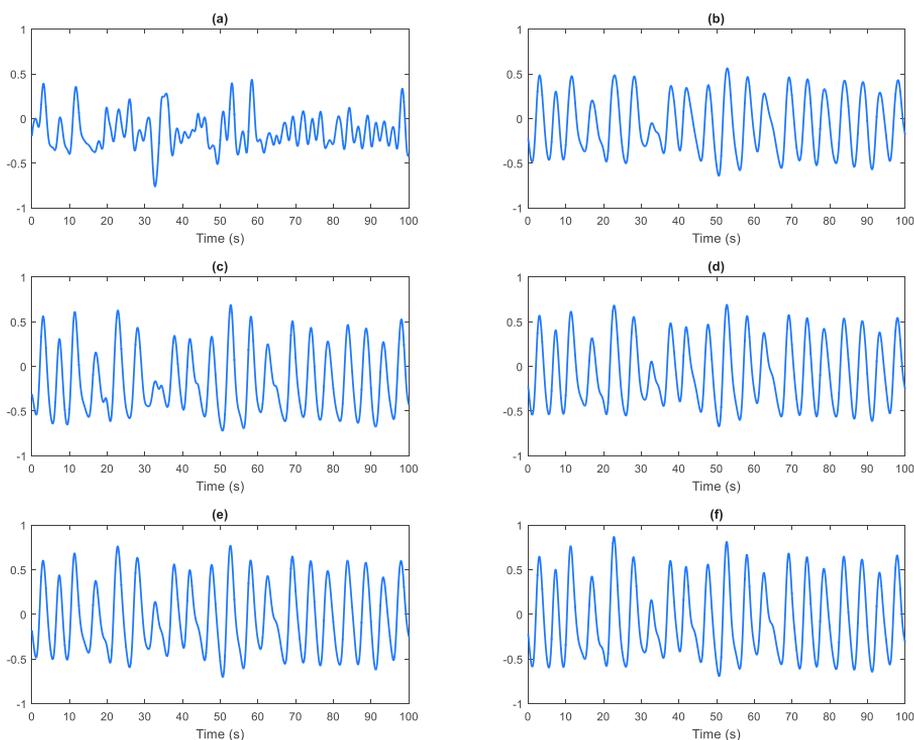
Fig. 8.5 Respiratory signals obtained from a subject after applying the filter and the interpolation (first 100 s): The signals (s) are named with subscripts according to the two reference marks numbers used for their calculation (see Fig. 8.1): (a) s_{13}; (b) s_{24}; (c) s_{15}; (d) s_{26}; (e) s_{35}; (f) s_{46}.

8.2.3   Synchronization of the signals

The synchronization of the respiratory signals was performed because of the camera and the inductive chest band were not aligned in time. This step was carried out by a cross-correlation analysis (Eq. 4.1) between the reference and video signals (Fig. 8.6).
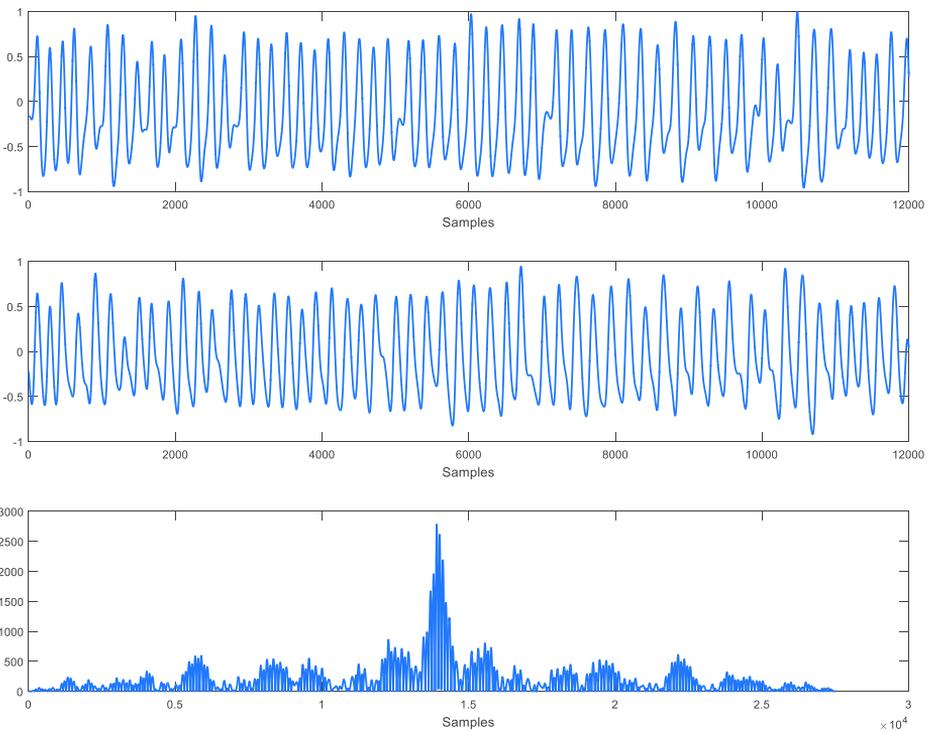
Fig. 8.6   Cross-correlation analysis performed between the respiratory signals: (top) reference signal; (middle) video signal; (bottom) result of the cross-correlation in absolute values.

The maximum absolute value of each cross-correlation and its corresponding time offset were calculated in the analysis. Thus, the synchronization of all signals was carried out by adjusting the time offset of the signal that achieved the highest correlation value.

8.2.4 Breathing rate time series

In order to calculate the BR time series, the local maxima were detected on the respiratory signals. The maximum possible value of a respiratory cycle was defined as 30 BPM, which corresponds to 2 seconds. Taking into account this minimum time between peaks of the respiratory signal, the local maxima were detected to obtain the BR intervals (Fig. 8.7).
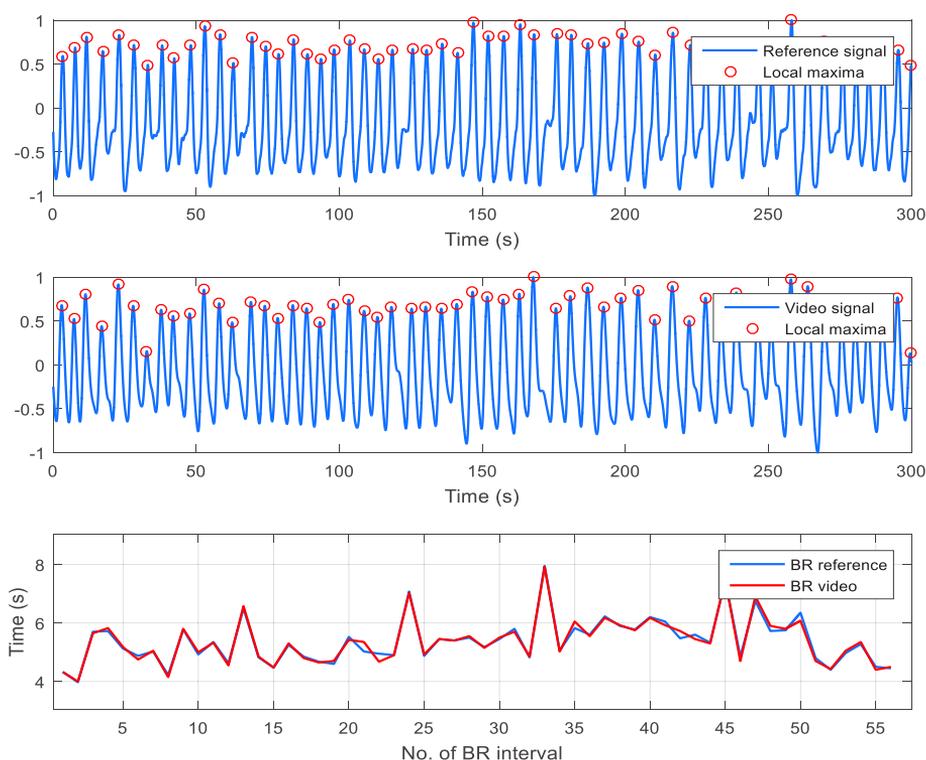
Fig. 8.7   Local maxima detection on the respiratory signals: (top) reference signal; (middle) video signal; (bottom) breathing rate (BR) time series of reference and video.

After the synchronization of the signals, which in a few cases resulted in non-synchronized intervals of reference and video, a synchronization of the BR intervals was carried out by means of a cross-correlation analysis (Eq. 4.1).

8.2.5   Selection of the respiratory signal

Due to the existence of several respiratory signals of video, only one of them was chosen in order to be compared with the reference data. To this end, the BR time series of video that achieved the minimum MAPE (Eq. 3.5) were chosen to perform the posterior statistical analysis. Moreover, the BR intervals of video that obtained the three minimum and three maximum MAPE were identified with the aim of finding the seat belt marks that achieved the best and worst results.

8.3   Results and discussion

Several parameters were calculated to evaluate the BR time series that achieved the minimum MAPE. Table 8.1 presents the statistical results of the BR time series obtained by each subject. As shown, all subjects achieved excellent results with correlation coefficients above 0.9. The $\bar{x} \pm s$ of the Pearson and ICC measures were 0.968 ± 0.029 and 0.967 ± 0.030, respectively. The mean of the errors achieved a MAPE = 1.94 % value, which corresponds to a MAE = 0.09 s or an RMSE = 0.15 s.

Fig. 8.8 and Table 8.2 show the correlation scatter plots and the corresponding statistical results of the mean and standard deviation of the breathing rate intervals (BRmean and SDBR), respectively. It is noted an excellent correspondence in these parameters which achieved very high correlation coefficients with values above 0.99. Fig. 8.9 shows the Bland-Altman plot with LoA of 95% representing the agreement between the BR intervals of reference and video. The mean difference was -0.004 s (0.01 breaths/min (BPM)), and the LoA from -0.33 to 0.32 s (-0.88 to 0.90 BPM), therefore, no large differences were found between both methods.

Table 8.1   Statistical results of the BR time series by subject

Subj.	r*	ICC*	MBE (s)	MBE (BPM)	MAE (s)	MAE (BPM)	MAPE (%)	RMSE (s)	RMSE (BPM)
1	0,988	0,988	0,00	-0,01	0,08	0,25	1,82	0,11	0,36
2	0,990	0,990	-0,02	0,04	0,16	0,19	2,17	0,24	0,29
3	0,991	0,991	0,00	0,00	0,08	0,16	1,43	0,11	0,23
4	0,995	0,995	0,00	0,00	0,03	0,14	0,80	0,04	0,18
5	0,988	0,988	0,00	0,02	0,04	0,22	1,27	0,06	0,31
6	0,992	0,992	0,00	0,00	0,03	0,18	0,93	0,04	0,25
7	0,904	0,901	-0,04	0,10	0,11	0,34	2,71	0,39	0,83
8	0,971	0,971	0,00	0,01	0,06	0,22	1,48	0,07	0,29
9	0,922	0,923	0,00	0,00	0,06	0,28	1,73	0,12	0,54
10	0,980	0,979	0,00	0,02	0,04	0,32	1,54	0,06	0,43
11	0,940	0,935	-0,08	0,06	0,47	0,24	4,27	0,59	0,31
12	0,918	0,918	0,00	0,03	0,14	0,39	2,95	0,21	0,58
13	0,914	0,911	0,00	0,02	0,05	0,35	1,73	0,10	0,60
14	0,984	0,985	0,00	0,01	0,12	0,25	2,18	0,16	0,33
15	0,954	0,951	0,00	0,01	0,10	0,56	3,07	0,18	0,75
16	0,991	0,991	0,00	0,00	0,03	0,13	0,86	0,04	0,17
17	0,993	0,993	0,00	0,00	0,04	0,14	0,96	0,06	0,19
18	0,987	0,987	0,00	-0,02	0,19	0,39	3,42	0,40	0,69
19	0,996	0,996	0,00	0,00	0,03	0,12	0,79	0,04	0,17
20	0,960	0,960	0,00	0,01	0,06	0,33	1,77	0,07	0,43
21	0,970	0,970	0,00	0,01	0,10	0,30	2,17	0,15	0,42
22	0,930	0,928	0,00	0,02	0,11	0,38	2,59	0,14	0,49
23	0,966	0,966	0,00	-0,01	0,10	0,43	2,59	0,13	0,60
24	0,984	0,983	0,00	-0,02	0,09	0,33	2,13	0,11	0,40
25	0,995	0,995	0,00	0,01	0,04	0,15	1,03	0,07	0,27
Mean	**0,968**	**0,967**	**-0,01**	**0,01**	**0,09**	**0,27**	**1,94**	**0,15**	**0,40**
S.D.	**0,029**	**0,030**	**0,02**	**0,03**	**0,09**	**0,11**	**0,90**	**0,13**	**0,19**

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.
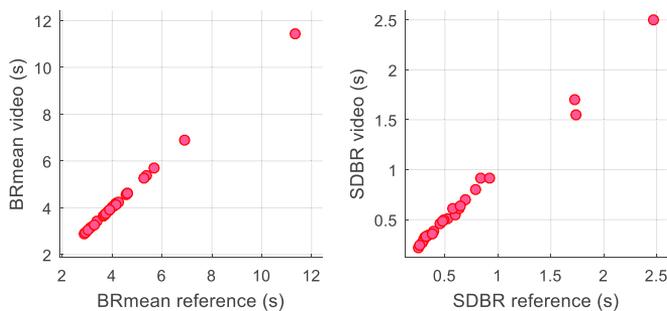

Fig. 8.8   Correlation scatter plots of the BRmean and SDBR parameters. Refer to Table 8.2 for r and ICC results.

Table 8.2   Statistical results of the BRmean and SDBR parameters

Parameter	r*	ICC*	MAE	RMSE
BRmean (s)	1,000	1,000	0,01	0,02
SDBR (s)	0,997	0,997	0,02	0,04

*: All p-values < 0.001. Refer to section 3.2 for statistical parameters definitions.
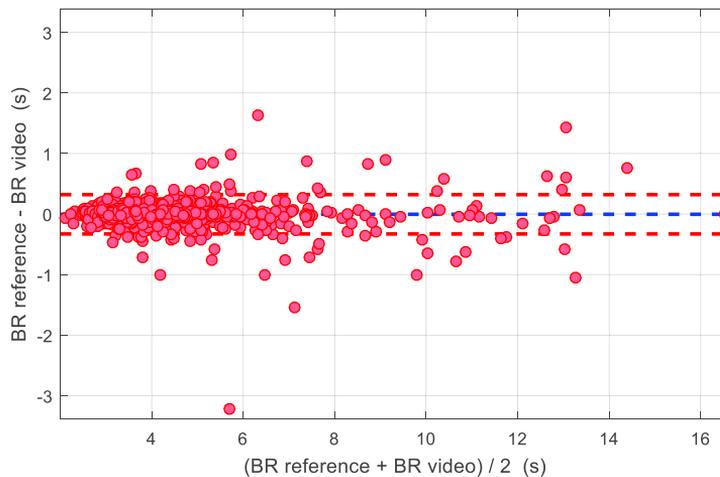
Fig. 8.9   Bland-Altman plot with LoA of 95% representing the agreement between the BR intervals of reference and video (data of all subjects combined in one dataset). The mean difference was -0.004 s (0.01 BPM) and the LoA from -0.33 to 0.32 s (-0.88 to 0.90 BPM).

Some authors proposed different non-contact video camera-based methods to estimate the instantaneous and average respiratory rate and reported statistical results of the measurements, which are summarized below. Bousefsaf et al. [29] proposed a motion-tolerant method to measure the instantaneous breathing rate from the face at rest and with predefined head movements. In this work, 12 healthy volunteers were recorded with a webcam during 35 s. In the two conditions of the study, the measurement of the average breathing rate achieved Pearson's correlation results of 0.99 and 0.98, respectively. In the case of the Bland-Altman plots, the results show LoA of 95% from -5.36 to 4.82 BPM at rest and during motion from -4.6 to 4.58 BPM.

In the study of van Gastel et al. [53], a non-contact camera-based method was presented to detect the instantaneous respiratory rate both in visible and dark lighting conditions. In this work, three subjects in a sitting position were recorded in the visible lighting condition during 120 or 150 s depending on different guided breathing scenarios. Including all these scenarios, the best statistical results reported in the visible light condition were r = 0.962, MAE = 1.74 BPM and RMSE = 2.67 BPM. In particular, the constant 20 BPM scenario achieved the best results with a MAE = 1.15 BPM and RMSE = 1.61 BPM.

Cobos and Abderrahim [58] proposed a method for the contactless measuring of the respiratory rate by employing the iPPG technique through wavelets. 10 healthy subjects in a sitting position were assessed in this work, in which 60 s were analyzed in their corresponding recordings. The Bland-Altman plot was used to evaluate the proposed method with results showing LoA of 95% from -1.8 to 2.0 BPM.

Alternatively, Al-Naji and Chahl [57] presented a non-contact monitoring system to simultaneously obtain the cardiorespiratory signal from a group of people. In total, 18 participants were recorded during 20 s with a digital camera. In this work, the respiratory rate per minute was measured in a stationary and non-stationary scenario by employing an iPPG-based method (color) and a motion-based method. The statistical results achieved in this study were as follows: (stationary scenario, iPPG-based method) r = 0.9893 and RMSE = 0.32 BPM; (stationary scenario, motion-based method) r = 0.9706 and RMSE = 0.64 BPM; (non-stationary scenario, iPPG-based method) r = 0.8455 and RMSE = 1.25 BPM; (non-stationary scenario, motion-based method) r = 0.6209 and RMSE = 2.08 BPM.

In comparison with the statistical data reported by these authors in stationary conditions, it is noted that our proposed method achieved better results.

Moreover, our study includes a larger number of subjects under analysis and the video recordings were longer. It is important to note that most of the works cited previously did not analyze the chest or abdomen area to obtain the respiratory signal (refer to section 2.4). Among these studies, only the work presented by Cobos and Abderrahim [58] analyzed the abdomen area in order to obtain the respiratory signal. On the other hand, it is not possible to compare the results obtained in these works in motion scenarios because our method was not evaluated in this condition.

On another note, the video signals that obtained the three minimum and three maximum MAPE were chosen by calculating the mode, with the aim of finding the seat belt marks that achieved the best and worst results. These results were divided according to the number of seat belt marks analyzed in the subjects (Table 8.3).

Table 8.3   Signals with the three minimum and three maximum MAPE of the BR intervals

No. seat belt marks	Subj. (N)	Minimum (best) MAPE			Maximum (worst) MAPE		
		1st	2nd	3rd	1st	2nd	3rd
6	8	s_{46}	s_{26}	s_{24}	s_{15}	s_{15}	s_{13}
8	12	s_{48}	s_{28}	s_{37}	s_{15}	s_{13}	s_{57}
10	5	s_{68}	s_{48}	s_{37}	s_{13}	s_{24}	s_{15}

Refer to Fig. 8.1 for examples of the reference seat belt marks (the location of the seat belt marks over the body depended mainly on the physical constitution of the participants). Refer to section 3.2 for statistical parameters definitions.

Based on the data presented in the table, it is noted that the signals which achieved better results were obtained from marks mainly located on lower-central positions (s_{46}, s_{48}, s_{68}), while the worst results were calculated from upper-central marks (s_{15}, s_{15}, s_{13}). In most of the subjects with 8 seat belt marks, who were the majority (N=12), and subjects with 10 marks, the signal s_{48} presented

low errors. These two reference points are not adjacent and it is also noted that six signals with minimum errors were obtained from non-adjacent reference points. Therefore, the acquisition of signals from reference points within a certain distance between them may be a good option.

8.4   Conclusions

In this chapter is presented a respiratory rate analysis by video using the KLT algorithm. The respiratory signals by video, which were obtained by using a seat belt of a car as a standard reference for all subjects, achieved a high correspondence compared to the reference signal. From the perspective of the video image, the body displacements caused by the respiration produce a variation visual effect on the distance between the reference marks. The inspiratory phase of the breath cause an approach of the reference marks to the camera that is interpreted as a longer distance between the reference points, and vice versa in the case of the expiratory phase. However, despite the high agreement between the measurements of the reference and video, it is important to note that the signals were obtained by measuring inductive changes and distances between reference points, respectively.

All subjects achieved excellent results in the BR time series with correlation coefficients above 0.9. The $\bar{x} \pm s$ of the Pearson and ICC measures were 0.968 ± 0.029 and 0.967 ± 0.030, respectively, and the mean of the individual MAPE results was 1.94 %. The BRmean and SDBR parameters achieved very high correlation coefficients with values above 0.99. In the case of the Bland-Altman plot, the measurements did not present large differences between the methods, which give the possibility to use them interchangeably, at least under the conditions of the study. Our proposed method achieved better results in comparison with those reported in other studies in stationary conditions. But, on

the other hand, it is not possible to compare the results obtained in motion scenarios in these works because of our method was not evaluated in this condition. Therefore, the acquisition of measurements in a non-stationary scenario is necessary to provide more clearly the scope and limitations of the proposed method in this condition in comparison with other related studies.

The video signals with the minimum and maximum errors assisted in identifying which regions of the body and distances between marks could provide better results. The signals that achieved better results were obtained from reference marks mainly located on lower-central positions, while the worst signals were calculated from upper-central marks. In general, the lower-central area of the torso could be the region that better reflects the movements of the respiration, although this depends on the subject. Moreover, the best results were obtained mostly from non-adjacent reference points. Therefore, the acquisition of signals from reference marks with a certain distance between them, as well as located on the lower-central region of the torso could be the best option using the proposed method. As an objective for future work, due to a seat belt of a car is used as a standard reference, an adaptation of the proposed method could be an alternative to measure the respiratory rate in subjects while driving, as well as in other conditions by tracking different reference marks.

Conclusions
and future work

This thesis presents a series of methods that aim at measuring physiological parameters such as heart rate, HRV, and respiratory rate through video imaging. The proposed methods employ image and signal processing techniques that focus on regions of the body that could provide more information on these physiological parameters and on the characteristics of the signals, respectively. Furthermore, an analysis of factors and conditions that influence the performance of this technique is carried out in this research study. The assessment of different cameras and settings, the acquisition of measurements in stationary and motion scenarios, and the evaluation of different light sources are some examples, as detailed throughout the chapters of this thesis.

The initial measurements of this research study are obtained by performing a frame averaging and a cross-correlation analysis achieving a good agreement on heart rate and HRV measurements. In general, these results are improved in comparison with the data reported by authors who employed a similar method. However, not all the reported results achieved a good correspondence on the measurements. A comparative analysis of low-resolution and high-resolution recordings shows that the video resolution has a significant impact on the overall results. Certainly, a higher resolution video presents more pixels to be analyzed and, therefore, a greater amount of data to obtain more reliable measurements. Moreover, the frame averaging could affect the results of the low-resolution recordings more, since the amount of data available to obtain the measurements was reduced.

Despite the good performance of the aforementioned method, it has the limitation that it can only be used in stationary conditions, thereby affecting its application in real-life scenarios. Then, a method for performing a video HRV analysis in motion conditions is proposed, in which the application of face detection approaches plays a key role in the achievement of this objective. The Viola and Jones and the KLT algorithms are employed in this work as part of the face detection and tracking method. Even at present, the Viola and Jones algorithm is widely used in several applications and research, since it is able to run in real-time and it is characterized by achieving high detection rates. In the case of the KLT algorithm, the approach is able to track the face of a subject during changes of scale and orientation, as occurred in the motion scenarios of this study.

The proposed method, which applies face detection and tracking approaches, aims to analyze the regions that contain a greater cardiac component (forehead and cheeks) and exclude those that may contribute with artifacts. A frame averaging is not necessary in this case, which allows having a greater amount of

data for the acquisition of the measurements. A first assessment included stationary and motion conditions that were performed by the participants. The statistical analysis shows a good agreement between the reference system and the proposed method by video imaging. The results obtained in stationary conditions are improved in comparison with the data reported in related works. In the case of the motion condition analysis, the reported results also achieved a good correspondence but with lower values in relation to the stationary condition. An overall comparative analysis of HRV parameters in motion conditions was more limited due to the lack of studies or studies containing insufficient data analysis.

Taking into account the similar results obtained in stationary conditions by employing the two methods mentioned above, the last one seems to be a better alternative for measuring heart rate and HRV parameters in conditions as those carried out in the study. Accordingly, the method could be applied while working in front of a PC or in other activities with similar moves. It is important to note that the measurements were performed by following specific movements, therefore, its application in a more real-life scenario was also proposed in this research study.

On the other hand, a video HRV analysis was carried out through the acquisition of the iPPG signal in real-time. This option offers a quick analysis of the HRV parameters by video, as well as the feasibility of real-time monitoring of the iPPG signal under influences such as the motion and the lighting conditions. In this analysis, the recordings were obtained using three different light sources in order to evaluate their influence on the measurements. Although good results were obtained using the three light sources, the statistical analysis shows that the sunlight achieved the best results in comparison with the LED and fluorescent light alternatives. The sunlight source has an energy spectrum that is more evenly distributed compared with the LED and fluorescent lights,

which present their energy more concentrated in some wavelengths. Thus, hypothetically, the energy level provided by the light source at the analyzed wavelength could be a factor that influences the quality of the signals. Even so, this study shows that some artificial lights can achieve very good and close results as those obtained with the sunlight source.

Furthermore, the interest in measuring physiological parameters in everyday activities by using alternative methods has increased in the last few years. Hence, a HCI scenario is performed in this work with the aim of assessing the proposed method in a real-life condition. Due to the wider variety of motion, some improvements are introduced in order to achieve a more robust face tracking of the participants. This motion scenario adversely affected the quality of the video signals and, thus, the reliability of the HRV measurements. A quality distinction of the video signal is proposed in order to only analyze the segments that contain a minor amount of motion artifacts. In general, the statistical analysis shows a lower agreement on the measurements in comparison with the results obtained in the preceding chapters. Likewise, based on the overall results, the proposed method achieved lower errors in comparison with those obtained in some related works.

Besides the measurement of heart rate and HRV, the respiratory rate analysis by video also has been explored in the last few years. In this study, the KLT algorithm is applied to track the body movements caused by the respiration in order to obtain the respiratory signal. All participants obtained excellent results in measuring the instantaneous respiratory rate, as well as in BRmean and SDBR parameters. These measurements achieved a better agreement in comparison with studies that were also performed in stationary conditions. However, the proposed method was not evaluated in a motion scenario, which does not allow knowing its scope and limitations in this condition. Since a seat belt of a car is used as a standard reference to obtain the respiratory signals, an adaptation of

the method could be an alternative to measure the respiratory rate in subjects while driving, as well as in other conditions by tracking different reference marks.

Based on the statistical analysis conducted in this thesis, the methods proposed herein could provide a low-cost, contactless, and reliable alternative for measuring physiological parameters in non-clinical environments. These results and data reported in related studies are strong evidence that the video imaging technique is a promising approach that could be used in the near future. Moreover, although the technique aims to be used by everyone, it may have a potential application in people with sensitive or fragile skin (neonates or patients with burn injuries, for example).

This contactless technique may eventually become an instrument to detect physical or psychological disorders in the future. Nevertheless, its use for this purpose will depend on the progress of the technique over the coming years, since it is still in the research and development phase. Therefore, more improvements are necessary to reach the reliability achieved by the current reference systems and, particularly, if its application in real-life scenarios is considered. The development of more robust algorithms is required in order to suppress, as much as possible, the contribution of artifacts present in a real environment. Consequently, the acquisition of measurements in several real-life scenarios, longer recordings, and the analysis of more factors that could influence the performance of the technique constitute some objectives for future work.

150

References

1. "The top 10 causes of death". *World Health Organization (WHO)*, 10 March 2017, www.who.int/.

2. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use", *European Heart Journal*, vol. 17, no. 3, pp. 354-381, 1996.

3. U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim and J. S. Suri, "Heart rate variability: a review", *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031-1051, 2006.

4. M. Baumert, L. Brechtel, J. Lock and A. Voss, "Changes in heart rate variability of athletes during a training camp", *Biomedizinische Technik*, vol. 51, no. 4, pp. 201-204, 2006.

5. J. F. Thayer, S. S. Yamamoto and J. F. Brosschot, "The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors", *International Journal of Cardiology*, vol. 141, no.2, pp. 122-131, 2010.

6. J. C. Cervantes, G. Rodas and L. Capdevila, "Heart-rate variability and precompetitive anxiety in swimmers", *Psicothema*, vol. 21, no.4, pp. 531-536, 2009.

7. G. Rodas, C. Pedret, J. Ramos, L. Capdevila, "Variabilidad de la frecuencia cardíaca: concepto, medidas y relación con aspectos clínicos (I)", *Archivos de medicina del deporte: revista de la Federación Española de Medicina del Deporte y de la Confederación Iberoamericana de Medicina del Deporte*, vol. XXV, no. 123, pp. 41-47, 2008.

8. Tortora, G. and Derrickson, B. *Principles of anatomy and physiology*. 12th ed. Hoboken, NJ: John Wiley & Sons, Inc., pp. 547-548, 2009.

9. A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram", *International Journal of Cardiology*, vol. 166, no. 1, pp. 15-29, 2013.

10. E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi and P. Laguna, "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions", *Physiological Measurement*, vol. 31, no. 9, pp. 1271-1290, 2010.

11. J. Kranjec, S. Beguš, G. Geršak and J. Drnovšek, "Non-contact heart rate and heart rate variability measurements: A review", *Biomedical Signal Processing and Control*, vol. 13, pp. 102-112, 2014.

12. Y. Sun and N. Thakor, "Photoplethysmography revisited: from contact to noncontact, from point to imaging", *IEEE Transactions on Bio-Medical Engineering*, vol. 63, no.3, pp. 463-477, 2016.

13. M. A. Hassan, A. S. Malik, D. Fofi, N. Saad, B. Karasfi, Y. S. Ali and F. Meriaudeau, "Heart rate estimation using facial video: A review", *Biomedical Signal Processing and Control*, vol. 38, pp. 346-360, 2017.

14. A. Al-Naji, K. Gibson, S. H. Lee and J. Chahl, "Monitoring of cardiorespiratory signal: principles of remote measurements and review of methods", *IEEE Access*, vol. 5, pp. 15776-15790, 2017.

15. T. Lister, P. A. Wright and P. H. Chappell, "Optical properties of human skin", *Journal of Biomedical Optics*, vol. 17, no. 9, 90901-1, 2012.

16. W. J. Cui, L. E. Ostrander and B. Y. Lee, "In vivo reflectance of blood and tissue as a function of light wavelength", *IEEE Transactions on Bio-Medical Engineering*, vol. 37, no. 6, pp. 632-639, 1990.

17. Y. Maeda, M. Sekine and T. Tamura, "The advantages of wearable green reflected photoplethysmography", *Journal of Medical Systems*, vol. 35, no. 5, pp. 829-834, 2011.

18. J. Lee, K. Matsumura, K. Yamakoshi, P. Rolfe, S. Tanaka and T. Yamakoshi, "Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1724-1727, 2013.

19. M. Z. Poh, D. J. McDuff and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation", *Optics Express*, vol. 18, no. 10, pp. 10762-10774, 2010.

20. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

21. R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection", *International Conference on Image Processing*, 2002.

22. H. Monkaresi, R. A. Calvo and H. Yan, "A machine learning approach to improve contactless heart rate monitoring using a webcam", *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1153-1160, 2014.

23. M. Z. Poh, D. J. McDuff and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam", *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 1, pp. 7-11, 2011.

24. Y. Sun, S. Hu, V. Azorin-Peris, R. Kalawsky and S. Greenwald, "Noncontact imaging photoplethysmography to effectively access pulse rate variability", *Journal of Biomedical Optics*, vol. 18, no. 6, 061205, 2013.

25. D. McDuff, S. Gontarek and R. W. Picard, "Improvements in remote cardiopulmonary measurement using a five band digital camera", *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2593–2601, 2014.

26. D. McDuff, S. Gontarek and R. W. Picard, "Remote detection of photoplethysmographic systolic and diastolic peaks using a digital camera", *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2948–2954, 2014.

27. J. Moreno, J. Ramos-Castro, J. Movellan, E. Parrado, G. Rodas and L. Capdevila, "Facial video-based photoplethysmography to detect HRV at rest", *International Journal of Sports Medicine*, vol. 36, no. 6, pp. 474-480, 2015.

28. K. Alghoul, S. Alharthi, H. Al Osman and A. El Saddik, "Heart rate variability extraction from videos signals: ICA vs. EVM comparison", *IEEE Access*, vol. 5, pp. 4711-4719, 2017.

29. F. Bousefsaf, C. Maaoui and A. Pruski, "Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate", *Biomedical Signal Processing and Control*, vol. 8, no.6, pp. 568-574, 2013.

30. M. Kumar, A. Veeraraghavan and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera", *Biomedical Optics Express*, vol. 6, no. 5, pp. 1565-1588, 2015.

31. C. H. Antink, H. Gao, C. Brüser and S. Leonhardt, "Beat-to-beat heart rate estimation fusing multimodal video and sensor data", *Biomedical Optics Express*, vol. 6, no. 8, pp. 2895-2907, 2015.

32. R. Y. Huang and L. R. Dung, "Measurement of heart rate variability using off-the-shelf smart phones", *BioMedical Engineering OnLine*, 15:11, 2016.

33. A. De Groote, M. Wantier, G. Cheron, M. Estenne and M. Paiva, "Chest wall motion during tidal breathing", *Journal of Applied Physiology*, vol. 83, no. 5, pp. 1531-1537, 1997.

34. S. Wiesner and Z. Yaniv, "Monitoring Patient Respiration using a Single Optical Camera", *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2740-2743, 2007.

35. W. Verkruysse, L. O. Svaasand and J. S. Nelson, "Remote plethysmographic imaging using ambient light", *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.

36. Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers and Y. Zhu, "Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise", *Journal of Biomedical Optics*, vol. 16, no. 7, 077010, 2011.

37. M. Bartula, T. Tigges and J. Muehlsteff, "Camera-based system for contactless monitoring of respiration", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2672-2675, 2013.

38. F. Bastianini, E. Schena, P. Saccomandi and S. Silvestri, "Accuracy evaluation of dynamic volume measurements performed by opto-electronic plethysmograph, by using a pulmonary simulator", *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 930-933, 2013.

39. L. Ding, H. Zhang and Y. Xie, "Respiratory motion tracking with a multi-camera vision system", *IEEE International Conference on Medical Imaging Physics and Engineering*, pp. 352-355, 2013.

40. F. Zhao, M. Li, Y. Qian and J. Z. Tsien, "Remote Measurements of Heart and Respiration Rates for Telemedicine", *PLoS ONE*, vol. 8, no. 10, 2013.

41. T. Lukáč, J. Púčik and L. Chrenko, "Contactless recognition of respiration phases using web camera", *24th International Conference Radioelektronika*, pp. 1-4, 2014.

42. M. Villarroel, A. Guazzi, J. Jorge, S. Davis, P. Watkinson, G. Green, A. Shenvi, K. McCormick and L. Tarassenko, "Continuous non-contact vital sign monitoring in neonatal intensive care unit", *Healthcare Technology Letters*, vol. 1, no. 3, pp. 87-91, 2014.

43. D. Alinovi, L. Cattani, G. Ferrari, F. Pisani and R. Raheli, "Spatio-temporal video processing for respiratory rate estimation", *IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, pp. 12-17, 2015.

44. C. Y. Fang, H. H. Hsieh and S. W. Chen, "A Vision-Based Infant Respiratory Frequency Detection System", *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1-8, 2015.

45. A. Venkitaraman and V. V. Makkapati, "Motion-based segmentation of chest and abdomen region of neonates from videos", *Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, pp. 1-5, 2015.

46. A. Chatterjee, A. P. Prathosh, P. Praveena and V. Upadhya, "Real-Time Visual Respiration Rate Estimation with Dynamic Scene Adaptation", *IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 154-160, 2016.

47. D. L. Dorantes, D. Santoyo, M. Á. Martínez, J. Rodríguez and Y. Oliveri, "Noninvasive monitoring system for early detection of apnea in newborns and infants", *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 494-498, 2016.

48. R. Janssen, W. Wang, A. Moço and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection", *Physiological Measurement*, vol. 37, no. 1, pp. pp. 100-114, 2016.

49. V. V. Makkapati and S. S. Rambhatla, "Remote monitoring of camera based respiration rate estimated by using occlusion of dot pattern", *IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1-5, 2016.

50. Y. Nam, Y. Kong, B. Reyes, N. Reljin and K. H. Chon, "Monitoring of Heart and Breathing Rates Using Dual Cameras on a Smartphone", *PLoS ONE*, vol. 11, no. 3, 2016.

51. D. M. Tveit, K. Engan, I. Austvoll and Ø. Meinich-Bache, "Motion based detection of respiration rate in infants using video", *IEEE International Conference on Image Processing (ICIP)*, pp. 1225-1229, 2016.

52. V. Upadhya, A. Chatterjee, A. P. Prathosh and P. Praveena, "Respiration Monitoring through Thoraco-Abdominal Video with an LSTM", *IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 165-171, 2016.

53. M. van Gastel, S. Stuijk and G. de Haan, "Robust respiration detection from remote photoplethysmography", *Biomedical Optics Express*, vol. 7, no. 12, pp. 4941–4957, 2016.

54. B. Xu, H. Madhu, R. Kulkarni, L. K. Mestha, S. Kyal and G. Pennington, "Evaluating and improving the robustness of a video-based PR/RR monitoring system for a clinical environment", *IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pp. 204-207, 2016.

55. P. S. Addison, D. Jacquel, D. M. H. Foo, A. Antunes and U. R. Borg, "Video-Based Physiologic Monitoring During an Acute Hypoxic Challenge: Heart Rate, Respiratory Rate, and Oxygen Saturation", *Anesthesia and Analgesia*, vol. 125, no. 3, pp. 860-873, 2017.

56. S. Alam, S. P. N. Singh and U. Abeyratne, "Considerations of handheld respiratory rate estimation via a stabilized Video Magnification approach", *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4293-4296, 2017.

57. A. Al-Naji and J. Chahl, "Simultaneous Tracking of Cardiorespiratory Signals for Multiple Persons Using a Machine Vision System With Noise Artifact Removal", *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 5, pp. 1-10, 2017.

58. J. C. Cobos and M. Abderrahim, "Measuring Heart and Breath Rates by Image Photoplethysmography using Wavelets Technique", *IEEE Latin America Transactions*, vol. 15, no. 10, pp. 1864-1868, 2017.

59. M. A. Hassan, A. S. Malik, D. Fofi, N. Saad and F. Meriaudeau, "Novel health monitoring method using an RGB camera", *Biomedical Optics Express*, vol. 8, no. 11, pp. 4838–4854, 2017.

60. C. Liu, Y. Yang, F. Tsow, D. Shao and N. Tao, "Noncontact spirometry with a webcam", *Journal of Biomedical Optics*, vol. 22, no. 5, 2017.

61. A. P. Prathosh, P. Praveena, L. K. Mestha and S. Bharadwaj, "Estimation of Respiratory Pattern From Video Using Selective Ensemble Aggregation", *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2902-2916, 2017.

62. B. A. Reyes, N. Reljin, Y. Kong, Y. Nam and K. H. Chon, "Tidal Volume and Instantaneous Respiration Rate Estimation using a Volumetric Surrogate Signal Acquired via a Smartphone Camera", *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 764-777, 2017.

63. Q. V. Tran, S. F. Su, C. C. Chuang, V. T. Nguyen and N. Q. Nguyen, "Real-time non-contact breath detection from video using adaboost and Lucas-Kanade algorithm", *Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS)*, pp. 1-4, 2017.

64. M. Villarroel, J. Jorge, C. Pugh and L. Tarassenko, "Non-Contact Vital Sign Monitoring in the Clinic", *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 278-285, 2017.

65. B. Wei, X. He, C. Zhang and X. Wu, "Non-contact, synchronous dynamic measurement of respiratory rate and heart rate based on dual sensitive regions", *BioMedical Engineering OnLine*, 16:17, 2017.

66. C. Ordóñez, C. Cabo, A. Menéndez, and A. Bello, "Detection of human vital signs in hazardous environments by means of video magnification", *PLoS ONE*, vol. 13, no. 4, 2018.

67. S. Sanyal and K. K. Nundy, "Algorithms for Monitoring Heart Rate and Respiratory Rate From the Video of a User's Face", *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1-11, 2018.

68. A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky and S. Leonhardt, "Neonatal non-contact respiratory monitoring based on real-time infrared thermography", *BioMedical Engineering OnLine*, 10:93, 2011.

69. J. Guirao Aguilar, J. G. Bellika, L. Fernandez Luque, V. Traver Salcedo, "Respiration Tracking Using the Wii Remote Game-Controller", *Studies in Health Technology and Informatics*, vol. 169, pp. 455-459, 2011.

70. Y. Peng, S. Vedam, S. Gao, P. Balter, "A new respiratory monitoring and processing system based on Wii remote: proof of principle", *Medical Physics*, vol. 40, no. 7, 071712, 2013.

71. S. L. Bennett, R. Goubran and F. Knoefel, "The detection of breathing behavior using Eulerian-enhanced thermal video", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7474-7477, 2015.

72. C. B. Pereira, X. Yu, V. Blazek and S. Leonhardt, "Robust remote monitoring of breathing function by using infrared thermography", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4250-4253, 2015.

73. A. H. Alkali, R. Saatchi, H. Elphick and D. Burke, "Thermal image processing for real-time non-contact respiration rate monitoring", *IET Circuits, Devices & Systems*, vol. 11, no. 2, pp. 142-148, 2017.

74. R. Chauvin, M. Hamel, S. Brière, F. Ferland, F. Grondin, D. Létourneau, M. Tousignant and F. Michaud, "Contact-Free Respiration Rate Monitoring Using a

Pan–Tilt Thermal Camera for Stationary Bike Telerehabilitation Sessions", *IEEE Systems Journal*, vol. 10, no. 3, pp. 1046-1055, 2016.

75. J. Rumiński, "Analysis of the parameters of respiration patterns extracted from thermal image sequences", *Biocybernetics and Biomedical Engineering*, vol. 36, no. 4, pp. 731-741, 2016.

76. X. He, R. Goubran and F. Knoefel, "IR night vision video-based estimation of heart and respiration rates", *IEEE Sensors Applications Symposium (SAS)*, pp. 1-5, 2017.

77. F. Deng, J. Dong, X. Wang, Y. Fang, Y. Liu, Z. Yu, J. Liu and F. Chen, "Design and Implementation of a Noncontact Sleep Monitoring System Using Infrared Cameras and Motion Sensor", *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 7, pp. 1555-1563, 2018.

78. M. Hu, G. Zhai, D. Li, Y. Fan, H. Duan, W. Zhu and X. Yang, "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation", *PLoS ONE*, vol. 13, no. 1, 2018.

79. M. C. Yu, J. L. Liou, S. W. Kuo, M. S. Lee and Y. P. Hung, "Noncontact respiratory measurement of volume change using depth camera", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2371-2374, 2012.

80. F. Benetazzo, A. Freddi, A. Monteriù and S. Longhi, "Respiratory rate detection algorithm based on RGB-D camera: theoretical background and experimental results", *Healthcare Technology Letters*, vol. 1, no. 3, pp. 81-86, 2014.

81. N. Bernacchia, L. Scalise, L. Casacanditella, I. Ercoli, P. Marchionni and E. P. Tomasini, "Non contact measurement of heart and respiration rates based on Kinect™", *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pp. 1-5, 2014.

82. S. H. Lim, E. Golkar and A. A. Abd. Rahni, "Respiratory motion tracking using the kinect camera", *IEEE Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 797-800, 2014.

83. A. Procházka, M. Schätz, O. Vyšata and M. Vališ, "Microsoft Kinect Visual and Depth Sensors for Breathing and Heart Rate Analysis", *Sensors*, vol. 16, no. 7, 996, 2016.

84. A. Procházka, H. Charvátová, O. Vyšata, J. Kopal and J. Chambers, "Breathing Analysis Using Thermal and Depth Imaging Camera Video Records", *Sensors*, vol. 17, no. 6, 1408, 2017.

85. D. Dei, G. Grazzini, G. Luzi, M. Pieraccini, C. Atzeni, S. Boncinelli, G. Camiciottoli, W. Castellani, M. Marsili and J. L. Dico, "Non-Contact Detection of Breathing Using a Microwave Sensor", *Sensors*, vol. 9, no. 4, pp. 2574–2585, 2009.

86. Y. S. Lee, P. N. Pathirana, C. L. Steinfort and T. Caelli, "Monitoring and Analysis of Respiratory Patterns Using Microwave Doppler Radar", *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 2, pp. 1-12, 2014.

87. T. Hall, D. Y. C. Lie, T. Q. Nguyen, J. C. Mayeda, P. E. Lie, J. Lopez and R. E. Banister, "Non-Contact Sensor for Long-Term Continuous Vital Signs Monitoring: A Review on Intelligent Phased-Array Doppler Sensor Design", *Sensors*, vol. 17, no. 11, 2632, 2017.

88. S. Fleming, L. Tarassenko, M. Thompson and D. Mant, "Non-invasive measurement of respiratory rate in children using the photoplethysmogram", *30th Annual*

International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1886-1889, 2008.

89. E. M. Lee, N. H. Kim, N. T. Trang, J. H. Hong, E. J. Cha and T. S. Lee, "Respiratory rate detection algorithms by photoplethysmography signal processing", *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1140-1143, 2008.

90. K. H. Chon, S. Dash and K. Ju, "Estimation of Respiratory Rate From Photoplethysmogram Data Using Time–Frequency Spectral Estimation", *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2054-2063, 2009.

91. S. Dash, K. H. Shelley, D. G. Silverman and K. H. Chon, "Estimation of Respiratory Rate From ECG, Photoplethysmogram, and Piezoelectric Pulse Transducer Signals: A Comparative Study of Time–Frequency Methods", *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1099-1107, 2010.

92. J. Li, J. Jin, X. Chen, W. Sun and P. Guo, "Comparison of respiratory-induced variations in photoplethysmographic signals", *Physiological Measurement*, vol. 31, no. 3, pp. 415-425, 2010.

93. D. Wu, G. Z. Liu, M. Y. M. Wong and Y. T. Zhang, "The accuracy of respiratory rate estimation using electrocardiography and photoplethysmography", *10th IEEE International Conference on Information Technology and Applications in Biomedicine*, pp. 1-3, 2010.

94. J. Lázaro, E. Gil, R. Bailón and P. Laguna, "Deriving respiration from the pulse photoplethysmographic signal", *Computing in Cardiology*, pp. 713-716, 2011.

95. M. Orini, M. D. Peláez-Coca, R. Bailón and E. Gil, "Estimation of spontaneous respiratory rate from photoplethysmography by cross time-frequency analysis", *Computing in Cardiology*, pp. 661-664, 2011.

96. W. Karlen, S. Raman, J. M. Ansermino and G. A. Dumont, "Multiparameter Respiratory Rate Estimation From the Photoplethysmogram", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946-1953, 2013.

97. D. Labate, F. L. Foresta, G. Occhiuto, F. C. Morabito, A. Lay-Ekuakille and P. Vergallo, "Empirical Mode Decomposition vs. Wavelet Decomposition for the Extraction of Respiratory Signal From Single-Channel ECG: A Comparison", *IEEE Sensors Journal*, vol. 13, no. 7, pp. 2666-2674, 2013.

98. L. M. Nilsson, "Respiration Signals from Photoplethysmography", *Anesthesia & Analgesia*, vol. 117, no. 4, pp. 859–865, 2013.

99. J. Lázaro, R. Bailón, P. Laguna, Y. Nam, K. Chon and E. Gil, "Respiratory rate influence in the resulting magnitude of pulse photoplethysmogram derived respiration signals", *Computing in Cardiology*, pp. 289-292, 2014.

100. W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino and G. A. Dumont, "Respiratory rate assessment from photoplethysmographic imaging", *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5397-5400, 2014.

101. J. Lázaro, Y. Nam, E. Gil, P. Laguna and K. H. Chon, "Smartphone-camera-acquired pulse photoplethysmographic signal for deriving respiratory rate", *8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)*, pp. 121-122, 2014.

102. Y. Nam, J. Lee and K. H. Chon, "Respiratory rate estimation from the built-in cameras of smartphones and tablets", *Annals of Biomedical Engineering*, vol. 42, no. 4, pp. 885–898, 2014.

103. W. Karlen, A. Garde, D. Myers, C. Scheffer, J. M. Ansermino and G. A. Dumont, "Estimation of Respiratory Rate From Photoplethysmographic Imaging Videos Compared to Pulse Oximetry", *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, pp. 1331-1338, 2015.

104. R. C. Peng, W. R. Yan, N. L. Zhang, W. H. Lin, X. L. Zhou and Y. T. Zhang, "Investigation of Five Algorithms for Selection of the Optimal Region of Interest in Smartphone Photoplethysmography", *Journal of Sensors*, vol. 2016, Article ID 6830152, 7 pages, 2016.

105. Y. Nam and Y. C. Nam, "Photoplethysmography Signal Analysis for Optimal Region-of-Interest Determination in Video Imaging on a Built-In Smartphone under Different Conditions", *Sensors*, vol. 17, no. 10, 2385, 2017.

106. E. Hjelmås and B. K. Low, "Face Detection: A Survey", *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236-274, 2001.

107. M. H. Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.

108. C. Zhang and Z. Zhang, "A survey of recent advances in face detection", *Technical report, Microsoft Research*, 2010.

109. S. Zafeiriou, C. Zhang and Z. Zhang, "A survey on face detection in the wild: Past, present and future", *Computer Vision and Image Understanding*, vol. 138, pp. 1-24, 2015.

110. P. Viola and M. Jones, "Robust real-time face detection", *Eighth IEEE International Conference on Computer Vision (ICCV)*, pp. 747-747, 2001.

111. Bradski, G. and Kaehler, A. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Inc., 2008.

112. M. Mathias, R. Benenson, M. Pedersoli and L. Van Gool, "Face Detection without Bells and Whistles", *European Conference on Computer Vision (ECCV)*, pp. 720-735, 2014.

113. B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", *7th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 674-679, 1981.

114. C. Tomasi and T. Kanade, "Detection and tracking of point features", *International Journal of Computer Vision*, 1991.

115. J. Shi and C. Tomasi, "Good features to track", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600, 1994.

116. "vision.PointTracker System object: Track points in video using Kanade-Lucas-Tomasi (KLT) algorithm". *MathWorks*®, 15 February 2018, www.mathworks.com/.

117. Z. Kalal, K. Mikolajczyk and J. Matas, "Forward-Backward Error: Automatic Detection of Tracking Failures", *20th International Conference on Pattern Recognition*, pp. 2756-2759, 2010.

118. "estimateGeometricTransform: Estimate geometric transform from matching point pairs". *MathWorks*®, 15 February 2018, www.mathworks.com/.

119. P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry", *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 138-156, 2000.

120. Fleiss, J. L. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons, Inc., pp. 1-32, 1999.

121. J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement", *The Lancet*, vol. 327, no. 8476, pp. 307-310, 1986.

122. J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm", *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230-236, 1985.

123. L. I-K. Lin, "A concordance correlation coefficient to evaluate reproducibility", *Biometrics*, vol. 45, no. 1, pp. 255-268, 1989.

124. N. Rodriguez-Ibañez, M. A. García-Gonzalez, M. A. F. de la Cruz, M. Fernández-Chimeno, J. Ramos-Castro, "Changes in heart rate variability indexes due to drowsiness in professional drivers measured in a real environment", *Computing in Cardiology*, pp. 913-916, 2012.

125. Y. L. Chang, Y. C. Feng and O. T. C. Chen, "Real-time physiological and facial monitoring for safe driving", *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4849-4852, 2016.

126. Z. Guo, Z. J. Wang and Z. Shen, "Physiological parameter monitoring of drivers based on video data and independent vector analysis", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4374-4378, 2014.

127. H. Qi, Z. J. Wang and C. Miao, "Non-contact driver cardiac physiological monitoring using video data", *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 418-422, 2015.

128. T. Blöcher, J. Schneider, M. Schinle and W. Stork, "An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection", *IEEE Sensors Applications Symposium (SAS)*, pp. 1-6, 2017.

129. S. Kyal, L. K. Mestha, B. Xu and J. Couderc, "A method to detect cardiac arrhythmias with a webcam", *IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pp. 1-5, 2013.

130. J. P. Couderc, S. Kyal, L. K. Mestha, B. Xu, D. R. Peterson, X. Xia, B. Hall, "Pulse harmonic strength of facial video signal for the detection of atrial fibrillation", *Computing in Cardiology*, pp. 661-664, 2014.

131. G. de Haan and V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG", *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2878-2886, 2013.

132. G. de Haan and A. van Leest, "Improved motion robustness of remote-PPG by using the blood volume pulse signature", *Physiological Measurement*, vol. 35, no. 9, pp. 1913-1926, 2014.

133. W. Wang, S. Stuijk and G. de Haan, "Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG", *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 415-425, 2015.

134. M. van Gastel, S. Stuijk and G. de Haan, "Motion Robust Remote-PPG in Infrared", *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 5, pp. 1425-1433, 2015.

Appendix

In this appendix is included supplementary information that is cited in the thesis.

A.1   Electrical diagram of the ECG acquisition board

Fig. A.1 shows the electrical diagram of the ECG acquisition board used to record the reference ECG signal (refer to section 3.1.1).
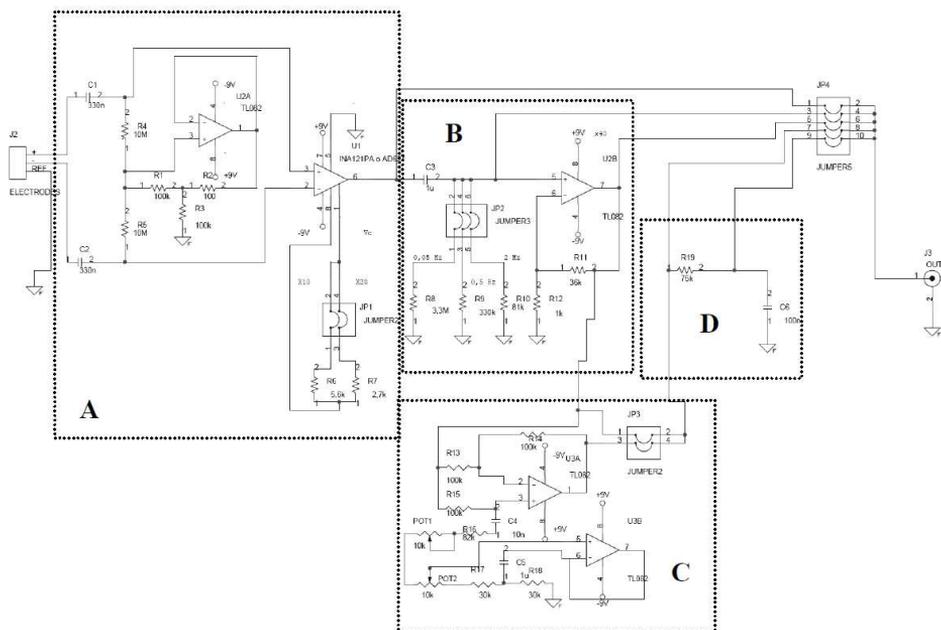
Fig. A.1   Electrical diagram of the ECG acquisition board: (A) Preconditioning and first amplification stage; (B) high-pass filter and second amplification stage; (C) notch filter; (D) low-pass filter.

Each stage of the shown electrical diagram is more detailed below:

A. **Preconditioning and first amplification stage:** this circuit stage has an instrumentation amplifier AD627 in which the gain depends on the resistor connected between the pins 1 and 8. The gain is set according to:

$$Gain = 5 + \left(200\frac{k\Omega}{R_G}\right) \qquad (Eq. A. 1)$$

in which R_G is selected between 5.6 kΩ or 2.7 kΩ with the jumper JP1.

B.   High-pass filter and second amplification stage: an RC high-pass filter is applied in order to eliminate low-frequency noise components such as random movements or the respiration component (Fig. A.2).
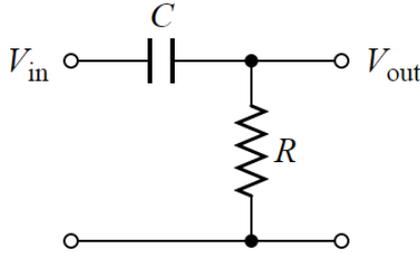

Fig. A.2   RC high-pass filter.

Its cutoff frequency is determined by:

$$f_c = \frac{1}{2\pi RC} \qquad \text{(Eq. A. 2)}$$

in which $C = 1$ μF and R is selected between 3.3 MΩ, 330 kΩ or 81 kΩ with the jumper JP2. Thus, the cutoff frequencies are 0.048 Hz, 0.482 Hz or 1.965 Hz, respectively. After the high-pass filter, an amplifier TL082 is used as part of a second amplification stage. The gain of the corresponding non-inverting amplifier configuration is set according to:

$$Gain = 1 + \left(\frac{R_F}{R_G}\right) \qquad \text{(Eq. A. 3)}$$

in which $R_F = 36$ kΩ and $R_G = 1$ kΩ.

163

C. **Notch filter:** this band-stop filter is used to eliminate the presence of the electrical power line noise (50 Hz). The frequency response shown in Fig. A.3 is set with the potentiometers POT1 and POT2. The POT2 sets the center frequency and the POT1 sets its attenuation.
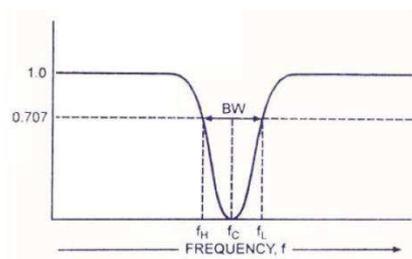

Fig. A.3 Frequency response of the notch filter.

D. **Low-pass filter:** an RC low-pass filter is applied in order to eliminate high-frequency noise components such as electronic noise or electromyogram (EMG) components (Fig. A.4). The values of the filter components are $R = 75$ kΩ and $C = 100$ nF, therefore, the cutoff frequency is 21.22 Hz.
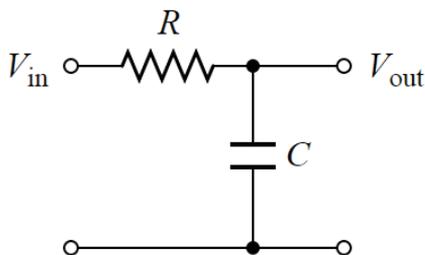

Fig. A.4 RC low-pass filter.

A.2   Questionnaire of the HCI scenario

The questionnaire completed by the participants during the HCI scenario (refer to section 7.1.1) is shown below. This questionnaire contains questions related to the psychological status, dietary habits, physical activity, and general data such as age, gender, weight, among others.

DATOS GENERALES

Instrucciones: rellene (datos entre paréntesis) o seleccione de acuerdo a la información solicitada. Se detallan las opciones de respuesta entre llaves en cada pregunta, cuando corresponda.

- Edad (años)
- Sexo {M; F}
- Peso (kg)
- Estatura (cm)
- ¿Fuma? ¿Cantidad de cigarros al día? {No fumo; Menos de 5; Entre 5 y 10; Entre 10 y 15; Más de 15}
- ¿Consume alcohol? ¿Cuántos días a la semana?{No bebo; 1; 2; 3; 4; 5; 6; 7}
- ¿Tiene alguna enfermedad cardiovascular? {Sí; No}
- ¿Está tomando actualmente algún tipo de medicamento? {Sí; No}

ESTADO PSICOLÓGICO

Instrucciones: seleccione en el rango de 0 a 10 (0=nada, 10=máximo) para indicar cómo te sientes en este momento.

- Tenso/a
- Enfadado/a
- Enérgico/a
- Nervioso/a
- Ansioso/a
- Deprimido/a

HÁBITOS ALIMENTICIOS

Instrucciones: *seleccione de acuerdo a la información solicitada. Se detallan las opciones de respuesta entre llaves en cada pregunta.*

- ¿Sigues una alimentación saludable? {Sí; Regular; No}

- En caso afirmativo ¿hace cuánto tiempo aproximadamente? {N/A; 1 semana; 1 mes; 3 meses; 6 meses; Más de 6 meses}

- ¿Cuántas veces comes al día? {1; 2; 3; 4; 5; Más de 5}

- ¿Cuál es la comida principal para ti? {Desayuno; Comida; Cena}

- ¿Regularmente en qué consiste tu comida entre semana? {Comida casera; Comida en bares/restaurantes; Comida precocinada y/o congelada; Otro}

- ¿Regularmente en qué consiste tu comida los fines de semana? {Comida casera; Comida en bares/restaurantes; Comida precocinada y/o congelada; Otro}

ACTIVIDAD FÍSICA

Instrucciones: *seleccione de acuerdo a la información solicitada. Se detallan las opciones de respuesta entre llaves en cada pregunta.*

- ¿Cómo es tu actual condición física? {Estoy sin forma; Débil; Regular; Buena; Perfecta}

- ¿Realizar algún deporte/actividad física de forma regular? {Sí; No}

- ¿Cuántos días por semana? {N/A; 1; 2; 3; 4; 5; 6; 7}

- ¿Cuánto tiempo al día (aproximadamente)? {N/A; Menos de 30 min; 30 min; 1 hora; 2 horas; Más de 2 horas}

- ¿Cuánto tiempo hace que lo practicas (aproximadamente)? {N/A; 1 semana; 1 mes; 3 meses; 6 meses; Más de 6 meses}

- ¿Qué nivel de esfuerzo realizas en la actividad física? {N/A; Ligero; Moderado; Intenso; Máximo}

A.3   Research publications

- A. Melchor Rodríguez and J. Ramos-Castro, "Video pulse rate variability analysis in stationary and motion conditions", *BioMedical Engineering OnLine*, 17:11, 2018.

- A. Melchor Rodríguez and J. Ramos-Castro, "Pulse rate variability analysis by video using face detection and tracking algorithms", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, pp. 5696-5699, 2015.

- A. Melchor Rodríguez and J. Ramos-Castro, "Análisis de la Variabilidad de la Frecuencia Cardíaca mediante Fotopletismografía por imagen", *XXXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB)*, Barcelona, 2014.